Teacher Training Curriculum Policies in Brazil: Possibilities of Wordsmith Tools

Clarissa Craveiro¹
Federal Fluminense University, Brazil
Visiting Scholar (CNPQ) at State University of Rio de Janeiro, Brazil
Felipe Aguiar²
State University of Rio de Janeiro, Brazil

Introduction
The purpose of this text is to contribute to the research in the Curriculum field with the use of a technological tool: WordSmith Tools software. Nowadays, this technology is well known in the field of Linguistics (cfr. Scott, 2016), creating the Corpus Linguistics. The use of the software WordSmith Tools for research is not recent (Browker, Pearson, 2002; Baker, 2006; Wodak, Meyer, 2009). In Brazil, studies carried through in the group of research “Curriculo, Sujeitos, Conhecimentos e Cultura” (Curriculum: Subjects, Knowledge and Culture) of Graduate Program of Education of State University of Rio de Janeiro, since 2012, have started to use of this program for research in the Curriculum field. Amongst these studies, the ones that had used a big amount of Brazilian curricular documents (cfr. Oliveira, 2012, Matheus, 2013, Craveiro, 2014) have signaled that this technological tool contributes to the selection and organization of the material making it faster and deeper, contributing in the process of document analysis. Thus, our goal is to present and instigate other researchers to make use of this software or similar ones, as well as the materials already organized by the research group with that technological support.

However, the use of this type of technology in the field of Education has not been recurrent. For example, from the data of Craveiro’s research (2014), we looked for researches in the field of Curriculum and Teacher's Education from 2012 to 2016. We have found 656 research papers. Sixty of them dealt simultaneously with two themes: curriculum and elementary school teachers’ education, because it was the educational level chosen by the researcher. These 60 papers were read in full. Amongst these theses and dissertations, 21 had worked with a significant amount of materials including revision of theses and dissertations, papers published in annual meetings such as for the National Association of Graduate Studies and Research in Education (ANPed) or the National Meeting of Didactics and Teaching Practices (Endipe) (important institutions for education in Brazil), because they focused on diverse aspects regarding Teachers' Education. Only one paper had worked with the technological tool for the election of materials, which means an underuse of tools such as the WordSmith in research of this field.

Our argument is that the possibility of working with such technological tool, independent of the theoretical method chosen by the authors for their analyses, can make possible to extend the amount of texts and the periods of analyses and, in a larger
universe of documents, to deepen these analyses. In fact, periods of extension larger
than ten years with documentary analysis are difficult to work with “manual” research.
The WordSmith Tools software facilitates the comparison and location of texts either by
means of words or by means of pieces of documents or yet by the placement of
documents by period of time independent of the volume of the selected material. Such
proposal may lead to further deepening in the analyses, because the researcher has the
possibility to test and to modify the ways chosen for the research with promptness and
an easy notation of the process chosen for the study. Beyond this aspect, both the papers
with more data in a specific theme and papers, which accumulate long historical
contexts, can contribute in uneven way in diverse academic fields. Therefore, we defend
the possibility to extend the research in the field of Curriculum and Education with the
use of technological programs, mainly the ones with documents and narratives in
extensive period research.

What is Corpus Linguistics?
Corpus Linguistics is a field of studies that deals with the creation and analysis
of Corpora (plural of Corpus), which are sets of texts collected with criteria and
formatted for computer reading. Another important characteristic of Corpus is to be
created from authentic and natural texts, that is, texts not produced artificially, but
collected from the language in use.

In accordance with the criteria of collection of each Corpus, we can divide them
into six basic categories:

1 - Specialized Corpus: formed by texts of a specific type, for example, the
Corpus organized by this research. This Corpus will be used to analyze how a type of
text is characterized and which are its language patterns.

2 - General Corpus: formed by texts of diverse types, generally used to represent
and describe or study a language.

3 - Learner's Corpus: formed by texts written or produced by learners of specific
language. They analyze the way apprentices acquire a language, and the production of
didactic materials.

4 - Parallel Corpus: formed by the original text and its translated versions. It is
used for studies related to translation.

5 - Comparable Corpus: formed by aligned texts of the same type in more than a
variant of the language or in different languages, and used generally to compare them.

6 - Monitor Corpus: formed by enormous amounts of texts of a language and
frequently fed, aiming at studying changes in the language and the creation of new
words.

An analysis of texts based in Corpus needs to be empirical, that is, to analyze
patterns in authentic and natural texts, according to the types of Corpora already
mentioned. Moreover, it needs quantitative and qualitative theoretical and analytical
foundations. This means that the computer does not do the analysis instead of the
researcher.

In accordance with Berber Sardinha (2006), the concept of Corpus dates of Old
Greece, when it was already referring to a set of texts. The first Corpus for linguistic
study, however, appeared in the decade of 1920, and had 4.5 million words, although
not computerized.

After the popularization of the computers in the 1960s, Corpus Linguistics
gained space in the research centers, and since then it keeps revolutionizing the way
language is investigated, providing the researcher with an unimaginable before amount
of data and precision.
Applications that make analysis of Corpus are important tools for the researcher of the area. Amongst them, the application suite WordSmith Tools stands out. Created by Mike Scott from the University of Liverpool, in the United Kingdom, this software allows the researcher to make analyses of frequency and occurrence of words in a Corpus, and to process the text files for the analysis itself. According to Berber Sardinha (2004), “the intention of the program is to serve as a tool that allows the achievement of tasks related to the analyses of Corpus. This means that it was not conceived to perform by itself a definitive analysis in the place of the user”.

The three main tools of the software are Wordlist, Concord and Keywords. The first, Wordlist, produces a list of all words of the file or selected files. This list of words contains important information of the Corpus, for example, its size and the absolute and percentile frequency of each word. The second, Concord, executes concordances, that is, listings of a specific word in which it is possible to analyze the context and list of collocates (words that appear around the searched word). The third, Keywords, calculates and produces lists of keywords, that is, words whose statistics are different of those found in another Corpus (Reference Corpus).

An example of research using the intersection of Corpus Linguistics with other areas is Aguiar (2013), in which the researcher analyzes the original and translated texts of a literary work. He investigates marks of the translator in the translated text, using parallel Corpus with the original texts and its translated versions and the software WordSmith Tools.

In his research, Aguiar analyzed a young readers work of literature in its diverse volumes (The Chronicles of Narnia - As Crônicas de Nárnia - of C. S. Lewis), seeking to verify the adopted patterns to portray its protagonist - the Aslan lion - in its translation to Brazilian Portuguese. He also aimed at establishing if the parallel, suggested for the author of the Chronicles, between Aslan and the figure of the Christian God is captured in the Brazilian Portuguese version. The results pointed to a notable removing from the translated text in relation to the original text as of the construction of the Aslan character, namely: changes in the semantic prosodies, changes of Force, omissions or additions that modify the meaning and change in the agency of the Verbal Processes.

Although originally the software has been created and thought for the study in the field of Linguistics, researchers and students around the world in various fields of study, such as Discourse Analysis, Cultural Studies and Translation Studies, have been using it. Therefore, we suggest a greater use of this tool in the field of Education, as we will discuss next.

**Why using this technology in the field of Curriculum?**

The Corpus Linguistics is directly linked to the technology and has got a great growth, while technological tools, like the personal computers and software that do collection and processing of texts, had appeared and become popular (Berber Sardinha, 2000). Such tools not only make possible the storage of data in electronic form, but also the exploration of them.

Since the compilation of Corpus Brown, in 1964 (Berber Sardinha, 2000), the computers have been used for the analysis of language. However, it was from the decade of 1990 on that the use of the computer had a significant increase in some areas of knowledge. Furthermore, the introduction of flexible and accessible programs to a bigger public provided that the field of Corpus Linguistics advanced, and with it the application of computerized technologies that have appeared as support.
With these programs, it became possible the counting of signifiers, the identification of its occurrences in the texts and the statistical calculation of the data found in the texts, for example, the lexical variety of a text, and the number of occurrences of a significant word of a set of texts. Starting with the sharing of this information through the field of Corpus Linguistics, the analysis of the language could be closer to other areas of knowledge, with the needed adjustments.

Oliveira (2012) reports on some issues in the transposition of the theoretical constructs of Corpus Linguistics for the works that make use of Theory of Discourse for the research. An important difference between the two theories is that the Discourse Analysis, field in which the Corpus Linguistics is traditionally inserted, presents marks of empiricism, of realism and the conception that it is possible to represent the totality of its linguistic uses. In the Theory of Discourse, on the second hand, the language assumes a problematic of political implications, in which it is possible to present meanings of the social in accordance with the context of the discourse. An approximation between the two theories is that the language presents itself as structure that constitutes the reality. The use of the WordSmith tools for this research was given to the fact that the empirical research of the material became more agile and powerful for the researcher. The researcher yet cites that the initial idea she had, to construct a Corpus of reference in the area where its work is located—was abandoned during the research, due to the low possibility of significant words appear as keywords when trying to compare two Corpus of so profound similarity.

Differently than Oliveira's work (2012), Craveiro research (2016) takes the challenge that is building a Corpus representative enough to serve as a Reference Corpus for the field of Curriculum. Later on, it will be possible to analyze the Corpus as a whole (searching for occurrences of significants in general in the curricular texts) or separately (in its several sub-Corpora that are, Young and Adults - EJA; Teacher's Education; Primary Education; national curricular proposals etc). These analyses will make possible the comparison of the main Corpus with sub-Corpora, or even comparing it with another Corpus of study compiled by the researcher.

In a similar way, Matheus (2013) chose to use WordSmith Tools in its research because the software “makes it possible, among other things, to map the occurrence of the signifiers and to identify links established between these and other signifiers” (P. 52). The analysis of the researcher of signifier quality in the education texts was made using the tool Keyword (comparing the Corpus of study with the reference Corpus), as well as using the search of collocates of quality (words that appear around the signifier searched, inside of a predefined limit). Through these other signifiers, it was possible in the research of Matheus (2013) to analyze the relation established with these conductors of the meanings of quality.

The possibilities of use of the software WordSmith Tools, for Matheus (2013), intersect with the Theory of Discourse (Laclau; Mouffe, 1985), since it allows a broad reconfiguration of the texts in innumerable possibilities of organization. Being thus, the researcher achieved making an analysis that made possible the decentering of the texts from the sequential organization of the Corpus, because the program provides the researcher with many ways to read the texts that form the Corpus. The signifiers, with their provisory closure, are understood in its discursive formations, including the articulations that they make between words and meanings.

Craveiro (2014a) guided its research in the Curriculum field using the Corpus Linguistics as a tool of access and organization of the compiled texts, with the aim of understanding the discourses of the curricular politics for the teacher's education. In addition, the official discourses of the pedagogical field were studied. Through this...
study, the researcher aimed to analyze the meanings of identification of the teachers and the fluctuations of meaning throughout the temporal limit of the research in spite of changes in politics.

For the researcher, then, the WordSmith Tools offers a variety of options of “inputs” in the texts and different readings. Amongst the tools, Keywords was chosen initially, and the Corpus of study is compared with the Reference Corpus, through the collocates bigrams and trigrams (keywords), bigrams (with two words) or trigrams (with three words).

Craveiro (2014a) highlights, in her thesis, that the program offers an “easiness of comparison and organization of the material” (p. 83). The research can also be extended in its possibilities because of the uncountable ways the texts can be organized by the computer in the software, without loss in the original texts. However, the software does not analyze data, and therefore the theoretical conception of the researcher is so important for the work with the selected material.

Some research carried through in the field of the Curriculum with WordSmith Tools

The research made by Craveiro (2014a) with the use of WordSmith Tools version 5 is one of the research papers that draws near that technological tool to the field of Curriculum. The author worked with curricular documents produced in the political contexts of the governments Fernando Henrique Cardoso (FHC) and Lula da Silva (Lula) making a search in the years of 1995 to 2011 in the thematic of teacher's education. She chose for that research the following organization of documents to be analyzed: A) for the pedagogical discourses of the period between 1995 and 2003, called FHC period, the research of Estado da Arte of André, Simões, Carvalho and Brzezinski (1999). B) the analyses of the theses and dissertations of André (2009) and Garrido and Brzezinski (2008). C) analyses of the politics of teacher's education of ten years of Freitas (2002). These works were selected for being considered works of reference for the academic community of researchers that work with Teacher's Education. In the period between 2003 and 2011, it was carried through a research of the theses and dissertations in the Capes website using the keywords Curriculum and Teachers' Education in years 2005 to 2012, in view that state of the art research had only been carried through until the year of 2004. The option of including the 2012 research aims to accumulate of studies made in the period of the government Lula understanding the malleability of the period of Thesis defenses and dissertations of the initiated and finished research papers yet in this period and, in this way, available for the public access. In this research, 612 registers were found from which 562 works of Master's degree, 30 PhD theses and 20 dissertations of Professional Master's degree. From this first survey, works have been selected because they presented some approximation with the thematic and the focus of the research. Initially the selection was by means of the titles and, not being possible to define only by this reading, the reading of the summaries was carried through.

In order to understand the addressed directions to the teaching identity in the political context of government FHC and the government Lula and the processes of constitution of these social discourses in the field of Curriculum, documents were analyzed and constructed a way of access to the curricular documents. Initially these Corpora of documents have been built with 35 documents and later (Craveiro, 2014b) extended to 119 curricular documents. Such documents have been accessed through the search in the websites: Google, Ministry of Education (MEC) (http://portal.mec.gov.br)
and curricular documents (Resolutions and Reports) made by the Chamber of Elementary Education, Meetings of Full Council and National Council of Education. In addition, they have been found, in the site of the Ministry of Education, the referring Laws and presidential Decrees to the thematic of teacher's education. We yet visited the site of the National Institute of Studies and Research (www.inep.gov.br), having accessed the link “Legislation”. This way, in 2015, we aimed to select the curricular documents from 1995 to 2014, which standardize Teachers’ Education in Brazil.

In the 2014 analyses, we claimed that the thematic of the teaching identity is central in the formation of teachers and the curricular documents that serve to guide this formation in a given historical context, by means of national, international or local designation. Beyond this aspect, this guidance suffers modifications in the different contexts for which it is addressed, producing other meanings. At the same time, they refer to traces of meaning of previous contexts and modify and reinterpret the orientations in a constant movement of production and interpretation of curricular politics. Because of the interest in comparing the discourses defended in curricular documents of these contexts, considering the amount produced in this period, WordSmith Tools was of great aid for the preparation and election of the material analyzed with the use of the tools Concord and Keywords.

We highlighted that the research regarding this curricular politics production and the delimitation of this field of study is recent in Brazil, as emphasize Lopes and Macedo (2011). They are placed from the 1990s in Brazil and the United States and, with more time in England, since the decade of 1980s. The studies that seek to understand the curricular politics beyond the diagnostic and pragmatic perspective, related to solutions and action of local policies, are carried through in the research related to the policies in the field of social science.

In the research on Curriculum policies, it was highlighted the focus on teacher's education marks the teaching leading role, as “a key part” in the changes of curricular policies. It is considered as one who re-signifies the knowledge, spreads and transforms the politic-pedagogical discourse in various educational instances. This space however is signified by the omission of its performance or formation, referring to a discourse of blaming teachers, marked by a space of “divinizing” and of signifying the teacher as “partner” in the changes and curricular projects considered by the national and international agencies. This movement of continuous discursive production of meanings leads us to assume determined social status, simultaneously producing “new global identifications and new local identifications” (cfr. Hall, 1997, P. 1). In this continuous production of meanings, the teacher's education is marked by a trend to naturalize certain meanings for the Curriculum, establishing an interface between pedagogical discourse and the discourse of curricular policies of teacher's education.

Meanings circulate for what is understood “to be a teacher”, what constitutes the teaching identity, what can assess a positive qualification or not for what the teaching performance in relation to what the society assumes of it.

Matheus (2013) investigated the discourse of education of quality in the curricular policies for the Elementary Education made in Brazil, in the period between 2003 and 2011, in the government of Luis Inácio Lula da Silva and at the beginning of the government Dilma Roussef, seeking to understand the links established between curriculum and quality. In order to do that, she investigated production of texts of the referred policy - the Ministry of Education (MEC), by the means of the reading of 57 documents signed and/or ordered by the Secretaria of Educação Básica (SEB) Board of Basic Education, for the National Council of Education (CNE), and for the National Institute of Studies and Pesquisas Educacionais - Educational Researches (INEP).
These texts had constituted a Corpus of study for her research. Such documents are available in the Internet in the websites of MEC (http://portal.mec.gov.br) and the National Institute of Studies and Educational Research - INEP (http://www.inep.gov.br) and the ones that had not been found in these sites had been located by the website Google (http://www.google.com). She tried to select the biggest number possible of texts, both the entitled as curricular and the broader texts about Elementary Education. With this segment, she aimed at understanding the process of signification of quality in its relation to the curriculum in the formulated curricular policies. The choice of the significant quality was out of an evaluation made when perceiving that this significant name a flag that goes beyond the educational specialized discourse into the common sense. This observation was confirmed when, by means of the program WordSmith Tools, it was verified that quality was a keyword inside of the set of selected texts to appear in the Corpus of study of the research. For that research, the following tools of the WordSmith Tools have been used: Concord, Collocates, Keywords and Clusters.

Yet Oliveira (2012) worked with the thematic of the circumscribed curricular policies to the analysis of articulatory processes in the subject field of History. Starting from one of the tools of Corpus Linguistics, the program WordSmith Tools 5, the researcher submitted to analysis, by one of its instruments - Concord -, the selected empirical material. This way, she identified the way these texts signify historical knowledge, highlighting the overlapping of meanings, in editions of the Revista Brasileira de História (Brazilian Magazine of History) and the História Hoje (History Today) Magazine - publications of the National Association of History (ANPUH), in Reports, Curricular Guideline, Orientations and Curricular Parameters, amongst other texts legitimized by the Brazilian State in the period between the years of 1960 and 2010, and in six interviews.

Creating possibilities of research with WordSmith Tools

Despite the majority of the examples of research mentioned in this text being made with the use of the technological software WordSmith Tools with the theoretical foundation of the Theory of Discourse (Laclau; Mouffe, 1985), these are only one among many possibilities of investigation in the field of Education and Curriculum. We understand that it would be impossible to include all of the countless possibilities of analyses and theoretical references that are developing in the field of Discourse. Mainly the French Discourse Analysis, the Critical Discourse Analysis, the Content Analysis and others that are close to the Theory of Discourse. These references do not focus in the linguistic functioning but on what happens by means of he language: relations of power, institution of social identities, processes of ideological unconsciousness and diverse human manifestations. Such consideration is supported in the understanding that in these different perspectives of analysis, the WordSmith Tools has the role of organizer and facilitator of ways by which the creativity and necessity of each researcher directs the movement of the research in accordance with its singularity.

With modifications in the form of operation of the analyzed material and with different theoretical foundations, we saw that researchers could benefit from the Corpus Linguistics and its tools for their analysis. In the same way, these instruments can favor other fields of knowledge that analyze texts, with its rapidity in the processing of the data and its diverse possibilities of collection and visualization of statistical data and patterns of the studied texts. As we mentioned previously, the WordSmith Tools and the instruments of the Corpus Linguistics can assist the researchers of diverse theoretical foundations in their analysis of texts. First, because the perception of the researchers is
that the empirical research of the material becomes faster and powerful. Second, because the software offers many entrances in the text and different readings.

In continuity to the vast possibilities of research, we highlight Corpus of Reference of Craveiro (2016) as a contribution to different modalities of curricular research in the field of the Education and Teaching. Such aspect can be signaled by the vast amplitude of curricular documents that composes this Corpus. Such as: National Common Base Library (Biblioteca da Base Comum Nacional); Publications of SECADI and the National Board of Education (Conselho Nacional da Educação); curricular documents of Adult's Education (EJA), of Elementary Education teaching Subjects (Licenciaturas), of Special Education, of Internships; of LDB; of Pro-Literacy (Prólogo-Letramento); MEC Publications, SEB publications; international agreements documents; the Educational Legislation; the Elementary Education Publications; the Middle School Education; those of Curriculum and others, comprising a great-sized Corpus, with more than 22 million words.

Although the prominence of the possibilities of research with the use of texts, the WordSmith Tools also can be useful to the research that uses a great amount of texts of narratives. Research in the area of literature already makes use of the tool Concord to analyze the construction of narratives through adjectives and verbs, and has a good history of the study of metaphors in diverse types of texts. Pine (2014), for example, analyzed with the WordSmith elements of attitude in narratives of learners of Spanish language, to discover how the students present its evaluations of the process of learning the language. This analysis, according to the researcher, is taken as a starting point for discussions on the context of learning, the Spanish teacher's education and the policies of language teaching foreseen by the curricular orientations for the language teaching.

Moreover, the software still offers other tools that have not been analyzed here, for example the tool “Plot”, that makes a graphic description of the occurrence of a given significant throughout the text, which can give indications on the construction of its meanings while the narrative develops. The tools “Clusters” and “Patterns” are also potential aids for the researcher to find repeated patterns of phraseology in the texts. While the latter shows the collocates of a given word in order of frequency before or after the searched word, the first computes words that appear around the research word in a horizon of words of distance both to the right and to the left, inside of the sentence.

Having said that, we consider WordSmith Tools a great tool for the studies of the field of Education and Teaching. In addition, we hope that with the spreading of this technological tool in Education and Curriculum studies and the construction of the reference Corpus of Curriculum, many other ways get signalized so that researchers extend the articulation of research data and make use of the benefits of this tool for their research.

Notes

1 clacraveiro@yahoo.com.br
2 felipeaguiar2002@gmail.com
3 www.lexically.net/wordsmith.
References

Submitted: October, 03rd, 2016

Approved: December, 18th, 2016