

# A UTILIZAÇÃO DO WORDSMITH NAS PESQUISAS DE EDUCAÇÃO E DE ENSINO

## THE USE OF WORDSMITH IN EDUCATION AND TEACHING RESEARCHES

Clarissa Craveiro<sup>1</sup>

### Resumo

Este artigo apresenta a utilização do programa WordSmith como possibilidade de contribuição para as pesquisas de Educação e de Ensino. O programa tem sido muito utilizado no campo da Língua, todavia, destacamos sua grande contribuição como ferramenta tecnológica para outros campos de pesquisa. Utilizando o programa como uma ferramenta de acesso e organização do material pesquisado ele possibilita leituras diferenciadas para os textos e agrupamentos de documentos. Trata-se de uma “porta de entrada” ou ferramenta tecnológica que auxilia as leituras dos textos, a análise dos deslizamentos de sentidos de documentos. À guisa de exemplificação, trazemos uma pesquisa do campo do currículo que utilizou esse programa analisando os discursos das políticas curriculares para a formação de professores e os discursos oficiais do campo pedagógico. Dentre outros aspectos, sinalizamos alguns caminhos para que pesquisadores abram novas frentes e articulações com seus pares e com outros grupos de pesquisa a partir da utilização do WordSmith.

**Palavras-chave:** Programa WordSmith, ferramenta tecnológica, pesquisas de Educação/Ensino.

### Abstract

This article presents the use of WordSmith program as a possible contribution to Education and Teaching researches. This program has widely been used in Linguistics field; however, we highlight its great contribution as a technological tool for other research fields. By using such program as an access tool and organization of the material studied it allows different readings for texts and documents groupings. It is a "gateway" or a technological tool that helps text readings, analysis of document sense slips. By way of example, we bring a curriculum field research that used this program analyzing the speeches of curriculum policies for teacher training and official speeches of pedagogical field. Among other things, we signaled some ways for researchers to open new fronts and joints with their peers and with other research groups from the use of WordSmith.

**Keywords:** WordSmith Program, technological tool, Education / Teaching research.

---

<sup>1</sup> Universidade Federal Fluminense

## Considerações iniciais

Esse texto busca contribuir com pesquisas no campo da Educação e do Ensino a partir da utilização de uma ferramenta tecnológica, o programa WordSmith Tools que até então, tem sido mais utilizada no campo da Linguística (cfr. SCOTT, 2016). Pesquisas utilizando grande quantidade de documentos curriculares (cfr. OLIVEIRA, 2012, MATHEUS, 2013, CRAVEIRO 2014) destacaram a contribuição dessa ferramenta tecnológica para esse tipo de estudo. Sendo assim, é nosso objetivo apresentar e instigar a que outros pesquisadores façam uso deste ou de programas semelhantes, bem como dos materiais já organizados pelo Grupo de Pesquisa Curricular (GPc) com este suporte tecnológico.

Argumentamos que a possibilidade de trabalhar com tal tecnologia, independente da abordagem teórica escolhida pelo autor para suas análises, pode possibilitar ampliar os períodos de análises com pesquisas documentais ou com pesquisas que trabalhem com narrativas por facilitar a organização e acesso dos textos que, realizados em pesquisas “manuais”, dispendem um tempo demasiado do pesquisador que, por essa via, pode investir mais tempo em análises ou outras etapas em sua pesquisa.

Para tal, apresentamos uma breve contextualização do programa WordSmith e possibilidades de uso do programa como ferramenta tecnológica no campo do currículo, da educação e do ensino.

## Introduzindo o WordSmith

[...] não há nenhum corpus que contenha toda a informação que eu quero explorar, mas, mesmo assim, todo corpus me ensinou coisas sobre a linguagem que eu não teria descoberto de nenhum outro modo (SARDINHA, 2000, p. 43).

O programa WordSmith Tools é um artefato eletrônico originado dos estudos da linguística de *corpus*. Segundo Sardinha (2000), possui 48 anos desde a criação do primeiro *corpus* de estudo denominado *corpus* Brown, criado com a finalidade de informatizar um conjunto grande de textos. O programa é conhecido como o precursor dos estudos da linguística de *corpus*, considerada “uma das áreas de pesquisa de linguagens mais ativas nos últimos anos” (SARDINHA, 2000, p. 2).

A linguística de *corpus* trabalha com a coleta e exploração de *corpora*, ou seja, “um conjunto de dados linguísticos textuais, coletados criteriosamente com o propósito de servirem para uma pesquisa de língua ou variedade linguística” (SARDINHA, 2000, p. 3). Segundo o autor, a palavra *corpus* remete ao significado de “corpo”, entendido como um conjunto de documentos. O autor

destaca, a partir de Sanchez (1995 *apud* SARDINHA, 2000, p. 8-9), a definição mais completa de *corpus*, por incorporar as principais características da organização dos dados:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

Esses estudos são desenvolvidos principalmente em várias partes da Europa, com predomínio na Inglaterra e países escandinavos, e, fora desses contextos, pesquisas com essa abordagem ainda são pouco desenvolvidas. Ainda de acordo com Sardinha, a história da linguística de *corpus* está condicionada à tecnologia não apenas para o armazenamento de *corpora*, mas principalmente para as possibilidades de exploração desses dados. Nesse campo, o WordSmith Tools versão 6 é considerado o programa mais completo e versátil por suas ferramentas computacionais para análise de *corpus*. Entende-se que o *corpus* “é um artefato produzido para a pesquisa [...], o *corpus* em si é artificial, um objeto criado com fins específicos de pesquisa” (2000, p. 14). Dessa maneira, destaca-se que a utilização do WordSmith Tools como uma ferramenta tecnológica, nas pesquisas e análises de perspectivas discursivas, é de grande auxílio. A ferramenta possibilita um acesso mais rápido aos textos selecionados que, se buscados manualmente, poderiam limitar o cruzamento de dados da análise, levando em conta o fator tempo de pesquisa e quantidade documental.

Concorda-se com Sardinha (2000) que, a partir dessas entradas nos discursos dos *corpora* de análise, é possível ter acesso a leituras, por diferentes entradas que o programa permite, o que possibilita comparar os discursos dos *corpora* de forma mais abrangente. A compreensão das aproximações e diferenças entre os textos acessados por uma gama maior de documentos pode contribuir, de fato, para algumas afirmações com relação à comparação entre textos selecionados. Isso se dá pela possibilidade de acessar os discursos com mais facilidade e rapidez por meio da repetição de ideias no texto, busca pela frequência da repetição, fácil localização das sentenças e identificação do documento a que se refere. Por meio dos caminhos de acesso que o programa oferece, diferentes testes de leituras realizados para análise de documentos facilitam a seleção e a identificação de algumas lutas, defesas, ataques, conflitos e omissões nos discursos dos textos selecionados. Além desse aspecto, acessar as diferentes partes dos textos e compará-las por meio tecnológico permite organizar o material digitalmente através dos significantes de maior interesse na temática da pesquisa e armazenar um extenso material de análise para dar continuidade a

futuras investigações. É possível que, de outra maneira, se leve muito mais tempo para analisar, selecionar e organizar esse material.

### **Utilizando o WordSmith em pesquisa no campo do currículo**

Utilizamos o WordSmith Tools versão 5, como ferramenta tecnológica, em uma pesquisa no campo do currículo. Nessa pesquisa (CRAVEIRO, 2014), trabalhamos com documentos curriculares produzidos nos contextos políticos dos governos Fernando Henrique Cardoso (FHC) e Lula da Silva (Lula) fazendo uma busca nos anos de 1995 a 2011 na temática da formação de professores.

A preparação dos *corpora* (MEC<sup>2</sup>, FHC, Lula) de documentos curriculares para o programa WordSmith Tools foi realizada de acordo com alguns passos que destaco a seguir, utilizando a descrição com o material empírico da pesquisa.

O WordSmith Tools não faz a seleção do material a ser trabalhado, por isso, é necessário que o pesquisador já tenha o material organizado para a análise que deseja realizar. Todo o processo metodológico documental é de responsabilidade do pesquisador, como a delimitação do material e a opção por organizar as temáticas dos *sub-corpora*. Sendo assim, em um primeiro momento da pesquisa, foi utilizado outro programa para seleção dos documentos, chamado *Boot Cat front-end*.

O *Boot Cat front-end* é um programa tecnológico de livre acesso na web que permite uma varredura em todos os *sites* da internet por meio de algumas “palavras-semente”, segundo nomenclatura dada por esse programa e que, na linguagem coloquial significa ‘palavras-chave’ ou ‘palavras de busca’. O programa sugere a inserção de, no mínimo, 5 (cinco) palavras-semente relacionadas com a temática que se deseja encontrar nos sites da web. Através dessas palavras-chave, a busca apresenta todos os sites disponíveis na web que contenham essas palavras. Os sites disponibilizados permitem ler os textos selecionados na íntegra na medida em que estão disponíveis na rede.

Para essa pesquisa, Craveiro (2014), as palavras-semente escolhidas foram: 1) formação de professores, MEC, INEP<sup>3</sup>, ensino fundamental, perfil; 2) professores, MEC, escola básica, parecer, diretrizes; 3) formação de professores, pedagogia, curriculares, perfil, MEC.

A cada busca, o programa apresentou vários *sites* que continham essas palavras-semente. Verificou-se cada um e foram selecionados os que apresentavam os documentos que estavam de acordo com a temática da pesquisa. Além da utilização desse programa, os *sites* do MEC e do

---

<sup>2</sup> Ministério da Educação (MEC).

<sup>3</sup> Instituto Nacional de Pesquisas Educacionais Anísio Teixeira.

INEP também foram consultados. Outros documentos foram acrescentados de acordo com as citações destes no próprio texto dos documentos já selecionados.

Esse foi o caminho escolhido para compor o *corpus* denominado MEC. Esse material foi finalizado com 35 (trinta e cinco) documentos curriculares. Atualmente esse corpus foi ampliado para 119 (cento e dezenove) documentos curriculares (CRAVEIRO, 2016a).

Após essa etapa, os documentos foram passados para o formato txt para a inserção dos documentos no programa WordSmith Tools.

A composição de cada *corpus* não foi aleatória, ainda que não tenham sido utilizados os princípios da linguística de *corpus*, conforme Sardinha (2000). Para a composição do material de análise, foi levado em consideração a experiência da organização do material segundo Sardinha (2000) e buscado previamente conhecer os documentos relacionados com a temática da pesquisa na íntegra, organizar a divisão dos documentos por *corpus* de maneira que pudesse atender a possibilidades de leituras que permitissem uma análise discursiva dos contextos políticos pelo viés da Teoria do Discurso de Ernesto Laclau, que foi a abordagem teórica escolhida pela autora.

A organização do material por meio do WordSmith Tools sugere alguns critérios que podem ser identificados: pela origem do material (documentos do MEC, documentos do MEC no período FHC, documentos do MEC no período Lula), a composição (documentos curriculares de formação de professores da Educação Básica), a formatação (legível pelo programa em formato txt), a representatividade (representativo na variedade de documentos curriculares e no campo pedagógico)<sup>4</sup> e a extensão (de acordo com a finalidade/intencionalidade da pesquisa que compreende o período a partir da LDBEN/96 ao final do governo Lula). Na construção dos *corpora* desta pesquisa, os dados apresentaram-se de modo escrito, em um tempo sincrônico de 1996 a 2011, de conteúdo documental, e com a finalidade de estudo (*corpus* FHC e *corpus* Lula) e de referência (*corpus* da língua portuguesa para fins de contraste com as palavras do *corpus* de estudo). Para esta pesquisa foi utilizado o Banco de Português, um *corpus* de português do Brasil que foi criado e é mantido pelo projeto Direct, fazendo parte dos bancos de dados do Centro de Pesquisa, Recursos e Informação em Linguagem (CEPRIL) e do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem (LAEL) da PUC/SP. Segundo a classificação do autor, esses *corpora* são de porte pequeno/médio, contendo entre 80 (oitenta) a 250 (duzentos e cinquenta) mil *tokens* ou palavras. O *corpus* MEC possui 223.581 (duzentos e vinte e três mil, quinhentos e oitenta e uma) palavras. Esse foi o *corpus* inicial, organizado com todos os documentos, que possibilitou reconhecer as palavras-chave originadas pelos documentos inseridos no WordSmith Tools que interessariam para o acesso aos textos na íntegra relacionados

---

<sup>4</sup> Ainda que Sardinha (2000) reconheça que não haja critérios objetivos por não haver limites na extensão do corpus.

ao objetivo dessa pesquisa. A partir das palavras-chave do *corpus* MEC, iniciou-se o acesso por meio das palavras-chave comuns aos dois *corpora*: o *corpus* FHC possui 127.848 (cento e vinte e sete mil, oitocentos e quarenta e oito) palavras e o *corpus* Lula que possui 114.517 (cento e quatorze mil, quinhentos e dezessete) palavras. Buscando dar continuidade a essa pesquisa, a autora organizou um corpus de Referência de Documentos Curriculares que será disponibilizado de domínio público (CRAVEIRO, 2016b).

Os *corpora* devidamente organizados foram inseridos separadamente no WordSmith Tools. As *keywords*<sup>5</sup> (ou palavras-chave) facilitam a busca de alguns significantes de maior ocorrência nos textos. Para esta pesquisa, considerou-se que o livre fluxo desses significantes nos textos pode estar associado a construções discursivas hegemônicas ou poderia apresentar a flutuação de sentidos a partir de um mesmo significante.

As possibilidades de entrada nos textos dão-se por meio das ferramentas: *collocates* (ou colocados), *keywords* (ou palavras-chave), WordList e Concord, cujas funções encontram-se detalhadas no Quadro 1.

**Quadro 1** – Comandos do programa WordSmith Tools utilizados

<i>Keyword</i>	Palavra-chave que foi marcada pelo programa pelo número de repetições. Pode marcar um gênero, uma prática discursiva.
<i>Collocates</i> (colocados)	Palavra à esquerda ou à direita da <i>keyword</i> ou nóculo.
<i>Cluster</i>	Amarrados ou feixes de palavras não são expressões fixas da língua, mas do texto. Indica o que é frequente ou repetitivo. É aleatório. É o que marca repetidamente o discurso, um padrão. É a língua em uso. Com 2 palavras, bigrama, com 3 palavras, trigrama e com 4 palavras, quadrigrama
<i>Concord</i>	A <i>keyword</i> localizada no contexto do texto.
<i>Corpus</i>	Grupo de textos organizados por alguma classificação
<i>Corpora</i>	Plural de corpus

Os *colocados* no programa WordSmith Tools têm a função de condução de sentidos por contribuir com o pesquisador na “orientação” do olhar na leitura dos textos por meio das *keywords*. Dessa maneira, a compreensão dos contextos pelas *keywords* no texto pode ser feita com maior rapidez, bem como a seleção dos textos relacionados com a temática pesquisada e que interessam ao estudo proposto. Ou seja, os *colocados*, ao indicarem as palavras mais próximas à *keyword* (palavra-chave) – que será a palavra de ‘entrada’ nos textos - facilitam a primeira seleção dos textos que estão de acordo com a pesquisa e o descarte dos textos que se distanciam da pesquisa por apresentarem sentidos diferentes ao que essa pesquisa se propõe. Esse aspecto é importante, pois facilita a seleção do material que interessa ser analisado para a pesquisa.

<sup>5</sup> Conforme nomenclatura utilizada pelo programa WordSmith para as palavras-chave ou palavras de busca.

Assim, o WordSmith Tools possibilita uma gama enorme de “entradas” nos textos pesquisados, de acordo com as diferentes opções de leitura. Uma das possibilidades de acessar os textos é através das palavras situadas ao lado da *keyword*, localizadas à esquerda ou à direita da *keyword*, denominadas de *collocates*. Inicialmente foram feitas várias leituras por meio do *collocates*. Entretanto, nessa pesquisa, a “entrada” ou análise inicial dos textos pesquisados não pode ser realizada por meio dos *collocates*, pois a *keyword identidade docente* não foi identificada inicialmente nos documentos pesquisados. Por isso, foi preciso identificar alguns significantes/*keywords* que pudessem remeter aos discursos de identidade docente, já que esses sentidos não foram identificados *a priori* no material selecionado para o *corpus* de pesquisa ou foram identificados em número bastante reduzido nesses documentos, o que não contribuiria com a pesquisa. Em outras palavras, buscaram-se outras *keywords* que apareceram e que pudessem trazer significados que interessassem para o tema da pesquisa, já que para a temática da *identidade docente* não apareceu como bigrama.

Aqui se destaca que é na seleção de qual ferramenta do programa será utilizada para a análise que o caminho de cada pesquisador manifesta as suas escolhas subjetivas, tanto para a seleção dos textos quanto na utilização do programa. Esse movimento advém de suas experiências e conhecimento do campo. Por isso, pode-se dizer que para essa pesquisa, o programa teve a função organizadora e facilitadora do material, sendo que as escolhas das *keywords*, seleção de textos e organização do material são de responsabilidade e opção do pesquisador. Nesse sentido, a utilização do programa WordSmith pode ser realizada por qualquer abordagem teórica metodológica.

Considerando que o foco desta pesquisa foi analisar os projetos das políticas curriculares para a formação de professores, compreendemos que essa busca é ampla e envolve várias dimensões da função docente como: atuação, projeção de formação, atividades docentes, avaliação da formação docente, e outros. Sendo assim, selecionamos, para nova busca, as seguintes *keywords*: *docente, docentes, docência, professores, professor, formação, formadores, currículo, currículos, curriculares e perfil*.

Essas *keywords* selecionadas e associadas aos *colocados* permitiram perceber outros sentidos para a *keyword*. Segundo o programa, quando a *keyword* é associada a um *colocado* (*keyword* associada a uma ou mais palavras), é denominada “bigrama” e, quando associada a dois *colocados*, é denominada “trigrama”. Essas *keywords*, bigramas ou trigramas, permitem outras entradas para as leituras dos textos. A seleção de *colocados* com as *keywords* funciona como “um jogo” de tentativa e erro nos textos, em que o pesquisador vai definindo quais são os bigramas e/ou trigramas e, em alguns casos, quadrigramas, que, no contexto dos documentos, permitem acesso a discursos que

são significativos para a pesquisa e que, no caso desta pesquisa, agregavam significados e sentidos à constituição da identidade docente.

Os *colocados* foram importantes para construir alguns bigramas ou trigramas. Por exemplo: com a *keyword* **docência**. Alguns *colocados* como **da, a, à, para e de**, quando associados à *keyword* **docência (a docência, à docência, para docência, de docência)** traziam outros sentidos ao entendimento da formação de professores: *quem são, que tipo de conhecimentos precisam ter, que funções exercem*. O *colocado* **própria** só se apresentou como significativo para a compreensão dos discursos dos documentos no trigrama **a própria docência**.

Alguns bigramas e trigramas não foram “montados” a partir de *keywords* e *colocados*, mas foram selecionados por meio de outra ferramenta do WordSmith Tools denominada Clusters (são as expressões mais repetidas no discurso) que também apresentaram possibilidades interessantes de análise nos documentos por apresentarem as expressões do discurso curricular mais repetidas.

Sendo assim, o acesso aos textos curriculares foi realizado por meio das *keywords*, bigramas ou trigramas e também pelos *clusters* que, conforme Quadro 1, são expressões compostas por duas ou mais palavras que manifestam expressões repetidas no texto. Essas expressões mobilizam sentidos que se repetem com frequência nos discursos dos textos. Para essa pesquisa, os *clusters* representaram alguns sentidos que poderiam ser hegemônicos por apresentarem-se repetidas vezes, significando um dado para análise para identificar uma ideia hegemônica que, por meio da repetição, buscava fixar sentidos no determinado contexto.

A busca por *keyword* também permitiu identificar que alguns bigramas ou trigramas eram mais significativos para analisar a constituição da identidade docente que outros. A leitura dos documentos por esse caminho possibilitou selecionar por qual *keyword* seria definida a busca do significante no contexto do documento já que alguns bigramas apareceram repetidos. Por exemplo, as *keywords*: *professores, profissionais, docentes*. Alguns bigramas da *keyword* *professor* traziam mais sentidos para a constituição da identidade docente quando lidos no contexto de *profissional*, o que permitiu escolher a entrada de leitura nos documentos por meio do significante *profissional* e não pelo significante *professor*.

Além disso, as *keywords* *formação/formadores*, selecionadas inicialmente como palavras de busca, foram eliminadas a partir da leitura dos documentos. Essas palavras sempre estavam associadas a outras como: *professores, profissionais* ou *docentes* e, sozinhas, não apresentaram sentidos nos discursos dos documentos que fossem relevantes para esta pesquisa. Esse entendimento da leitura dos documentos é feito por meio das ferramentas do WordSmith Tools.



A partir das *keywords* selecionadas do *corpus* MEC, apresenta-se o número de ocorrências de cada *keyword*, os *colocados* selecionados como significativos nesse contexto e os bigramas, trigramas e quadrigramas encontrados pela ferramenta *Concord* (que localiza a *keyword* no meio do texto completo). Essas mesmas *keywords* do *corpus* MEC foram selecionadas para a análise do *corpus* FHC e *corpus* Lula com o objetivo de comparar a modulação de sentidos desses significantes nos dois *corpora*. A proposta de analisar o *subcorpus* Lula e o *subcorpus* FHC busca identificar as fixações, as flutuações e/ou os distanciamentos nos sentidos endereçados nas políticas curriculares para a formação de professores da Educação Básica, da Educação Infantil aos anos iniciais do Ensino Fundamental, tendo como marco histórico-político o período de vigência desses dois governos.

Com a utilização desse programa, busca-se assim uma outra modalidade de leitura de documentos para analisar os sentidos dos discursos hegemônicos para a formação de professores nesses dois grandes períodos de propostas governamentais conhecidos por seus distanciamentos político-partidários.

Esses acessos ao texto podem trazer diferentes modulações de significados para o *ser professor* a partir da leitura desses discursos curriculares. Após várias tentativas de leitura dos textos entre as *keywords*, *collocates* e *clusters*, notou-se que, por não ter um acesso definido *a priori* para buscar os sentidos e os deslizamentos para a formação da identidade docente nesses textos de política curricular, a possibilidade de entrada que mais oportunizou discursos relacionados ao tema da pesquisa foi por meio das *keywords* comuns aos dois *corpora*, já destacados acima.

As *keywords* que se repetem e aparecem nos dois *subcorpora* apresentando sentidos próximos ao da temática dessa pesquisa são: *docente*, *docentes*, *docência*; *professores*, *professor*; *formação*, *formadores*; *perfil*; *currículo*, *currículos* e *curricular*. Essas *keywords* foram delimitadas a partir da leitura dos documentos por meio das *keywords* comuns aos dois *corpora* e constituem a organização dos textos lidos e analisados.

### **Distanciamentos do campo da Linguística de *corpus***

A análise dos projetos de políticas curriculares para a formação de professores no período destacado para essa pesquisa foi organizada por meio de grupos de documentos preparados para a leitura pelo programa WordSmith Tools versão 5. Todavia, a análise desses documentos não é centrada na linguística de *corpus*, conforme estudos desenvolvidos por Sardinha (2000) a respeito da organização documental por *corpus*. O programa WordSmith foi utilizado em nossa pesquisa, conforme Craveiro (2014), somente como uma ferramenta de acesso e organização do material que possibilita uma leitura diferenciada para esses textos e agrupamentos de documentos. Trata-

se de uma “porta de entrada” ou ferramenta tecnológica que auxiliou as leituras dos textos, a análise dos deslizamentos de sentidos, visando ao entendimento dos discursos das políticas curriculares para a formação de professores e os discursos oficiais do campo pedagógico. Muito embora, a opção tenha sido a de utilizar algumas nomenclaturas comuns à linguística de *corpus*.

Inicialmente foi trabalhada a organização de todos os documentos curriculares selecionados mediante várias buscas no *site* do MEC e do INEP no período de 1996 a 2011, tendo como recorte a formação de professores na Educação Básica. A partir desse material, foram organizados alguns subgrupos ou *subcorpora* para a proposta de análise, conforme a nomenclatura utilizada pela linguística de *corpus*, quais sejam: *corpus* de documentos curriculares MEC de 1996 a 2011; *corpus* de documentos curriculares no período denominado FHC, de 1995 a 2003; *corpus* de documentos curriculares no período denominado Lula, de 2003 a 2011.

Os documentos curriculares organizados no *corpus* MEC de 1996 a 2011 foram selecionados com o intuito de conhecer a produção dos principais textos que endereçaram sentidos para a formação de professores nesse período e, a partir dessa busca, organizar a análise comparativa dos sentidos que circularam em torno da constituição da identidade docente por intermédio dos documentos curriculares dos contextos FHC e Lula.

O *corpus* de documentos denominado FHC e o *corpus* de documentos denominado Lula foram organizados separadamente para facilitar a análise entre os dois contextos de produção curricular buscando perceber as modulações de sentidos na constituição da identidade docente desses períodos. Essa organização do material favoreceu a análise comparativa da produção do campo pedagógico com a produção do campo curricular.

A utilização do programa WordSmith Tools por meio das *keywords* mais recorrentes nos documentos curriculares buscou analisar os discursos e compreender as articulações políticas defendidas pelos diferentes grupos no período selecionado que endereçam sentidos para a formação de professores.

Além desse aspecto, a análise desses *corpora* teve o intuito de possibilitar uma compreensão do que se entendia por professor, do que era defendido como necessário para a formação de professores e de quais eram os projetos de constituição de identidade para os professores nesses períodos por abarcarem propostas governamentais discursivamente diferentes. Entende-se, como Lopes (2004), que grupo político articulado em torno de Lula, ao transitar do contexto de oposição para o de situação, enfrentou várias dificuldades e, dentre elas, o “desmonte de todos os dispositivos de regulação das práticas curriculares estabelecidos em oito anos de Governo Fernando Henrique Cardoso” (p. 109). Por isso, pode-se situar as duas propostas de

governo com endereçamentos de sentidos diferentes, senão antagônicos, para os projetos nacionais curriculares.

As propostas curriculares podem delimitar o espaço de ação, mas é na articulação de demandas, que a fixação de discursos vai se formando e se hegemonizando. Por conseguinte, a luta por conquistar uma identidade nacional docente não será a luta por uma identidade universal, mas a luta pela hegemonia de sentidos de uma identidade particular que, no caso do contexto Lula, está impregnada do discurso de autonomia, cidadania e democracia. Sendo assim, procura-se analisar os sentidos de identificação docente, buscando compreender quais os sentidos que sofreram modificações ao longo desse tempo e quais os sentidos que permaneceram apesar das mudanças políticas. A partir dessas análises, procuramos identificar alguns discursos hegemônicos que são defendidos no campo curricular como necessários à formação de professores e ao processo de qualificação da identificação docente. Acreditamos que esses resultados poderiam contribuir com as futuras análises das políticas curriculares de formação de professores da Educação Básica.

### **Contribuições e limites do WordSmith para o campo do Currículo**

A utilização do programa WordSmith Tools, versão 6<sup>6</sup>, pode contribuir para as análises no campo do currículo no Brasil e também para pesquisas em outros países, pela facilidade de comparação e organização do material. Esclareço que isso é possível, na medida em que há grupos de pesquisa da área da linguística trabalhando em distintos países na elaboração de banco de dados dos seus respectivos idiomas. Por exemplo, o banco de dados utilizado nessa pesquisa foi o da PUC/SP, do grupo de pesquisa Estudos da Linguagem (LAEL). Em Portugal, a Universidade de Lisboa trabalha com um banco de dados com o Português Contemporâneo (CRPC, v 2.3) e outro Falado em Moçambique de 1986 a 1987. Isso é necessário para o trabalho comparativo com documentos, textos ou entrevistas, pois a base de comparação do banco de dados deve ser inserida na língua em que o material a ser analisado esteja redigido.

A possibilidade de trabalhar com grande número de textos, a organização do material em diferentes *corpus* de análise – por períodos históricos, por temáticas ou significantes - amplia possibilidades para a pesquisa. Por exemplo, armazenamento de material na temática de política de formação de professores com vários países; organização de documentos curriculares temáticos; formação de professores, documentos por disciplinas, produção acadêmica por universidades, políticas curriculares de diferentes níveis de ensino. Enfim, várias maneiras de

---

<sup>6</sup> Em CRAVEIRO (2014) foi utilizado a versão 5 do WordSmith. Em 2016 utilizamos a versão 6 do programa.

organização de materiais extensos que permitam o armazenamento de dados que possam subsidiar futuras pesquisas.

Todavia, conforme mencionado anteriormente, a organização desses dados é utilizada apenas para fins de localização e acesso rápido ao material. Destaco que o programa facilita diferentes leituras do material e da organização.

Tendo em vista que o acesso ao material organizado nos *corpora* se dá por meio das palavras, a concepção teórica que o pesquisador opta é fundamental para delimitar a forma com que vai utilizar e trabalhar com esses dados.

Para a Teoria do Discurso, utilizada em nossas análises como abordagem teórico-metodológica, compreende-se que as “palavras” dos materiais inseridos no *corpus* são significantes que endereçam sentidos por estarem sendo lidos em uma dada formação discursiva. Entrevistas, documentos curriculares ou qualquer outro tipo de texto que componha um *corpus* de análise permite acessar esses significantes e a análise de uma dada formação discursiva possibilita a leitura desses sentidos. Nessa perspectiva, a organização de um *corpus* significa a possibilidade de compreender uma aproximação discursiva, partindo do pressuposto de que as suas significações não dependem de como está estruturado esse campo discursivo por exemplo, no caso das políticas curriculares de formação de professores – o MEC que prescreve o texto curricular, não prescreve um texto único, há lutas de poder no discurso prescrito; esse órgão institucional também modifica sentidos nos seus discursos de acordo com as articulações de poder e que a significação é uma possibilidade de entendimento, de analisar as relações de poder. Segundo Laclau (2009, p. 5),

es la noción de ‘discurso’ como una totalidad significativa que trasciende la distinción entre lo lingüístico y lo extralingüístico. Como hemos visto, la imposibilidad de una totalidad cerrada desliga la conexión entre significante y significado.

Isso quer dizer que o discurso compreende relações de poder e que a significação das ‘palavras’ acessadas nos discursos do *corpus* não é dada de acordo com as regras e sintaxe gramaticais. O discurso, do qual fazem parte esses significantes, endereçam sentidos das relações de poder e por isso, ainda que não seja possível abarcar a totalidade dos sentidos dos significantes, através da flutuação de seus sentidos é possível compreender algumas articulações defendidas nos contextos, o fortalecimento de algum significante na cadeia discursiva e a diferença/aproximação de sentidos entre os contextos.

En ese sentido, hay una proliferación de ‘significantes flotantes’ en la sociedad, y la competencia política puede ser vista como intentos de las fuerzas políticas rivales de fijar parcialmente esos significantes a configuraciones significantes particulares. Las luchas discursivas sobre las formas de fijar el significado de un significante como ‘democracia’, por ejemplo, son centrales para explicar la

semántica política de nuestro mundo político contemporáneo. Esta fijación parcial de la relación entre significante y significado es lo que se denomina en estos trabajos ‘hegemonía’ (LACLAU, 2009, p. 5).

Por isso, organizar os bigramas ou trigramas que são construídos a partir das *keywords* foi uma opção de utilização de uma ferramenta do programa para analisar as flutuações de sentidos que esses significantes sofreram e/ou produziram determinados sentidos. Mas os sentidos não estão lá no texto e não serão “desvelados”, não se traz à tona o que está oculto ou que não está “claro”. A proposta é entender e analisar as articulações políticas que “permitiram” a fixação de algum sentido particular no texto. Outra possibilidade é analisar o porquê algumas construções discursivas permanecerem hegemônicas por determinados períodos.

Dessa maneira, a análise dos contextos dos discursos de FHC e Lula parte do princípio que há distintas forças de poder capazes de possibilitar certos discursos compreendidos por meio da análise desses *corpora*. As unidades que são fixadas nos significantes são parciais, pois a “unidade é um significante flutuante, porque seus significados só são fixados pelos conteúdos concretos dados pelas forças antagônicas” (LACLAU, 2011, p. 143).

Com esse entendimento da utilização do programa WordSmith Tools, reafirmo que este programa é uma ferramenta que pode contribuir para futuras pesquisas no campo da política curricular. A utilização desse programa pode potencializar a organização dos dados e, a partir dessa etapa, contribuir para

indicar os valores comunitários positivos que derivam da limitação dos agentes históricos, da contingência das relações sociais e dos arranjos políticos pelos quais a sociedade organiza a gestão de sua própria impossibilidade (LACLAU, 2011, p. 155).

## **Possibilidades do WordSmith nas pesquisas da Educação e do Ensino**

A proposta desse texto foi dar a conhecer o programa WordSmith Tools e suas possibilidades para o campo da Educação e do Ensino. Em nossos últimos estudos, CRAVEIRO (2016b), destacamos que no campo da Educação, apesar de muitos pesquisadores utilizarem um recorte de análise de temas em períodos de dez ou mais anos, tais análises tem sido realizadas de forma manual e, poucos estendem as análises por períodos maiores de dez anos. É de se supor que a utilização de uma ferramenta tecnológica possa auxiliar os pesquisadores no aumento da extensão temporal e no quantitativo de documentos em análises para as abordagens teóricas discursivas.

Além desse aspecto, a utilização do WordSmith como proposto nessas páginas, apenas sinaliza possibilidades de análises através das várias ferramentas que o programa oferece bem

como, deixa a cargo do pesquisador a escolha da sua abordagem teórica para sua pesquisa. Por exemplo, “a Análise do Discurso Francesa (AD), a Análise Crítica do Discurso (ACD), a Análise de Conteúdo (AC) ou a Teoria do Discurso (TD) que, aproximam-se por não focalizarem o funcionamento linguístico e sim o que há por meio dele: relações de poder, instituição de identidades sociais, processos de inconsciência ideológica e diversas manifestações humanas” (Craveiro, 2016 b, p.10).

Outra possibilidade a se destacar é a construção de bancos de referência para pesquisas com temáticas específicas como é o caso do banco de referência de documentos curriculares que construímos em Craveiro (2016 a), de livre acesso a ser disponibilizado em breve. Da mesma forma, outros pesquisadores poderão construir outros bancos de referência de acordo com as características do banco destacadas pela autora. Essa organização de dados permite pesquisas comparadas em um campo delimitado do conhecimento, por exemplo, no campo do currículo com temáticas/modalidades específicas: Educação de Jovens e Adultos (EJA), licenciaturas, estágio, ensino fundamental e outros.

Todavia, é mister destacar a impossibilidade de elencar todas as possibilidades de pesquisa a partir do WordSmith. Nesse texto, sinalizamos apenas alguns caminhos para que pesquisadores abram novas frentes e articulações com seus pares e com outros grupos de pesquisa.

## Referência

CRAVEIRO, Clarissa B. **Políticas curriculares para a formação de professores: processos de identificação docente (1995- 2010)**. 2014. Tese (Doutorado em Educação). UERJ, Rio de Janeiro.

CRAVEIRO, Clarissa (org.), 2016a. **Corpus de Documentos Curriculares**. Proped (UERJ), Rio de Janeiro. 1ª. ed. Fonte: [www.curriculo-uerj.pro.br](http://www.curriculo-uerj.pro.br). (em processo de disponibilização no site do Proped).

CRAVEIRO, Clarissa; AGUIAR, Felipe. **Possibilidades do software WordSmith Tools nas Políticas de Currículo de Formação de Professores no Brasil** (2016b no prelo).

LACLAU, Ernesto. **La razón populista**. Buenos Aires: Fondo de Cultura Económica, 2009.

LACLAU, Ernesto. **Emancipação e diferença**; coordenação e revisão técnica geral, Alice Casimiro Lopes e Elizabeth Macedo. Rio de Janeiro: EdUERJ, 2011, 222p.

LOPES, Alice Casimiro; MACEDO, Elizabeth (orgs.). **Currículo de Ciências em Debate**. Campinas, SP: Papyrus, 2004.

SARDINHA, Tony Berber. Linguística de Corpus Histórico e Problemática. **D.E.L.T.A.** [online]. 2000, vol.16, n.2, pp. 323-367. Disponível em: <http://dx.doi.org/10.1590/S0102-44502000000200005>.

SCOTT, Mike, **WordSmith Tools Tools Help**. Disponível em:  
<[http://www.lexically.net/downloads/version7/HTML/proc\\_tag\\_handling.htm](http://www.lexically.net/downloads/version7/HTML/proc_tag_handling.htm)>. Acesso em  
18 de maio de 2016.