

A TERMINOLOGIA DOS DOCUMENTOS ESCOLARES INDIVIDUAIS: FERRAMENTAS COMPUTACIONAIS, DIFICULDADES E SOLUÇÕES

Glória Pinotti Assumpção¹

RESUMO: Em estudos linguísticos, a atividade terminológica exige procedimentos e etapas tais como constituir um *corpus*, identificar os termos, verificar concordâncias, atestar o emprego por meio de referências precisas e difusão do conhecimento para diferentes públicos. Para isso, considera-se essencial, já no projeto, estabelecer o conjunto de procedimentos acompanhado de ferramentas computacionais, uma vez que o trabalho manual não mais tem sido utilizado em pesquisas recentes. O uso delas é consenso entre os terminólogos pelo ganho de produtividade, de qualidade e de acessibilidade sem precedentes. No decorrer do desenvolvimento de nossa pesquisa terminológica no domínio dos *documentos escolares individuais* deparamo-nos com uma série de dificuldades com relação às ferramentas computacionais para extração dos termos, confirmação dos contextos e difusão do conhecimento. O objetivo principal deste artigo é apresentar o modo como se procedeu à identificação e a difusão do conjunto terminológico no domínio dos *documentos escolares individuais* em Português do Brasil utilizando ferramentas computacionais de processamento, as dificuldades e as soluções encontradas.

PALAVRAS-CHAVE: terminologia; ferramentas computacionais; documentos escolares.

Terminology of individual academic documents: computational tools, problems and solutions

ABSTRACT: In linguistic studies, the terminological activity requires a series of procedures and steps such as to constitute a *corpus*, identify the terms, check concordances, attest employment through contexts and make knowledge the domain. Therefore, it is essential, since the project, establish a set of procedures including computational tools, since manual labor is rarely used nowadays. Their use is consensus among terminologists by productivity gains, quality and accessibility. During the development of our research in the field of terminology individual academic documents, we face a number of difficulties with respect to computational tools for extracting terms, confirmation of contexts and knowledge diffusion. The main objective of this paper is to present how we proceeded to the identification and dissemination of terminology in the field *individual academic documents* in Portuguese of Brazil using computational tools for processing, the difficulties and solutions found.

¹ UNESP/IBILCE/São José do Rio Preto. SP, Brasil. gloriapinotti@terra.com.br

KEYWORDS: terminology; computational tools, academic documents.

INTRODUÇÃO

Procedimentos que circunscrevem ferramentas computacionais de processamento na pesquisa terminológica descritiva são particularmente verdadeiros em pesquisas linguísticas pelo ganho de produtividade, de qualidade e de acessibilidade, além de possibilitar trabalhar com um *corpus* considerável, item sempre recomendável para a confirmação dos dados representativos do conjunto terminológico de um dado domínio.

Para se identificar a terminologia de uma determinada área de especialidade, considera-se essencial, já no projeto, estabelecer o conjunto de procedimentos que sirvam de referência e guiem essa tarefa, acompanhado de ferramentas computacionais de processamento, uma vez que o trabalho manual não tem sido usado com frequência em pesquisas atuais. Definimos, já em primeiro instante, um esboço dos recursos de que potencialmente necessitaríamos ao longo do trabalho, por exemplo, processador de textos, formato dos arquivos em texto puro digitalizado, programa de concordâncias, ferramenta para conferência e cotejamento dos dados pós-coleta e para a criação de uma base de dados.

Assim, optamos pelas ferramentas computacionais ao alcance do usuário, ou seja, as conhecidas que fazem parte do *Microsoft Office Access* e *Excel* e as disponíveis livremente na *internet*, o *Adobe*, os formatos [.txt] (texto puro) e [pdf] (*reader*), o programa *concordance*. Utilizar essas ferramentas em suas devidas funções seria importante para darmos conta das etapas: constituição do *corpus*; extração dos candidatos a termo, organização dos dados, preparação das tabelas e quadros demonstrativos, cotejamento das unidades (candidatos a termo) em contexto, revisão e confirmação dos contextos e, por fim, a difusão.

O objetivo principal deste artigo é apresentar o modo como se procedeu à identificação e a difusão do conjunto terminológico no domínio dos *documentos escolares individuais*² em português do Brasil utilizando ferramentas computacionais, as dificuldades com que nos deparamos e as soluções encontradas. Pensamos que um trabalho com estas

² Chamamos de individuais os documentos escolares entregues por candidatos em processo de seleção. A pesquisa maior intitulada *A Terminologia dos Documentos Escolares Individuais* integra o projeto O Léxico da Tradução Juramentada ó LexTraJu III, financiado pelo CNPq, em desenvolvimento na UNESP/IBILCE.

características visa contribuir para a técnica de trabalho sistemático em terminologia na área de estudos linguísticos. Nesse espírito, demonstramos o conjunto de procedimentos práticos que nos levaram a reconhecer as possíveis contribuições de programas computacionais mais acessíveis, que não sejam restritos ao ambiente acadêmico. Também, contribuir para o aperfeiçoamento do modelo metodológico de pesquisas em áreas de especialidade, para diferentes usuários e pesquisadores, tradutores e consultantes de várias profissões, técnicos administrativos e gestores escolares. Além de auxiliar pontualmente aos interessados em identificar o conjunto terminológico de uma área de especialidade, com base nos postulados da Teoria Comunicativa da Terminologia (TCT), de Cabré (1993).

A TEORIA COMUNICATIVA DA TERMINOLOGIA

Terminologia é o estudo científico dos termos utilizados nos textos e nos discursos de uma área de especialidade e tornou-se a base da comunicação especializada, sobretudo em situações em que a precisão é mais importante que a expressividade. O objetivo principal da Terminologia é organizar e divulgar os termos técnicos científicos como forma de favorecer a univocidade da comunicação especializada, Cabré (1993, 1999) e Barros (2004).

Como prática, a Terminologia nasceu da necessidade de se organizar o conhecimento científico e técnico acompanhando a era do desenvolvimento industrial no mundo; quando impulsionada pelo desenvolvimento científico e tecnológico, surgiu a necessidade de se ter designações próprias para atender à demanda de conceitos criados. Como ciência, somente a partir dos anos 1930, passou a ser reconhecida como ciência dos termos.

Boutin-Quesnel (1985, p. 17) considera a terminologia uma disciplina em seu funcionamento social, que estuda a denominação dos conceitos pertencentes a domínios especializados da experiência humana, enquanto Rey (1999, p. 54) aponta três espécies de necessidades do estudo em Terminologia:

- b)** Para a descrição sistemática dos conjuntos de termos necessários à formação dos discursos sobre um domínio socialmente distinto.

- c) Para a transmissão e difusão dos conhecimentos num domínio através de sua terminologia.
- d) Por necessidade de normas que se apliquem a todo o uso linguístico, à formação teórica, à prática complexa, à transmissão do saber.

De acordo com Pavel e Nolet (2002, p. xvii),

Terminologia significa um conjunto de palavras técnicas pertencentes a uma ciência, uma arte, um autor ou um grupo social. ... em sentido mais restrito e especializado, uma disciplina linguística consagrada ao estudo científico dos conceitos e termos usados nas línguas de especialidade. ... utilizada para proporcionar uma comunicação sem ambiguidade numa determinada área do conhecimento ou da prática, com base num vocabulário e em usos linguísticos específicos desse campo.

Os domínios, segundo Depecker (2005, p. 6), correspondem aos setores de atividade humana, campos de conhecimentos e determinados objetos próprios. Em certas áreas, a sistematização da terminologia própria é determinante para recuperar ou caracterizar a área. Uma língua de especialidade é um sistema de comunicação oral e escrita, usado por uma comunidade de especialistas de uma área particular do conhecimento.

Pelo dito acima, é possível entender que uma língua de especialidade já se encontra constituída no universo dos estudos da linguagem. Na nossa pesquisa, a língua de especialidade se encontra com grande parte dos termos na língua comum utilizados em contextos específicos. O que os torna termos de uma língua de especialidade serão os construtos teóricos, a metodologia, as análises, a circulação em diferentes comunidades e variantes que diferem parcialmente um termo que designa o mesmo conceito. (PAVEL & NOLET, 2002, p. 133; BOUTIN-QUESNEL, 1985, p. 23)

Na Teoria Comunicativa da Terminologia (TCT), de Cabré, o termo é uma unidade poliédrica, com uma face semântica, uma pragmática e outra formal. A TCT nos possibilita investigá-lo no domínio dos documentos escolares individuais, uma língua de especialidades nas situações comunicativas em que ocorre. A TCT busca evidenciar o caráter comunicativo dos termos nas áreas de especialidade, com o objetivo de:

descrever formal, semântica e funcionalmente as unidades que podem adquirir valor terminológico, dar conta de como são ativados e explicar suas relações com outros tipos de signos do mesmo ou distinto sistema, para fazer progredir o conhecimento sobre a comunicação especializada e as unidades que nela se usam. (CABRÉ, 1999, p. 133)

Ainda segundo a autora, a Teoria Comunicativa da Terminologia (TCT) é de base linguístico-comunicativo-comunicacional, com uma forte ascendência das relações mente/linguagem e fixa-se como campo de encontro da linguística, da ciência cognitiva, da ciência da informação, da comunicação e da informática.

O uso de ferramentas computacionais para o processamento de textos especializados, segundo Sardinha (2004, p. 5), surgiu em 1980 com a entrada em cena de microcomputadores pessoais. Nas pesquisas terminológicas, é comum o entendimento entre os terminólogos de que usá-las resolvem o problema da lentidão, precisão e concisão dos dados, pois possibilitam isolar os termos em seus contextos de uso com elementos à direita ou à esquerda, o que também possibilita atualizar conceitos relativos à empregabilidade captada pela ferramenta, conseqüentemente, facilitando as análises. Almeida *et al.* (2007, p. 414) apontam o crescimento de estudos e pesquisas na área de Linguística Computacional e Linguística de *Corpus* e o conseqüente aprimoramento e desenvolvimento de ferramentas computacionais voltadas para o processamento de língua natural (PLN) do português começando a interferir diretamente na prática terminológica.

A difusão é uma das etapas da metodologia em pesquisas terminológicas. Conforme salienta Rondeau (1984, p. 64), algum produto deve ser colocado à disposição de usuários, como forma de conceituar, estruturar e representar o conhecimento de um domínio. No nosso caso, construímos uma base de conhecimento informatizada no domínio dos *Documentos Escolares Individuais*, o que, de acordo com Barros (2004, p. 266), diferencia-se das bases de dados comuns pelas informações que se apresentam relacionadas entre si por meio de relações lógicas mais ou menos complexas e por serem compostas por dois tipos de base:

- uma *base de fatos*, constituída por dados conhecidos e devidamente formalizados, definições, classificações etc., a que se chama genericamente de fatos e que inclui também as relações estabelecidas entre esses elementos ou fatos.
- Uma *base de regras*, que consiste num conjunto de regras que permite, a partir dos fatos conhecidos, estabelecer novos fatos ou relações, inclusive, deduzir conclusões.

Cada termo inserido nesta base foi extraído de documentos da comunicação *in vivo* e confirmados na bibliografia especializada (apresentados na Tabela 1 e Quadro 1 adiante).

CORPORA E FERRAMENTAS

Os *corpora* no domínio dos *documentos escolares individuais* foram constituídos em *corpus* de estudo e bibliografia especializada, com documentos de tipologias diferentes coletados em visitas aos estabelecimentos de ensino, por abordagem pessoal e da *internet*.

De posse desses *corpora*, separamos onze gêneros textuais diferentes, totalizando o *corpus* de estudo com cento e quarenta e oito textos da comunicação *in vivo*.

TABELA 1 ó *Corpus* de estudo

Nº de ordem	Gêneros textuais	Comunicação <i>in vivo</i>
1	Ata	3
2	Atestado	15
3	Boletim	3
4	Certidão	5
5	Certificado	27
6	Declaração	17
7	Diploma	25
8	Ficha de matrícula	10
9	Histórico Escolar	31
10	Menção Honrosa	9
11	Requerimento	3
Totais	148

E constituímos a bibliografia especializada com trinta e seis documentos legítimos formais, gerados nas esferas federal, estadual e municipal e estabelecimentos de ensino públicos e particulares, formando o complexo teórico-legislativo a ser utilizado para a confirmação dos termos em contexto(s).

A organização da bibliografia especializada em grupos foi necessária para melhor identificar o nível hierárquico em que os termos foram gerados. Essa organização possibilitou a etapa da pesquisa de conferência dos dados, a disposição dos dados coletados em quadros e as tabelas no programa Excel.

Os textos de gêneros diferentes, tanto da comunicação *in vivo* como da bibliografia especializada, são reconhecidos no domínio por caracterizar precisão, tecnicidade e fornecer possibilidades de descrição e sistematicidade na área. São também autênticos, ou seja, um *corpus* composto de conteúdo especializado, em linguagem natural, representativo do domínio, equilibrado, pertencente ao uso escrito formal, tal como recomenda Sardinha (2004, p. 18).

QUADRO 1 ó Bibliografia Especializada

COMPLEXO TEÓRICO-LEGISLATIVO	
NOME	ESPECIFICIDADE
Grupo 1	Contém as leis de diretrizes e bases da educação nacional, leis para o ensino do Exército, ensino da Aeronáutica e ensino da Marinha; leis que normatizam a expedição de documentos escolares, a matrícula obrigatória aos seis anos de idade e regulamenta o Ensino Fundamental de nove anos.

Grupo 2	Contém decretos, pareceres e normas reguladoras (regulamento geral) tratando de política de ensino e do programa de defesa nacional, normas reguladoras dos cursos de graduação e pós-graduação <i>lato sensu</i> e <i>stricto sensu</i> . Credenciamento de instituições, a emissão de diplomas no ensino superior, sobre a educação profissional e avaliação de aprendizagem e transferência de alunos.
Grupo 3	Contém portarias e resoluções que tratam das diretrizes do mestrado profissional, categorias de docentes na pós-graduação, arquivamento e eliminação de documentos escolares, desburocratização dos documentos escolares, escrituração escolar. Também do Ensino Médio na rede estadual de ensino, avaliação de aprendizagem e transferência de alunos, cursos de especialização <i>lato sensu</i> .
Grupo 4	Contém manuais técnicos, guia para orientações, procedimentos, arquivamento, legislação no nível do estabelecimento, modelos de documentos, vocabulário técnico especializado e rotinas administrativas em secretarias escolares.

Os arquivos mencionados na tabela 1 e no quadro 1 foram preparados no formato digital, legível para computador, convertidos no formato [.txt] e em [.pdf], de modo a facilitar as etapas de exploração e extração dos dados de forma semiautomática formando as listas de concordâncias.

Para Sardinha (2004, p. 16), arquivo é o mesmo que depósito de texto sem organização prévia. Daí ser fundamental a apropriação do arquivo para o formato [.txt], livre das *stoplist*, de todo tipo de erro ou caractere que possa interferir na leitura da concordância de uma palavra ou caractere com outra(o), ou de ambos.

A seguir temos a imagem de arquivos preparados em [txt].

[List of input files]
 File1=C:\Users\Home Premium\Desktop\GLORIA_DO\CORPUS TXT\Atestados_1.txt
 File2=C:\Users\Home Premium\Desktop\GLORIA_DO\CORPUS TXT\Atas_1.txt
 File3=C:\Users\Home Premium\Desktop\GLORIA_DO\CORPUS TXT\Certificados_1.txt

File4=C:\Users\Home Premium\Desktop\GLORIA_DO\CORPUS TXT\Declarações_1.txt
 File5=C:\Users\Home Premium\Desktop\GLORIA_DO\CORPUS TXT\Diplomas_1.txt

FIGURA 1 ó Arquivos em [txt]

Uma vez estando limpo o *corpus* de pesquisa e preparado na sua totalidade em formato digital, recorreremos ao método de extração semiautomática utilizando o programa *Concordance 3.3* (WATT, 2009). Trata-se de um programa que permite fazer listas de palavras, extrair concordâncias, selecionar palavras específicas, verificar contextos, contabilizar o número de ocorrências, verificar as percentagens de distribuição, verificar as concordâncias (até quatro) à esquerda ou à direita de uma unidade, lematizar uma lista de palavras, entre outros recursos, inclusive as estatísticas.

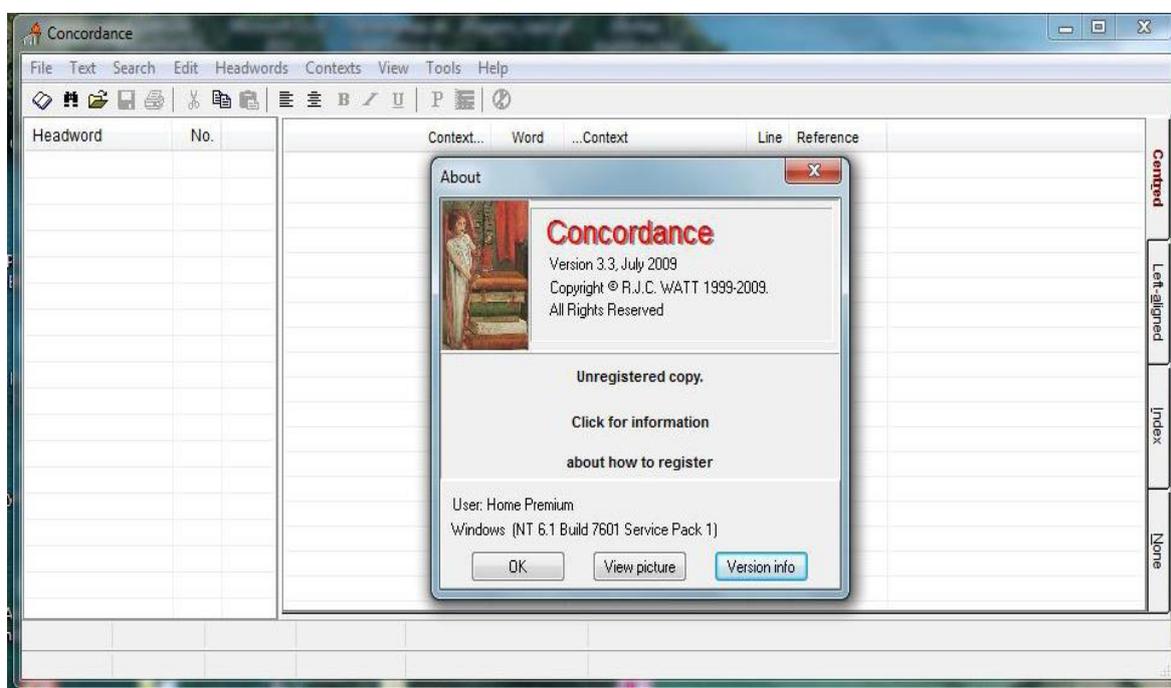


FIGURA 2 ó Interface do Programa Concordance

A listagem de unidades recolhidas no *corpus* a que procedemos foi semiautomática, tendo por base critérios de frequência, isto é, todas as palavras foram registradas, bastando aparecer uma única vez. Foi possível, então, chegarmos à seguinte tabela de concordâncias.

TABELA 2 6 Concordâncias por documento

Nº de ordem	Tipologia Textual	Candidatos a termo
1	Ata	101
2	Atestado	192
3	Boletim	94
4	Certidão	147
5	Certificado	288
6	Declaração	221
7	Diploma	176
8	Ficha de matrícula	234
9	Histórico Escolar	482
10	Menção Honrosa	83
11	Requerimento	149
Totais	2.171

A bibliografia especializada, num total de trinta e seis arquivos, também precisou passar por tratamento. Os arquivos foram convertidos para o formato [.pdf pesquisável], com alta qualidade em resolução de modo a permitir o cotejamento, etapa da pesquisa em que foi utilizada a ferramenta Adobe Professional 7.0, com o auxílio da função Localizar.

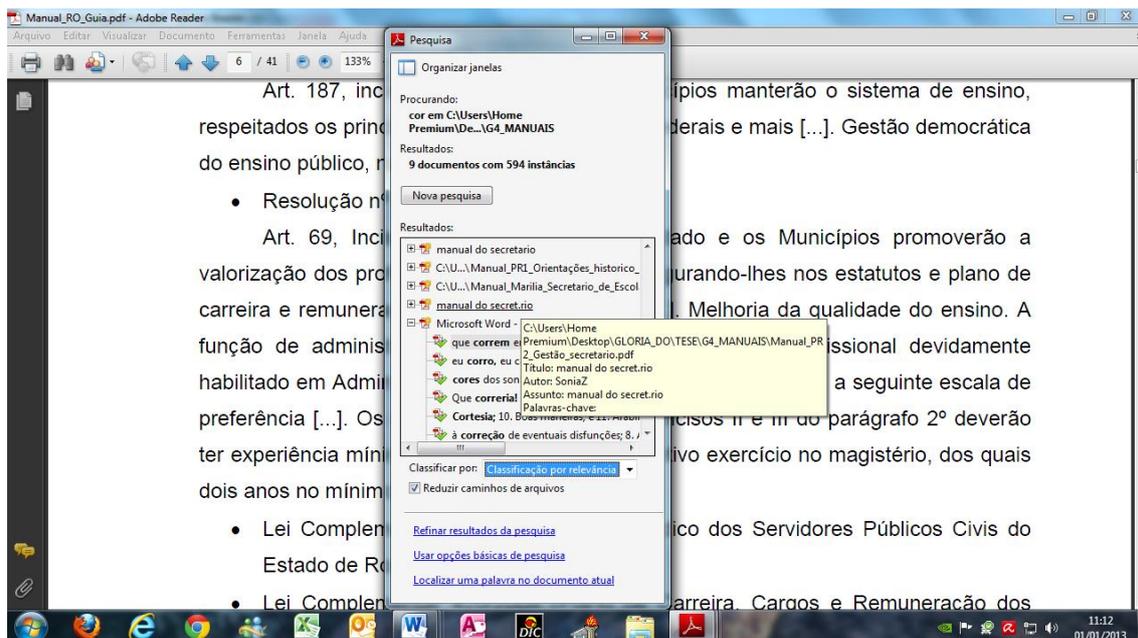


FIGURA 3 ó O cotejamento - Localizar

Do Microsoft Office 2010 ainda foram utilizados os programas Excel e Access. O Excel, para elaboração dos quadros demonstrativos dos candidatos a termos que seriam elevados ao estatuto de termos simples ou dos termos complexos na etapa cotejamento. Os candidatos a termo, num total de dois mil cento e setenta e um, foram submetidos à verificação do contexto de uso na bibliografia especializada.

QUADRO 10 - Cotejamento de termos simples

Termo simples	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	
34 bacharel																																				
35 bacharelado																																				
36 bairro																																				
37 banca																																				
38 bases																																				
39 básica																																				
40 bimestre																																				
41 biociências																																				
42 boletim																																				
43 bom																																				
44 bolsa																																				
45 bolsista																																				
46 brasileira																																				
47 brasileiro																																				
48 câmpus																																				
49 candidato																																				
50 cargo																																				
51 certidão																																				
52 carimbo																																				
53 casado																																				
54 celular																																				
55 chamada																																				
56 chefe																																				
57 ciclo																																				

FIGURA 4 ó O cotejamento - Quadros

Com a ferramenta *Access* está sendo preparada uma base de conhecimento para armazenamento dos termos em fichas terminológicas.

A base de conhecimento da terminologia dos documentos escolares individuais foi concebida com a ferramenta *Microsoft Access* 2010 para o registro descritivo dos termos, com as características particulares e delimitadas no domínio. A base integra termos simples, termos complexos e as formas braquigráficas (siglas e abreviaturas) quando estas retratarem fielmente a forma de expressão de um conceito.

A criação da base de conhecimento com o *Access* permite uma exploração rápida e eficaz dos termos, através da consulta dos registros existentes, a criação de tabelas de referência, a visualização dos registros, a escolha de campos a contabilizar, elaboração de gráficos e o fornecimento de relatórios de vários formatos.

Dificuldades e soluções encontradas

Como a maioria das pesquisas em terminologia, a utilização das ferramentas computacionais apresentou dois tipos de situação: tanto auxiliou, e reconhecemos, sem elas

seria impossível alcançarmos o resultado a que chegamos, como também surgiram várias dificuldades. As soluções possíveis foram buscadas uma a uma com a verificação pontual, a chamada extração manual, forma que somente o contexto nos permitiu chegar.

a) O Concordance

A função *õcriarõ* no programa *Concordance 3.3* (WATT, 2009) gera todas as concordâncias possíveis nos textos e não escolhe somente as palavras e, sim, caracteres identificáveis. O programa *Concordance* apenas contabiliza as ocorrências e não procede nenhum tipo de reconhecimento no padrão linguístico de classificação, nem de categoria, nem de estrutura. Dito de outra forma, são contabilizadas separadamente as formas do plural e singular, os nomes e adjetivos, verbos no infinitivo e flexionados, as formas braquigráficas numerais, abreviaturas e siglas, sinais gráficos como traço, barras, parênteses, chaves, colchetes, dois pontos, ponto e outras formas e caracteres não identificados.

Isso implicou recorrer às listas de concordância inúmeras vezes para conferência do dado pesquisado. A solução foi gerar um arquivo com listas por documento e um contendo todas as ocorrências. Esses dois arquivos foram utilizados para confrontar os candidatos a termo, um a um. Assim, a verificação do contexto foi feita pontualmente sob o olhar do terminólogo.

b) O Adobe

O Adobe Professional 7.0 foi utilizado para a fase cotejamento, na função localizar, cruzando candidatos a termo nos contextos de uso nos quatro grupos da bibliografia especializada (Quadro 1). As dificuldades encontradas, porém, foram de diferentes tipos.

Na fase de revisão dos contextos de uso, nossa atenção esteve voltada para a forma como cada candidato a termo aparece na bibliografia especializada. Essa operação se deu um a um. Deparamo-nos, então, com a resolução do [.pdf], qualidade cujo percentual de legibilidade precisa ser muito alto, caso contrário não há leitura pelo computador. Alguns textos recolhidos no formato [.pdf] eram visíveis para o pesquisador mas não legíveis pelo computador devido à baixa resolução. Refizemos os textos quando possível ou trabalhamos com estes manualmente.

Outro problema verificado foi com palavras de três dígitos como õataö, õatoö, õcorö, õmêsö e õpaiö, as quais não são reconhecidas como unidades léxicas e, sim, parte de unidades maiores. O Adobe também não diferencia plural de singular e/ou acentos como em õhorasö, õpaísö e õmêsö; o õsö dessas palavras é sempre confundido com o plural. Dessa forma, pesquisa-se õmêsö e aparece õexameö. Unidades como õexameö e õartesö são identificadas em õexaminadorö e õpartesö, problema também percebido com relação a õsecretáriaö, õsecretariaö e õsecretariadoö, õpaiö e õpaísö.

Há, ainda, de se observar que a ferramenta oferece o recurso õrefinar a pesquisa - localizar palavra inteiraö. Porém esse recurso isola a unidade, eliminando os contextos de uso. Foi necessário, então, utilizar as duas formas separadamente: localizar palavra inteira para a íntegra e localizar para o contexto, este feito pontualmente.

Reforçando, é preciso observar o contexto de fato. O termo õcorö, por exemplo, apareceu em 118 contextos apenas no grupo 2, e em 518 contextos no grupo 4, os quais continham nove documentos da bibliografia de especialidade e nenhuma ocorrência de fato. O termo õarteö apareceu em 31 contextos; õartesö cai para 17 e apenas uma ocorrência de fato em toda bibliografia especializada. Isso prova que uma ocorrência nem sempre tem relação com a frequência, melhor dizendo, foi preciso verificar o aspecto semântico dos termos inseridos em contexto para se abstrair a falsa relação alta frequência e termo.

Com os termos complexos, essa situação fica ainda mais evidente. Ao buscar o termo õdistinçãoö aparecia õdistinção de documentosö (processo de conferência) e não exatamente o termo õaprovado com distinçãoö (procedimento de classificação adotado por alguns estabelecimentos de ensino correspondente a grau e atribuído ao aluno em formato de diploma). E ainda temos exemplos como em õleiö que faz combinações com õbrasileiros (nato)ö, õreeleição (do colegiado)ö e õeleitorö; õpaiö com õprincipaisö; õprovaö com õaprovação (por merecimento)ö, õcomprovar (título)ö e õaprovados (lista de)ö. Sem contar õbancaö com õbancáriaö (conta) e õbancáriosö (Técnico em Serviços Bancários). Nesses casos, somente a conferência do contexto feita por extração manual conseguiu localizar o termo no contexto que procurávamos õbanca examinadoraö.

c) O Access

O programa *Microsoft Access* foi utilizado para a criação da base de conhecimento, fase em que tivemos dificuldades com relação à ordem numérica sequencial. A base de conhecimento não permite cancelamento de lançamentos, ou seja, não aceita eliminar um registro depois de salvo.

Uma correção feita significa um novo registro e, assim, altera-se a ordem numérica, o que impossibilita verificar a quantidade exata de termos inserida na base. Registramos casos em que tivemos de eliminar cadastro, mesmo o termo apresentando, no *corpus*, frequência superior a três ocorrências, ou inserir, mesmo apresentando frequência inferior a três ocorrências. Tivemos que proceder a algumas eliminações após o cadastramento e tentamos inserir no mesmo espaço, mas a ordem numérica não foi obedecida pelo sistema. Dito por um informata, isso ocorreu devido à chave primária no programa registrar o campo autonumeração, a sequência ser controlada pelo programa e não pelo usuário.

Outro problema detectado foi com relação à ficha terminológica Autor. Como dito, cada autor só pode ser registrado uma vez. O autor Brasil, por exemplo, aparece em um grande número de registros, dado que a bibliografia especializada é um complexo teórico-legislativo. Para resolvermos esse tipo de problema, optamos por inserir a especificidade de que trata o texto, ou seja, buscamos dar um registro com as informações do contexto. Então, o autor registrado na ficha terminológica tornar-se-ia BRASIL, matrícula 6 anos.

d) O Excel

O Excel foi utilizado para os quadros demonstrativos, pois comporta um grande número de informações e, em princípio, pareceu-nos o ideal para os quadros demonstrativos. O inconveniente foi com a quantidade de documentos consultados da bibliografia especializada e informações extraídas de cada documento, para cada candidato a termo. Listar toda a bibliografia especializada facilitaria a conferência, mas tornou-se inviável para a

leitura. Optamos, então, pela organização em grupos separados (Quadro 1), o que facilitou a visualização e possibilitou uma melhor impressão dos quadros comparativos nos contextos de uso. (Figura 4 ó o cotejamento).

À GUIA DE CONCLUSÕES

À guisa de conclusões, nossas observações são para a extração semiautomática *versus* a extração manual feita diretamente das fontes impressas, sob o olhar do terminólogo. A extração manual, embora seja muito morosa, é importante e necessária, pois permite ao terminólogo aprofundar o conhecimento da área, questionando a pertinência dos textos e termos no domínio sob variados aspectos linguísticos.

Pensamos também que a extração manual tende a influenciar o terminólogo no que concerne distinguir as unidades da língua geral das unidades das linguagens especializadas. Porém, trata-se de um trabalho que depende da boa formação linguística ou um cabedal de conhecimento adquirido com as leituras e aprofundamento da área, assim como do contato com especialistas que têm um papel fundamental neste processo.

Neste artigo apresentamos as ferramentas utilizadas em nossa investigação científica, as dificuldades e as soluções encontradas, sempre com a preocupação de mostrar que as ferramentas computacionais são fundamentais para o desenvolvimento do trabalho terminológico, mas nunca distanciadas do olhar crítico e criterioso do terminólogo, quaisquer que sejam as circunstâncias.

Identificamos 2.167 candidatos a termo, dos quais 385 foram elevados a categoria de termos simples e 170 a de termos complexos, considerados o conjunto representativo no domínio. As formas braquigráficas totalizaram setenta e quatro abreviaturas, trinta e três siglas e sessenta e nove numerais cardinais e ordinais. Todos os termos foram examinados sob o ponto de vista semântico constituídos por sua referencialidade nos textos oficiais e nos documentos.

A base de conhecimento teve os registros representativos do domínio em contextos, com todos os subsídios relevantes para uma busca do consulente, primeiramente, podendo

fornecer dados de pesquisa básica e futura publicação de um dicionário técnico. Porém, o mais relevante foi pensar na formação do conhecimento no *domínio dos documentos escolares individuais* e em modos de utilização por diferentes profissionais, principalmente os de tradução juramentada e de práticas administrativas.

A extração semiautomática, ao contrário da manual, permite uma maior rapidez em todo o processo de identificação e seleção dos termos, é inegável, quer porque temos acesso às frequências, quer porque permite que observemos os contextos, quer porque possibilita selecionar o mais adequado para o propósito do trabalho. Utilizar ferramentas computacionais na pesquisa terminológica é fundamental para garantir a efetiva localização do termo em pesquisas de terminologia descritiva levadas a cabo nos textos de especialidade.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, G. M. B; ALUISIO, S. M.; OLIVEIRA, L. H. M. *O método em Terminologia: revendo alguns procedimentos*. In: ISQUERDO, A. N; ALVES, I. M (orgs). Ciências do Léxico Lexicologia, Lexicografia, terminologia. v. III. Campo Grande/São Paulo. UFMS/Humanitas, 2007. pp. 409-420.

BARROS, L. A. *Curso básico de Terminología*. São Paulo: EDUSP, 2004.

BOUTIN-QUESNEL, R. *et al. Vocabulaire systématique de la terminologie*. Québec: Publications du Québec. 1985. 40 p. Cahiers de l'Office de la langue française. Disponível em:

http://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/voc_systematique_terminologie.pdf. Acesso em: 03 jun. 2010.

CABRÉ, M. T. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Editorial Antártica / Empuréis, 1993.

_____. *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona: IULA-UPF, (Sèrie Monografies, 3), 1999.

DEPECKER, L. *Contribution de la terminologie à la linguistique*. In: La terminologie: nature et enjeux. Langages. 157. Larousse. 2005. pp. 6-13

PAVEL, S. e NOLET, D. *Manual de Terminologia*. Adaptação para língua portuguesa de Enilde Faulstich. Canadá: Translation bureau / Bureau de la traduction. Direcção de terminologia e normalização, departamento de tradução do governo canadense, 2002.

REY, A. *La Terminologie: nons et notions*. Paris: PUF, 1979 (coll. Que sais-je?)

RONDEAU, G. *Introduction à la Terminologie*. 2. Ed. Québec: Gaëtan Morin, 1984.

SARDINHA, T. B. *Linguística de Corpus*. Manole. Barueri: 2004.

WATT. R. J. *Programa Concordance* 3.3. 2009. Disponível em: <<http://concordancesoftware.co.uk>>. Acesso em: 3 jan. 2012

Recebido em 30/03/2014.

Aceito em 15/05/2014.