Análise textual informatizada ou análise lexicográfica de produções acadêmicas sobre Natureza da Ciência (NdC)

Computerized textual analysis or lexicographic analysis of academic productions on Nature of Science (NOS)

Marsílvio Gonçalves Pereira
Universidade Federal da Paraíba (UFPB)
marsilvioeduc@gmail.com

Sílvia Luzia Frateschi Trivelato
Faculdade de Educação - Universidade de São Paulo (FEUSP)
slftrive@usp.br

Resumo

Este trabalho apresenta uma análise textual informatizada, lexicográfica ou lexical com o uso do *ALCESTE* e do *IRAMUTEQ* em publicações nacionais sobre Natureza da Ciência (NdC) em dois eventos nacionais (ENEBIO/EREBIO E ENPEC); cinco periódicos Qualis-A da Área de Educação em Ciências e em dissertações e teses de dois bancos de dados BDTD/IBICT e CEDOC/UNICAMP) e tem como objetivo classificar e caracterizar tendências da pesquisa brasileira sobre NdC. A análise informatizada dos dados foi realizada considerando os descritores analíticos apreendidos através dos títulos, resumos, palavras-chave de 259 publicações científicas online e procedeu-se a "Análise Hierárquica Descendente" ou "Classificação Hierárquica Descendente" e o "Índice de Similitude ou Semelhanças de Palavras". Os resultados indicam quatro classes lexicais ou tendências nessas publicações: pesquisas sobre "História e NdC - abordagens históricas e ensino"; investigações sobre "aspectos teórico-metodológicos de pesquisa sobre NdC"; "concepções de alunos sobre NdC" e pesquisas sobre "formação e concepções do professor e ensino de NdC na educação em Ciências".

Palavras chave: alceste, iramuteq, estado da arte, análise lexicográfica, natureza da ciência.

Abstract

This work presents a computerized, lexicographic or lexical textual analysis using *ALCESTE* and *IRAMUTEQ* in national publications on Nature of Science (NOS) in two national events (ENEBIO / EREBIO E ENPEC); five Qualis-A journals in the area of Science Education and in dissertations and theses from two BDTD / IBICT and CEDOC / UNICAMP databases) and aims to classify and characterize trends in Brazilian NOS research. The computerized analysis of the data was performed considering the analytical descriptors apprehended through the

titles, abstracts, keywords of 259 scientific publications online and proceeded to “Descending Hierarchical Analysis” or “Descending Hierarchical Classification”. The results indicate four lexical classes or trends in these publications: research on “History and NOS - historical approaches and teaching”; investigations on “theoretical and methodological aspects of research on NOS”; “Students' conceptions about NOS” and research on “teacher education and conceptions and NOS teaching in science education”.

Key words: alceste, iramuteq, state of the art, lexicographic analysis, nature of science.

Introdução

O conceito de Natureza da Ciência (NdC) engloba uma diversidade de aspectos sobre o que é a ciência, seu funcionamento interno e externo, como constrói e desenvolve o conhecimento que produz, os métodos que usa para validar tal conhecimento, os valores implicados nas atividades científicas, a natureza da comunidade científica, os vínculos com a tecnologia, as relações da sociedade com o sistema tecnocientífico e, vice-versa, as contribuições deste à cultura e ao progresso da sociedade (VÁSQUEZ et al. 2007). É um tema recorrente na área de Ensino de Ciências, ao mesmo tempo desafiante, sendo visto como um objeto a ser investigado e com valor para a educação científica (MARTINS; RYDER, 2014). Esses aspectos destacados e mostrados aqui, têm gerado um volume de pesquisas e publicações que precisam ser analisadas, a exemplo de trabalhos de estado da arte e de revisão sistemática no contexto da área de Educação em Ciências e Ensino de Biologia.

Alguns trabalhos que tratam especificamente sobre o tema NdC na perspectiva de estado da arte ou de revisão sistemática (HARRES, 1999; SILVA et al., 2015; AZEVEDO; SCARPA, 2017; ALCÃNTARA; BRAGA, 2017), revelam a necessidade do presente trabalho, pois na busca realizada, a pesquisa de literatura mostrou que nenhum estudo específico sobre a produção de conhecimento em NdC de modo amplo envolvendo publicações em eventos nacionais, periódicos, dissertações e teses foi localizado até a finalização desta pesquisa e apresenta uma inovação no estudo de Estado da Arte incluindo uma análise textual informatizada ou lexicogáfica através de dois softwares específicos (*ALCESTE* e *IRAMUTEQ*) como trabalho pioneiro nessa abordagem de investigação na Área de Educação em Ciências.

A análise textual informatizada é um tipo específico de análise de dados, ou seja, um tratamento analítico de material verbal transcrito, que podem ser textos produzidos em diferentes condições tais como: transcrição de entrevistas e questionários, documentos, redações, textos originalmente escritos, textos de publicações na forma de artigos em eventos e periódicos e produções acadêmicas, entre outras fontes usadas tradicionalmente em Ciências Humanas e Sociais (CAMARGO, 2005; CAMARGO; JUSTO, 2013) e que são submetidos a um tratamento analítico através de programas informatizados específicos (softwares de uso específico em pesquisas científicas quantitativas e qualitativas).

O uso de softwares na análise textual tem aparecido cada vez mais em estudos e pesquisas das Ciências Humanas e Sociais, principalmente tem auxiliado bastante quando se tem um corpus bastante volumoso (CHARTIER; MEUNIER, 2011) como é o caso do conjunto de dados utilizado nesta pesquisa.

O ALCESTE (Analyse Lexicale par Context d’un Ensemble de Segments de Texte

[ALCESTE], 2009; REINERT, 1990) ou (Análise Lexical Contextual de um Conjunto de Segmentos de Texto), um programa informático inovador porque se diferenciou em possibilitar a recuperação do contexto em que as palavras ocorriam, de modo a realizar uma análise do tipo Classificação Hierárquica Descendente (CHD), que permite uma análise lexical do corpus textual e oferece contextos (classes lexicais), caracterizados por um vocabulário particular e pelos segmentos de textos que compartilham este vocabulário (CAMARGO, 2005).

Outro programa que surgiu recentemente como alternativa ao ALCESTE, é o IRAMUTEQ (Interface de R pourles Analyses Multidimensionnelles de Textes et de Questionnaires), criado pelo pesquisador francês Pierre Ratinaud (2009) que usa o mesmo algorítmico do ALCESTE (REINERT, 1990) para realizar análises estatísticas de textos. Lahlou (2012) realça o potencial que o IRAMUTEQ tem porque além de incorporar e realizar a Classificação Hierárquica Descendente proposta por Reinert (1990) realiza outras análises lexicais que o ALCESTE não faz (CAMARGO; JUSTO, 2013). O IRAMUTEQ tem uma vantagem em relação ao ALCESTE porque é um software livre ou gratuito e desenvolvido sob a lógica da open source e ancora-se no ambiente estatístico do software R e na linguagem python e realiza diferentes tipos de análise de dados textuais, das mais simples, como a lexicografia básica (cálculo de frequência de palavras), até análises multivariadas (classificação hierárquica descendente, análises de similitude). Também organiza a distribuição do vocabulário de forma facilmente compreensível (análise de similitude e nuvem de palavras) e visualmente clara (CAMARGO; JUSTO, 2013). A análise de similitude se baseia na teoria dos grafos, ou seja, uma teoria que viabiliza as relações entre os objetos de um determinado conjunto. Esse tipo de análise possibilita identificar as co-ocorrências entre as palavras, auxiliar na identificação da estrutura de um corpus textual, pode indicar a conexidade entre as palavras e pode diferenciar as partes comuns e as especificidades em relação às variáveis descritivas na análise (MARCHAND; RATINAUD, 2012; CAMARGO; JUSTO, 2013).

Este trabalho tem como questões principais: O que tem sido mais investigado sobre NdC? Quais as tendências das pesquisas sobre NdC?

Portanto esta investigação tem como objetivo caracterizar a produção brasileira de conhecimento sobre NdC através de trabalhos de pesquisa publicados no ENEBIO, ENPEC, em periódicos nacionais, e sob a forma de dissertações e teses, evidenciando características e tendências de pesquisa como aspectos do atual "estado do conhecimento" de pesquisas sobre NdC no âmbito da área de Educação em Ciências e Ensino de Biologia tomando como base a análise textual informatizada do tipo Classificação Hierárquica Descendente (CHD) e Índice de Similitude.

Metodologia

Neste estudo procedeu-se a uma análise textual informatizada ou análise lexicográfica no âmbito de um estudo de estado da arte com uso do *ALCESTE* e do *IRAMUTEQ*, e articula abordagens qualitativas e quantitativas.

Foi realizado um levantamento e identificação de produções científicas e acadêmicas sobre o tema "Natureza da Ciência no Ensino de Ciências e Biologia", publicados nos Anais do I ao V Encontro Nacional de Ensino de Biologia com seus respectivos Encontros Regionais de Ensino de Biologia (ENEBIO/EREBIO), no período de 2005 – 2014, promovido e organizado

pela Associação Brasileira de Ensino de Biologia (SBEnBio); em Actas e Anais do I ao X Encontro de Pesquisa em Educação em Ciências (ENPEC), no período de 1997 – 2015, promovido e organizado pela Associação Brasileira de Pesquisa em Educação em Ciências (ABRAPEC); em cinco periódicos nacionais Qualis A-CAPES da área de Educação em Ciências (Revistas: Alexandria, Ciência & Educação, Ensaio: Pesquisa em Educação em Ciências, Investigações em Ensino de Ciências, e Revista Brasileira de Pesquisa em Educação em Ciências), no período de 1997 à 2015; e em Dissertações e Teses com acesso online na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) do Instituto Brasileiro de Ciência e Tecnologia (IBICT), no período de 2002 - 2017 e com base em um extrato de trabalhos acadêmicos sobre NdC inventariados e classificados por Augusto (2015), quando realizou um Estado da Arte sobre as produções com temas de pesquisa em História e Filosofia da Biologia, tomando como base o Banco de Dados do CEDOC/UNICAMP, no período de 1983 – 2010. O período de análise das produções no geral compreende os anos de 1983 – 2017, totalizando 34 anos de publicações.

Os dados foram analisados tomando como base a análise textual ou Análise Lexical Contextual de Conjunto de Segmentos de Textos através da Classificação Hierárquica Descendente (CHD), com o uso do software *ALCESTE* (2010) e através do Índice de Similitude com o uso do software *IRAMUTEQ*.

Com base no trabalho de busca e identificação de produções sobre NdC localizamos um total de 280 publicações. Na organização do corpus para uso do software, foram considerados apenas os trabalhos disponibilizados online, contendo resumos, e que estivessem escritos em português. Com base nesses critérios nosso corpus ficou constituído por 259 (92,5%) unidades de contexto inicial (UCI) que foram submetidas aos procedimentos analíticos.

A análise informatizada dos dados foi realizada considerando os descritores analíticos apreendidos através dos títulos, resumos, palavras-chave das produções científicas ou publicações e através do uso dos softwares especializados procedeu-se a "Análise Hierárquica Descendente" ou "Classificação Hierárquica Descendente" e o "Índice de Similitude".

A partir da análise da distribuição do conjunto de vocábulos nos textos o ALCESTE realiza: (1) a descrição da freqüência e do percentual das palavras, aqui a significância estatística do pertencimento das palavras a uma dada classe lexical é medida por uma estatística qui-quadrado ($\chi 2$), ao nível de significância de 5% (medida da relação entre as palavras, de acordo com os padrões de co-ocorrência entre as classes e (2) a classificação hierárquica descendente (CHD) das classes de palavras encontradas com base na proximidade de conteúdos do total do *corpus*, resultando em um gráfico na forma de dendrograma (SARAIVA, COUTINHO; MIRANDA, 2011).

A análise de similitude ou de semelhança é um tipo de análise lexical que se baseia na teoria dos grafos. Um grafo é um modelo matemático que se presta para o estudo das relações entre objetos discretos de qualquer tipo e viabiliza identificar as co-ocorrências entre as palavras e o seu resultado, apresenta indicações da conexidade entre as palavras, auxiliando na identificação da estrutura de um corpus textual, diferencia também as partes comuns e as particularidades em função das variáveis descritivas identificadas na análise (RATINAUD; MARCHAND, 2012). Segundo Mendes et al. (2016, p. 347), essa análise de semelhanças possibilita visualizar a relação entre as palavras e a sua conectividade dentro de cada classe e por outro lado a ligação entre as várias classes.

Resultados e discussão

Na análise realizada, o ALCESTE processou o corpus e reconheceu 259 unidades de contexto inicial (UCI), o que corresponde aos títulos e resumos dos trabalhos sobre NdC considerados. De acordo com este processamento, o corpus contém um total de 54000 ocorrências ou palavras, das quais 6543 são ocorrências de palavras diferentes, tendo em média 8 ocorrências por palavra e um número de hapax alto no valor de 3536, o que indica uma heterogeneidade do vocabulário utilizado pelos autores dos trabalhos que constituem o corpus (CAMARGO, 2005) que tem uma riqueza muito grande de vocabulário (98,99%).

Classificação Hierárquica Descendente (CHD)

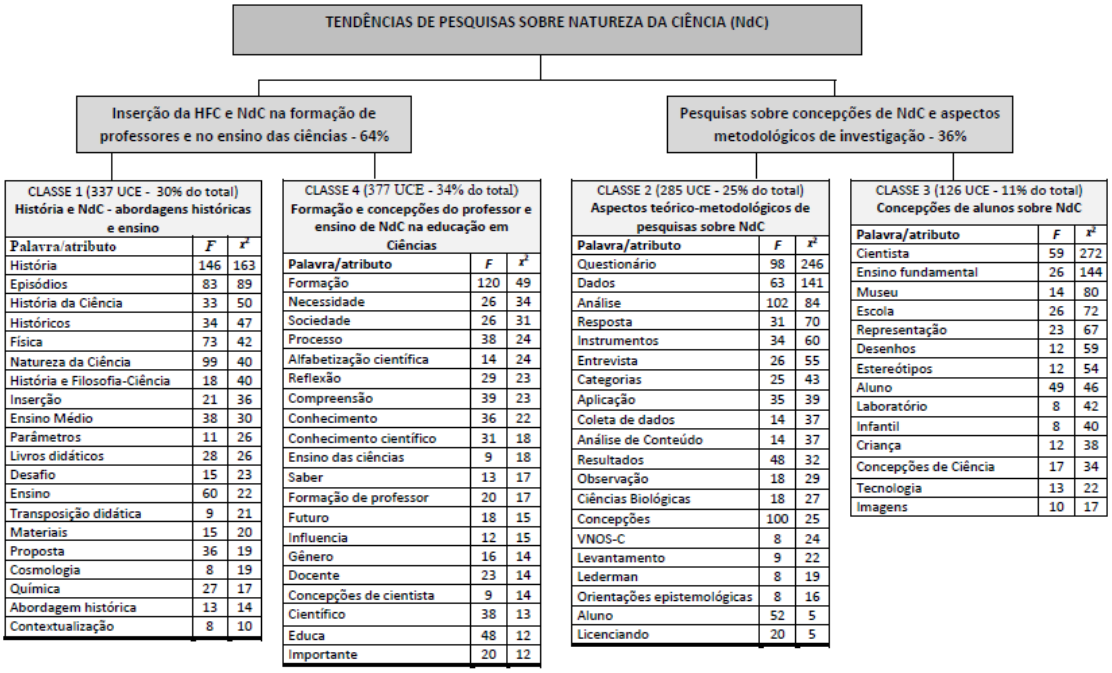
Na Classificação Hierárquica Descendente (CHD) processada, considerou-se a análise do vocábulo e dos vocabulários e foram utilizados alguns critérios lexicográficos, como por exemplo, as palavras com frequência igual ou superior à média que em nosso caso é igual a 8, ou seja, ($f \geq 8$) e com qui-quadrado ($\chi 2 \geq 3,84$). A divisão do corpus apresentou 1457 unidades de contexto elementar (UCE), contendo 6543 palavras ou vocábulos distintos, após a redução do vocabulário às suas raízes, foram encontradas 1180 palavras reduzidas e analisáveis. Como indicam Kronberger e Wagner (2002), o critério de qui-quadrado é utilizado nesta pesquisa como uma medida de relação existente entre palavras, de modo que se procura separar da maneira mais nítida possível padrões de co-ocorrência entre as classes.

A Classificação Hierárquica Descendente (CHD) reteve 77% do total das UCE ou segmentos de texto do corpus, evidenciando uma análise interessante com um índice de aproveitamento satisfatório, apreendendo 1125 das UCE do corpus, que foram organizadas em quatro classes apreendidas a partir do recorte dos 259 títulos e resumos dos trabalhos analisados. O dendrograma (figura 1) que se segue faz alusão à distribuição das quatro classes que apontam tendências das pesquisas brasileiras em NdC no período de 1983 – abr/2017, manifestas nos vários tipos de produção de conhecimento analisados.

No dendrograma, consta o título de cada uma das classes, seguido pelo número de UCE (unidade de contexto elementar) que compõe a descrição da classe, bem como as palavras de maior associação com a referida classe, segundo os critérios lexicográficos estabelecidos citados anteriormente (($f \geq 8$ e $\chi 2 \geq 3,84$). As classes (figura 1), após a categorização resultante da estruturação dos discursos categorizados pelo ALCESTE, resultaram na organização da seguinte maneira: Classe 1 – História e NdC - abordagens históricas e ensino; Classe 2 – Aspectos teórico-metodológicos de pesquisa sobre NdC; Classe 3 – Concepções de alunos sobre NdC e Classe 4 – Formação e concepções do professor e ensino de NdC na educação em Ciências. Observa-se que a distribuição das UCE entre as classes apresentou-se relativamente diferenciada entre os dois sub-corpus.

São encontradas na literatura as duas formas de escrita: dendrograma (NASCIMENTO; MENANDRO, 2006; SARAIVA; COUTINHO; MIRANDA, 2011; GLAP; BRANDALISE; ROSSO, 2014; FONSECA et al., 2015) e dendograma (BASTOS FILHO et al., 2017; AZEVEDO; COSTA; MIRANDA, 2013; CAMARGO, 2005). Conceitualmente, o dendograma ou dendrograma, é um diagrama de classificação ou árvore de classificação que exibe os grupos formados por agrupamento de observações (em nosso caso dos léxicos ou palavras) em cada passo que compõe a análise quantitativa e qualitativa e em seus níveis de similaridade ou similitude.

Figura 1 - Dendrograma da Classificação Hierárquica Descendente (CHD) de tendências de pesquisas sobre NdC em Educação em Ciências e Biologia no Brasil.



CLASSE 1 (337 UCE - 30% do total) **História e NdC - abordagens históricas e ensino**

Palavra/atributo	*F*	x^2
História	146	163
Episódios	83	89
História da Ciência	33	50
Históricos	34	47
Física	73	42
Natureza da Ciência	99	40
História e Filosofia-Ciência	18	40
Inserção	21	36
Ensino Médio	38	30
Parâmetros	11	26
Livros didáticos	28	26
Desafio	15	23
Ensino	60	22
Transposição didática	9	21
Materiais	15	20
Proposta	36	19
Cosmologia	8	19
Química	27	17
Abordagem histórica	13	14
Contextualização	8	10

CLASSE 4 (377 UCE - 34% do total) **Formação e concepções do professor e ensino de NdC na educação em Ciências**

Palavra/atributo	*F*	x^2
Formação	120	49
Necessidade	26	34
Sociedade	26	31
Processo	38	24
Alfabetização científica	14	24
Reflexão	29	23
Compreensão	39	23
Conhecimento	36	22
Conhecimento científico	31	18
Ensino das ciências	9	18
Saber	13	17
Formação de professor	20	17
Futuro	18	15
Influencia	12	15
Gênero	16	14
Docente	23	14
Concepções de cientista	9	14
Científico	38	13
Educa	48	12
Importante	20	12

CLASSE 2 (285 UCE - 25% do total) **Aspectos teórico-metodológicos de pesquisas sobre NdC**

Palavra/atributo	*F*	x^2
Questionário	98	246
Dados	63	141
Análise	102	84
Resposta	31	70
Instrumentos	34	60
Entrevista	26	55
Categorias	25	43
Aplicação	35	39
Coleta de dados	14	37
Análise de Conteúdo	14	37
Resultados	48	32
Observação	18	29
Ciências Biológicas	18	27
Concepções	100	25
VNOS-C	8	24
Levantamento	9	22
Lederman	8	19
Orientações epistemológicas	8	16
Aluno	52	5
Licenciando	20	5

CLASSE 3 (126 UCE - 11% do total) **Concepções de alunos sobre NdC**

Palavra/atributo	*F*	x^2
Cientista	59	272
Ensino fundamental	26	144
Museu	14	80
Escola	26	72
Representação	23	67
Desenhos	12	59
Estereótipos	12	54
Aluno	49	46
Laboratório	8	42
Infantil	8	40
Criança	12	38
Concepções de Ciência	17	34
Tecnologia	13	22
Imagens	10	17

Fonte: Dados originais a partir do Relatório do *ALCESTE*.

É possível observar que a distribuição das UCE entre as classes se apresentou de maneira diferenciada, com predomínio das classes 1 e 4, denotando que as tendências identificadas nas produções de conhecimento nacionais estudadas são mais expressivas em função da relação do tema NdC com abordagens didáticas de ensino e aprendizagem e com formação e concepções de professores voltadas para o trabalho do tema NdC na perspectiva da alfabetização científica, com as quais se desenham pesquisas com intervenções ou ações na escola, como se dá pelo viés de abordagens ou narrativas históricas ou uso de episódios históricos na perspectiva da inserção da História e Filosofia da Ciência nos currículos escolares de formação e profissionalização de professores e de processos de ensino e aprendizagem de NdC de modo contextualizado. Outras duas tendências que aparecem estão voltadas para aspectos metodológicos de pesquisa sobre NdC com destaque para os estudos de concepções ou representações de alunos sobre NdC e aí sobressaem-se as preocupações e discussões da área de Educação em Ciências e Ensino de Biologia com instrumentos e abordagens teórico-metodológicas, como é o caso de vários questionários disponíveis pela literatura na tentativa de capturar imagens de ciência e de cientista entre os estudantes.

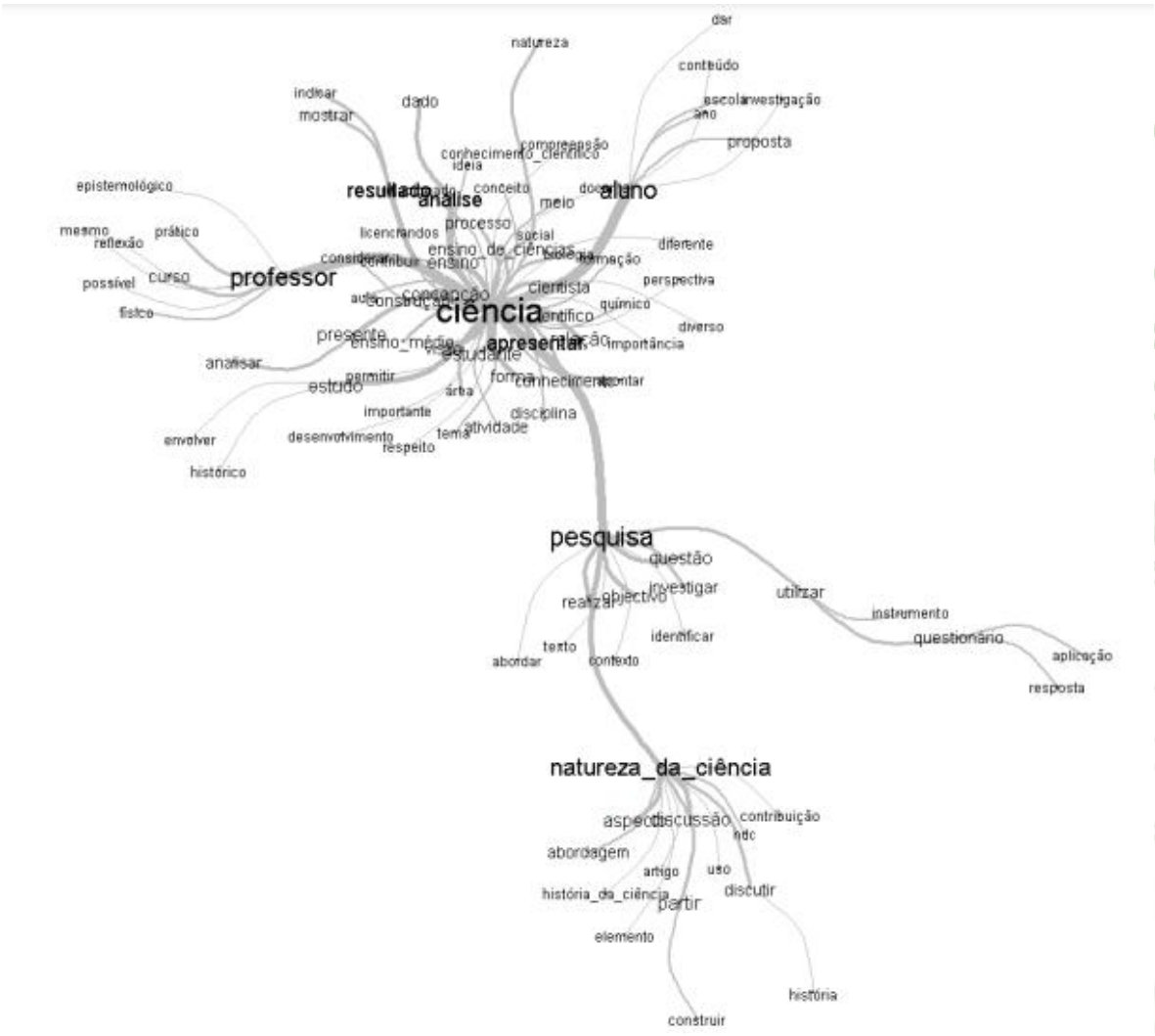
Análise de Similitude ou de Semelhança

Com base nos resultados dessa análise de similitude ou de semelhança percebe-se como as pesquisas brasileiras sobre NdC, através de suas produções textuais, relacionam as várias palavras para descreverem traços característicos e as principais tendências de suas produções de conhecimento em NdC, ou seja, é uma análise que auxilia na identificação da estrutura lexical representada pelos títulos e resumos dos trabalhos analisados. A partir da representação dispõem-se palavras fixadas a uma região central e outras em zonas periféricas.

Através da análise de similitude (figura 2) é possível identificar a estrutura, o núcleo central e sistema periférico do esquema semântico de co-ocorrências que aponta a palavra ciência

como palavra central, cujas conexões mais fortes se deram com as palavras professor, aluno, pesquisa e natureza da ciência. As palavras resultado, análise, apresentar, concepção, cientista, ensino de ciências, estudante também apresentaram ligações importantes com o léxico central ciências, além daquelas que a este termo se ligaram independentemente, como ilustra a figura 2.

Figura 2- Resultado da Análise de Similitude do *Corpus* constituído de títulos e resumos de produções científicas e publicações brasileiras sobre NdC.



Fonte: Dados originais a partir do *IRAMUTEQ*.

Observamos que a palavra ciência é o léxico mais agregador e organizador que emerge da árvore de co-ocorrências e parece indicar uma tipologia de pesquisas que partem de vários aspectos que são discutidos em torno da ciência ou do conhecimento científico. E isto envolve várias questões, aspectos, sujeitos e espaços que se articulam na produção de conhecimento sobre o tema NdC, como concepção de ciência e de cientista, ciência como processo de produção de conhecimento científico que é apresentado e divulgado a sociedade, como instituição de importância social, como disciplina escolar quando se estuda, se ensina e se

aprende sobre ciência como cultura científica. Assim, podemos perceber que os agrupamentos resultantes da Classificação Hierárquica Descendente (CHD) já apresentada e discutida anteriormente, ganha reforço através da análise de similitude, pois fica perceptível na árvore de co-ocorrências, quatro agrupamentos principais, que se dão em torno dos léxicos professor, aluno, pesquisa e natureza da ciência, que assumiram o papel de palavras centrais, se aproximando dos agrupamentos resultantes da CHD.

Conclusões

A tendência da maioria dos trabalhos brasileiros aqui analisados seguem a tendência internacional de trabalhos que investigam concepções de NdC como constatado por Azevedo e Scarpa (2017) e se debruça, principalmente, sobre o estudo das características e concepções de alunos e/ou de professores sobre NdC e concepções de NdC em materiais escolares e recursos didáticos.

De acordo com a Classificação Hierárquica Descendente (CHD), como esquema semântico ou lexical, os resultados apresentados mostraram quatro classes lexicais ou categorias que representam quatro tendências das pesquisas brasileiras sobre NdC, agrupadas em dois eixos: um eixo relacionado a pesquisas que dão atenção a aspectos da "Inserção da HFC e NdC na formação de professores e no ensino das ciências" (64%) e o segundo eixo que agrupa os trabalhos em "Pesquisas sobre concepções de NdC de alunos e aspectos metodológicos de investigação (36%).

Como resultado final da análise de similitude ou de semelhança foi produzido um esquema semântico de co-ocorrências que aponta a palavra ciência como léxico central, com conexões mais fortes com as palavras professor, aluno, pesquisa e natureza da ciência, mas também com conexões com as palavras resultado, análise, apresentar, concepção, cientista, ensino de ciências, estudante. Esse esquema lexical sugere os agrupamentos de trabalhos nas quatro classes ou tendências de pesquisas em NdC, confirmando a classificação hierárquica descendente.

Do ponto de vista metodológico, a análise textual informatizada ou lexicográfica e lexical apresentada assume uma possibilidade a mais nos trabalhos que se voltam para uma análise de estado da arte em pesquisas da área de Educação em Ciências com uso de softwares especializados.

Agradecimentos e apoios

CNPq, FEUSP, GEPEB, Grupo de Pesquisa LINCE da USP-Ribeirão Preto e Programa de Pós-Graduação em Educação/CE/UFPB.

Referências

ALCÂNTARA, M. C.; BRAGA, M. Natureza da Ciência: um estudo das influências teóricas em trabalhos publicados em periódicos brasileiros. **Enseñanza de las Ciencias**, n. extra, p. 3643-3648, 2017.

AUGUSTO, T. G. S. **O ensino de ciências e a epistemologia da biologia: o estado da arte das pesquisas realizadas no Brasil de 1972 a 2010: relatório final do estágio probatório**.

Jaboticabal-SP: Universidade Estadual Paulista (UNESP), Campus de Jaboticabal-SP. Faculdade de Ciências Agrárias e Veterinárias, 2015, 88 p.

AZEVEDO, D. M.; COSTA, R. K. S.; MIRANDA, F. A. N. Uso do ALCESTE na análise de dados qualitativos: contribuições na pesquisa em enfermagem. **Rev. Enferm**. UFPE on line., n. 7 (esp), p. 5015-22, jul., 2013.

AZEVEDO, N. H.; SCARPA, D. L. Revisão Sistemática de Trabalhos sobre Concepções de Natureza da Ciência no Ensino de Ciências. **Revista Brasileira de Pesquisa em Educação em Ciências**, v. 17, n. 2, p.579–619, agosto – 2017.

BASTOS FILHO; R. A.; PINTO, N. M. A.; FIÚZA, A. L. C.; BARROS, V. A. M. Segregação sócioespacial: uma meta-análise dos trabalhos publicados em periódicos a partir da aprovação do estatuto da cidade (2001-2017). **Holos**, v. 08, p. 298-320, 2017.

CAMARGO, B. V. Alceste: Um programa informático de análise quantitativa de dados textuais. In MOREIRA, A. S. P.; CAMARGO, B. V.; JESUÍNO, J. C.; NÓBREGA, S. M. (Eds.). **Perspectivas teórico-metodológicas em representações sociais**. João Pessoa, PB: Editora Universitária da Universidade Federal da Paraíba, 2005, p. 511-539.

CAMARGO, B. V.; JUSTO, A. M. IRAMUTEQ: Um software gratuito para análise de dados textuais. **Temas em Psicologia**, v. 21, n. 2, p. 513-518, 2013.

CHARTIER, J-F.; MEUNIER, J-G. Text Mining Methods for Social Representation Analysis in Large Corpora. **Papers on Social Representations**, v. 20, n. 37, p. 1-47, 2011.

FONSECA, E. S.; SILVA, E. P.; MAFRA, S. C. T.; FREITAS, N. C.; SAMPAIO, J. F. Viabilidade do uso do software ALCESTE na análise dos resumos disponíveis no banco de dados da CAPES: contribuições para a pesquisa sobre envelhecimento demográfico no Brasil. **Oikos: Revista Brasileira de Economia Doméstica**, Viçosa, v. 26, n. 1, p. 4-17, 2015.

GLAP, G.; BRANDALISE, M. A. T.; ROSSO, A. J. Análise da produção acadêmica sobre a avaliação na/da Educação Infantil do período 2000-2012. **Práxis Educativa**, Ponta Grossa, v. 9, n. 1, p. 43-67, 2014.

HARRES, J. B. S. H. Uma revisão de pesquisas nas concepções de professores sobre a Natureza da ciência e suas implicações para o ensino. **Investigações em Ensino de Ciências**, v. 4, n. 3, p. 197-211, 1999.

KRONBERGER, N.; WAGNER, W. Palavras-chave em contexto: análise estatística de textos. In: BAUER, M. W.; GASKELL, G. (Ed.). **Pesquisa qualitativa com texto, imagem e som**. 2ª ed. Petrópolis-RJ: EditoraVozes, 2002. p. 416-441.

LAHLOU, S. Text Mining Methods: An answer to Chartier and Meunier. **Papers on Social Representations**, v. 20, n. 38, p. 1.-7, 2012.

MARCHAND, P.; P. RATINAUD. L'analyse de similitude appliqueé aux corpus textueles: les primaires socialistes pour l'election présidentielle française. Em: **Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles**. JADT 2012. P. 687–699. Presented at the 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT 2012, Liège, Belgique.

MARTINS, A. F. P.; RYDER, J. Há realmente um consenso acerca da Natureza da Ciência no ensino de Ciências? In: **Anais** XV Encontro de Pesquisa em Ensino de Física – Maresias – 2014.

MENDES, F. R. P.; ZANGÃO, M. O. B.; GEMITO, M. L. G. P.; SERRA, I. C. C. Representações sociais dos estudantes de enfermagem sobre assistência hospitalar e atenção primária. **Rev. Bras. Enferm**, v. 69, n. 2, p. 343-50, 2016.

NASCIMENTO, A. R.; MENANDRO, P. R. M. Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. **Estudos e Pesquisas em Psicologia**, UERJ, RJ, ano 6, n. 2, p. 72-88, 2006.

SARAIVA, E. R. A.; COUTINHO, M. P. L.; MIRANDA, R. S. O emprego do software ALCESTE e o desvendar do mundo lexical em pesquisa documental. In: COUTINHO, M. P. L; SARAIVA, E. R. A. **Métodos de pesquisa em psicologia social**: perspectivas qualitativas e quantitativas. João Pessoa, PB: Editora Universitária, 2011, p. 67 – 94.

SILVA, B. V. C.; SOUSA, E. C.; NASCIMENTO, L. A.; CARVALHO, H. R. Um estudo exploratório sobre a inserção da natureza da ciência na sala de aula em revistas da área de ensino de ciências. **Holos**, ano 32, v. 7, p. 265-280, 2015.

VÁZQUEZ, A.; MANASSERO, M.A.; ACEVEDO, J.; ACEVEDO, P. Consensos sobre la naturaleza de la ciencia: la comunidad tecnocientífica. **Revista Electrónica de Enseñanza de las Ciencias**, v. 6, n. 2, p. 331-363, 2007.