

Uma proposta de vocabulário bilíngue português/inglês para o Turismo de Negócios e Eventos direcionada pelo *corpus*: da teoria à prática

A proposal of a Portuguese/English bilingual vocabulary for Business and Events Tourism directed by the corpus: from the theory to the practice

Danila Alves Carvalho*

RESUMO: Este artigo tem o objetivo de evidenciar as etapas de construção de um vocabulário bilíngue direcionado pelo corpus. Antes de demonstrar esta sequência de etapas metodológicas, abordam-se, ainda que rapidamente, as bases teóricas que orientam a produção desta obra terminológica: a Teoria Comunicativa da Terminologia (1995, 2006); a tipologia das obras lexicográficas e definição de campo de Barbosa (1990); os estudos de Krieger e Finatto (2004) sobre Terminologia; a tese de Fromm (2007) que trabalha a construção de vocabulários eletrônicos para aprendizes de tradução e os pressupostos teóricos da Linguística de Corpus, apresentados por Teixeira (2008). Depois da revisão teórica, as etapas metodológicas são descritas: a criação da árvore de domínio do turismo – esta elaborada a partir da realidade brasileira e de base motivacional; a Linguística de Corpus enquanto metodologia para a coleta e tratamento de dados via suíte WordSmith Tools, e a inserção de dados nas fichas terminológicas da plataforma Votec, que trabalha com a gestão terminológica online e também é fonte de inspiração deste trabalho (<http://www.pos.voteconline.com.br/>; FROMM, 2007), para a construção das definições. Este relato de pesquisa visa ajudar os iniciantes em projetos de Terminologia com foco em construção vocabulários, contribuindo com estudos científicos relacionados à área.

PALAVRAS-CHAVE: Vocabulário Bilíngue. Terminologia. Linguística de Corpus. Turismo de Negócios e Eventos.

ABSTRACT: This article aims to highlight the building steps of a bilingual vocabulary directed by the corpus. Before demonstrating this sequence of methodological steps, we discuss briefly the theoretical foundations that guide the production of this terminological work: the Communicative Theory of Terminology (1995, 2006); the typology of lexicographical works and field setting from Barbosa (1990); the studies from Krieger and Finatto (2004) about Terminology; the thesis from Fromm (2007) that works with the building of electronic vocabularies for translation learners and the theoretical assumptions of Corpus Linguistics from Teixeira (2008). After the theoretical review, the methodological steps are described: the creation of the concept map of the tourism field – this one related to the Brazilian reality and from motivational basis; the Corpus Linguistics as a methodology for collecting and processing data via WordSmith Tools, and the inclusion of data on the terminology records of VoTec platform (<http://www.pos.voteconline.com.br/>; FROMM, 2007) for the construction of the definitions. This research report aims to help beginners in terminology focused on vocabulary building, contributing to scientific studies related to the area.

KEYWORDS: Bilingual Vocabulary. Terminology. Corpus Linguistics. Business and Event Tourism.

* Mestre em Linguística pelo PPGEL/UFU.

1. Introdução

Uma característica inerente ao Turismo como um todo é que o mesmo acontece porque há deslocamento de pessoas; estas pessoas são os turistas, que se movimentam de um lugar para o outro por diferentes razões, dentre elas, o trabalho. De acordo com dados do Portal Brasil (2013):

O turismo de negócios é um dos segmentos mais importantes e de maior vitalidade para a economia turística do país. Ocupa a primeira posição entre os que mais aumentaram seu faturamento em 2012: cresceu 23,3% em relação ao ano anterior, de acordo com a 9ª Pesquisa Anual de Conjuntura Econômica do Turismo (Pacet), realizada pela Fundação Getúlio Vargas.

Na cidade de Uberlândia, Minas Gerais, o Turismo de Negócios e Eventos é o principal tipo de turismo local. Em publicação virtual do Uberlândia Convention & Visitors Bureau (2015), temos: “Destaque no cenário nacional de negócios e turismo, Uberlândia oferece infraestrutura de recepção, hospedagem, alimentação, comunicação e apoio logístico para eventos, congressos, seminários e feiras.” Como habitante desta cidade, observamos um crescimento constante desta área não somente por Uberlândia ser um elo estratégico entre as diversas regiões do país, mas também por ser centro logístico do Brasil, que conta com grandes empresas atacadistas e multinacionais de diferentes setores.

Assim, o trânsito de pessoas de diferentes nacionalidades aumenta, e a demanda por um vocabulário bilíngue português/inglês para o Turismo de Negócios e Eventos se faz constante. Esta proposta de vocabulário tem o objetivo de ajudar estudantes e profissionais iniciantes dos cursos de Tradução e Turismo a conhecerem os termos da área do Turismo de Negócios e Eventos no português e também no inglês.

Para este trabalho, adotamos a nomenclatura *Turismo de Negócios e Eventos*, por acreditarmos que o evento é o principal aliado dos negócios. Além disso, é ferramenta fundamental para que as pessoas se encontrem, reúnam e realizem suas trocas de conhecimento, bens e serviços em reuniões, feiras, treinamentos, convenções, rodadas de negócio, etc.

Na primeira parte do artigo, apresentamos a base teórica que é necessária ser estudada para a produção de um vocabulário, como a Teoria Comunicativa da Terminologia (TCT), o fazer terminográfico e os aspectos fundamentais da Linguística de *Corpus* (LC). Já, na segunda parte do artigo, tratamos da metodologia utilizada na

proposta de produção de um vocabulário bilíngue, como: a elaboração da árvore domínio, a compilação de dados do *corpus* de estudo e, em seguida, a análise do *corpus* de estudo e o processo de criação das definições via fichas terminológicas para se chegar ao formato que será exibido online na seguinte página da WEB – Vocabulário Técnico Online: <http://www.pos.voteconline.com.br> (FROMM, 2007).

2. Uma breve passagem por aspectos teóricos relevantes

Como parte das Ciências do Léxico, a Terminologia, grafada com T maiúsculo, é a área de estudos baseada em teorias e análises que abordam os termos a partir de um recorte científico. Já a terminologia, com t minúsculo, designa o próprio conjunto de termos de determinada especialidade. Tanto a área de conhecimento da Terminologia quanto as unidades terminológicas e suas funções devem ser ponderadas para a construção de vocabulários, uma vez que o recorte teórico do seu tratamento contribui para a feição do vocabulário especializado identificado e vice-versa.

Como bem explicam Krieger e Finatto (2004, p.22):

Essa dupla face, teórica e aplicada, que reúne tanto a descrição e a explicação dos termos, fraseologias e definição terminológica quanto o conjunto de diretrizes metodológicas para o tratamento desses objetos configura a identidade da disciplina denominada Terminologia.

Ainda de acordo com as autoras, além de ter como objeto de estudo os termos, a Terminologia também se ocupa das fraseologias e definições terminológicas, de modo que a conceituação e a utilização dos termos em textos também interessam.

A teoria terminológica na qual baseamos a proposta de criação deste vocabulário bilíngue é a Teoria Comunicativa da Terminologia (TCT), proposta por Maria Teresa Cabré e seus colaboradores do IULA - Espanha. A TCT defende a tridimensionalidade do termo ao caracterizá-lo como unidade linguística, cognitiva e comunicacional, o que amplia o olhar do terminólogo para a criação da própria definição terminológica para os vocabulários. Para Krieger e Finatto (2004, p.36): “A TCT, ao introduzir de modo sistemático uma visão linguística nos estudos terminológicos, tem impulsionado um maior conhecimento sobre a estrutura e o funcionamento do termo, objeto primeiro da Terminologia.”

Para Cabré (*apud* Almeida, 2006, p.86-87):

Um projeto terminológico vinculado teoricamente à TCT deve refletir na sua prática os seguintes pressupostos gerais: a) o objeto central da Terminologia são as unidades terminológicas e não os conceitos. Eleger as unidades como objeto central significa reforçar uma perspectiva linguística e uma abordagem semasiológica; b) não há uma diferença a priori entre termo e palavra, o que há são signos linguísticos que podem realizar-se no discurso como termo ou palavra dependendo da situação comunicativa c) os níveis lexical, morfológico, sintático e textual podem veicular conhecimento especializado; d) os termos devem ser observados no seu ambiente natural de ocorrência, ou seja, nos discursos especializados; e) a variação conceitual e denominativa deve ser considerada; f) do ponto de vista cognitivo, as unidades terminológicas: i) estão subordinadas a um contexto temático; ii) ocupam um lugar preciso num mapa conceitual; iii) o seu significado específico é determinado pelo lugar que ocupam nesse mapa.

Resumidamente, podemos dizer que a TCT tem o objetivo de entender que há variação linguística a partir de diversidades relacionadas ao uso, é caracterizada pela descrição e denominação, utiliza de um método de pesquisa que é semasiológico (partimos do nome (termo) para criar a definição do conceito) e considera a tridimensionalidade do termo (unidade linguística, cognitiva e comunicacional).

Assim contextualizada, a TCT guia o desenvolvimento desta proposta de vocabulário, pois também acreditamos na preponderância da descrição de usos reais dos termos e na variação do significado e da forma do termo, que comunica e significa tanto na área de especialidade em que está inserido quanto na língua enquanto sistema gramatical.

Mas afinal, o que é um vocabulário? Segundo a concepção de Barbosa (1990), o vocabulário pode ser compreendido também como conjuntos lexicais manifestados em uma área de especialidade e permite que escolhas sejam feitas dentro do sistema linguístico geral, ou seja, dentro de toda a língua. O vocabulário de uma especialidade terá como unidades de trabalho para um analista os termos, que apresentam significado restrito e frequências peculiares. As acepções de sentido de um dado termo, em diferentes situações de uso, integrarão um verbete de um dicionário ou levantamento de termos de uma determinada área de especialidade.

Para este trabalho, entendemos que os termos, seus sentidos e usos estão ligados a uma área de especialidade, o Turismo; mais pontualmente exploraremos o Turismo de Negócios e Eventos.

A concepção de vocabulários usufrui também do entendimento e uso criterioso da Terminografia que promove uma reflexão sobre a prática da criação de obras

terminológicas. Para Krieger (2004, p.332), a Terminografia é uma atividade prática, uma vez que oferece aporte teórico para “o estabelecimento de diretrizes metodológicas às aplicações terminológicas”. Também é possível, através dela, refletir sobre problemas relacionados à identificação e tratamento das terminologias (KRIEGER, 2004, p. 332).

Entendemos que a Terminografia é o processo que inclui desde a coleta e compilação de dados à seleção, estruturação e criação de definições para os termos da área de especialidade que se estuda a partir de princípios, critérios e objetivos pré-estabelecidos, que têm como resultado final uma descrição do uso do termo e sua definição. Ainda como parte do fazer terminográfico, a criação da definição, a partir de fontes confiáveis, e o preenchimento de fichas terminológicas são etapas fundamentais do processo que serão detalhadas em nossa descrição metodológica.

Outra compreensão teórica importante para a criação de vocabulários está relacionada ao entendimento da Linguística de *Corpus* (doravante LC) que estuda a língua e seus aspectos específicos a partir da probabilidade gerada pelo tratamento dos dados. Neste trabalho, usamos a LC, primeiramente, enquanto abordagem, pois as evidências linguísticas presentes no *corpus* é que guiam o desenvolvimento deste vocabulário e todo o processo foi realizado via procedimentos auxiliados pelo computador, outra característica importante dessa abordagem.

Como metodologia, a LC representa uma sequência de ações a serem seguidas criteriosamente, respaldadas por programas computacionais, que contribuem para a análise empírica da língua. Essas ações também podem guiar a construção de um vocabulário de especialidade.

Para Tagnin (2010, p. 358-359) há dois tipos de estudo possíveis de serem realizados a partir do *corpus*:

Estudo baseado em *corpus*: estudo em que o *corpus* é usado (a) para comprovar (ou não) uma hipótese ou (b) para extrair exemplos.

Estudo direcionado pelo *corpus*: estudo que se desenvolve conforme dados apresentados pelo *corpus*, sem pressuposições teóricas.

Neste estudo específico, toda a coleta, conhecimento da área, seleção e tratamento dos termos não visa validar uma hipótese já pronta e definida, mas pelo contrário, o *corpus* construído é que permite com que a área trabalhada seja aos poucos descoberta.

A Linguística de *Corpus* é de extrema importância para criação de vocabulários, uma vez que contribui para a coleta eletrônica específica de textos reais, de onde partirão posteriores análises. Segundo Bevilacqua (2013, p.17),

a Linguística de *Corpus*, além de estabelecer os princípios e critérios para a compilação de *corpora* (...) também oferece recursos e ferramentas que auxiliam nas diferentes etapas metodológicas terminográficas: desde a própria compilação de *corpora*, passando pela identificação de candidatos a termos e fraseologias e chegando à identificação de elementos que permitem a elaboração de uma definição.

Até este ponto entendemos, então, a importância da Terminologia, da Terminografia e da Linguística de *Corpus* na criação de obras terminográficas. É preciso haver um dado ponto de vista, teórico e metodológico, para embasar o reconhecimento e o trabalho de levantamento de usos terminológicos em um âmbito especializado, valendo isso também para o Turismo de Negócios e Eventos.

Para finalizar esta passagem pelas bases teóricas que fundamentam a proposta de um modelo de vocabulário bilíngue para o Turismo de Negócios e Eventos, vale revisar a tese de doutorado de Fromm (2007). Nela, o autor criou o VoTec, um ambiente de gestão terminológica que visa auxiliar os aprendizes de tradução na minuciosa tarefa de traduzir e onde explicou o passo a passo da construção da plataforma, que está disponível na internet (<http://www.pos.voteconline.com.br/>) e oferece vocabulários bilíngues nas áreas de Computação e Linguística.

O que marca e diferencia o VoTec em relação a outros recursos criados no cenário brasileiro é a variação de possibilidades de consulta dos termos propostos. Os modos de consulta, por exemplo, podem ser normal (leiaute de uma obra impressa) ou descritivo (o termo é descrito em detalhes em linhas sequenciais); enquanto os tipos de consulta da microestrutura do termo podem ser: Total (apresenta todos os campos disponíveis no banco de dados para o entendimento do termo), Tradutor (apresenta os campos mais relevantes para o entendimento do termo e que foram apontados pelos tradutores pesquisados,) e Modular (apresenta os campos escolhidos pelo consulente para o entendimento do termo).

A nossa proposta de vocabulário bilíngue português/inglês para o Turismo de Negócios e Eventos – identificada pelo nome VocTur, é direcionada pelo *corpus* e baseada nos passos sugeridos pelo VoTec. Esses passos passam pela construção da árvore

de domínio, coleta e tratamento de dados pela *suíte WordSmith Tools*, além da utilização das fichas terminológicas do VoTec para a criação das definições.

3. Metodologia - Etapas terminológico-terminográficas para a construção de vocabulários bilíngues

3.1 A Árvore de Domínio

Para chegarmos ao conhecimento de um campo a ser explorado e entendermos em que pontos ou categorias os termos desse campo podem ser classificados, a criação da *Árvore de Domínio/Conhecimento* (ou ainda *ontologia*¹) é fundamental. Nas palavras de Krieger e Finatto (2004, p.134) “Uma árvore de domínio é um diagrama hierárquico composto por termos-chave de uma especialidade, semelhante a um organograma.”.

Como não há uma árvore de domínio oficial do turismo, então foi necessário criá-la a partir de leituras e buscas por tipologias relacionadas ao Turismo no Brasil. Para tanto, realizamos pesquisas online que foram fundamentais para explorar as tipologias existentes de Turismo, além de pesquisas avançadas de artigos, dissertações e teses; a exploração de web sites diversos (empresas, organizações e governo) ajudou a ampliar as possibilidades de segmentação do Turismo no Brasil.

Para a realização deste trabalho, optamos pela **motivação da viagem** do turista (ou dos turistas) como base para o desenvolvimento da *Árvore de Domínio*.

Organizar as tipologias dentro de uma *Árvore de Domínio* não significa dizer que estas áreas não se relacionam; as tipologias do Turismo interagem entre si de maneira tal que um tipo de turismo pode e terá características de outras tipologias da área. Abaixo (Figura 1) apresentamos a *Árvore de Domínio* do Turismo de base motivacional:

¹ Há uma distinção, em Processamento de Linguagem Natural, sobre Ontologias, mapas conceituais e árvores de domínio. Para maiores detalhes, ver, por exemplo, uma publicação específica sobre Ontologias e redes semânticas em *corpora* de Ciências de Saúde da *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde* em <http://www.reciis.icict.fiocruz.br/index.php/reciis/issue/view/65>

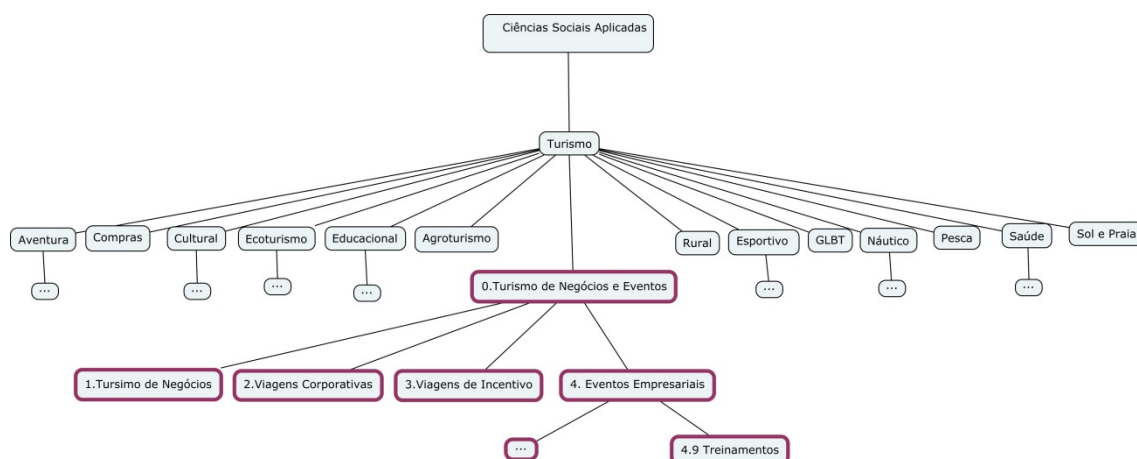


Figura 1 – Árvore de Domínio do Turismo de base motivacional.

Fonte: Elaboração própria. Autora: Danila Alves Carvalho

3.2 Compilação do *Corpus* de Estudo

A compilação do *corpus* de qualquer trabalho terminológico é um processo que deve ser sistematizado e criterioso.

Como bem apontado por Oliveira e Muller (2013, p.54)

Para fins terminológicos, o primeiro critério a ser seguido para a construção dos *corpora* de especialidade é garantir que os textos sejam completos, para que não se perca toda a riqueza de termos e de conhecimentos inerentes ao seu conteúdo.

Para a construção deste vocabulário bilíngue compilamos *corpora* comparáveis (pois os textos são similares em duas línguas) em português e inglês, a partir de sites de agências de turismo, sites diversos que tratam do tema e também sites governamentais, além de artigos, teses e dissertações que tratam do Turismo de Negócios e Eventos.

Esta compilação do *corpus* em português obedeceu à seguinte sequência:

- Busca online a partir das subáreas ligadas ao Turismo de Negócios e Eventos.
- Verificação de cada link exibido na página para compor o *corpus* em construção.
- Se no *link* encontramos informações importantes (que tratavam especificamente da área como números e descrição do setor, guias, etc.) para compor o *corpus*, copiamos o arquivo na ferramenta Bloco de Notas (formato txt.) e o salvamos com a codificação *Unicode* (melhor leitura pelo programa *WordSmith Tools*) na pasta à qual pertence.

Para a busca de material que compôs o *corpus* de caráter acadêmico, foi realizada uma pesquisa mais avançada a partir do próprio Google. Então, após escolhido o termo da Árvore que ajudará na busca por dados que vão compor o *corpus*, à direita podemos

selecionar a opção Pesquisa avançada como mostra a Figura 2:

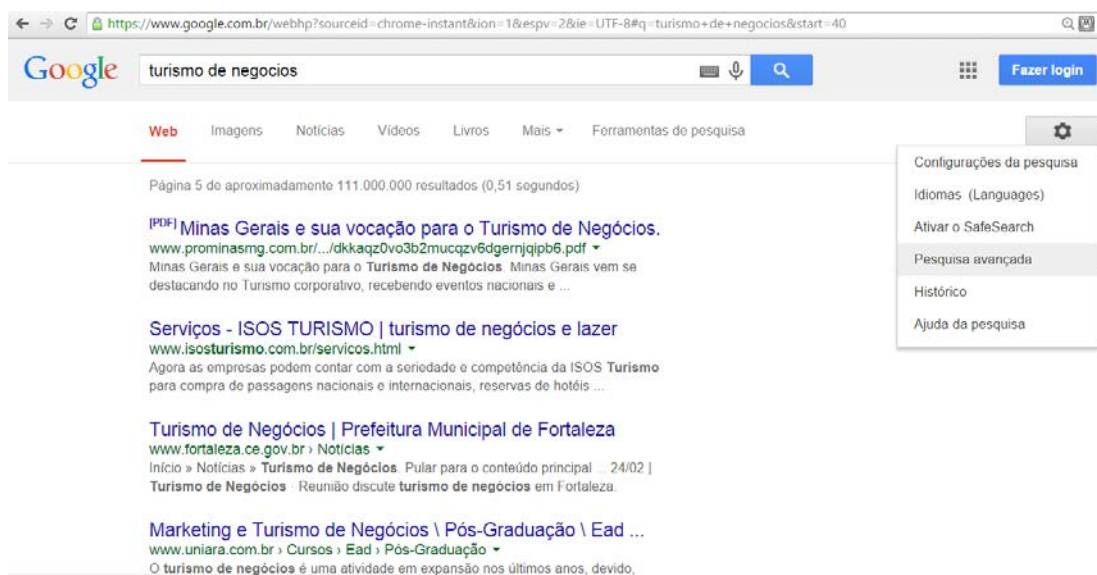


Figura 2 – Primeiro passo para uma pesquisa avançada no Google.com.br

Fonte: <https://www.google.com.br/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=turismo+de+negocios&start=40> Acesso em: 02 Feb 2015

Depois de acessarmos a página, outros campos devem ser preenchidos (“esta expressão ou frase exata”; idioma; região e tipo de arquivo – pdf), como mostra a Figura 3:

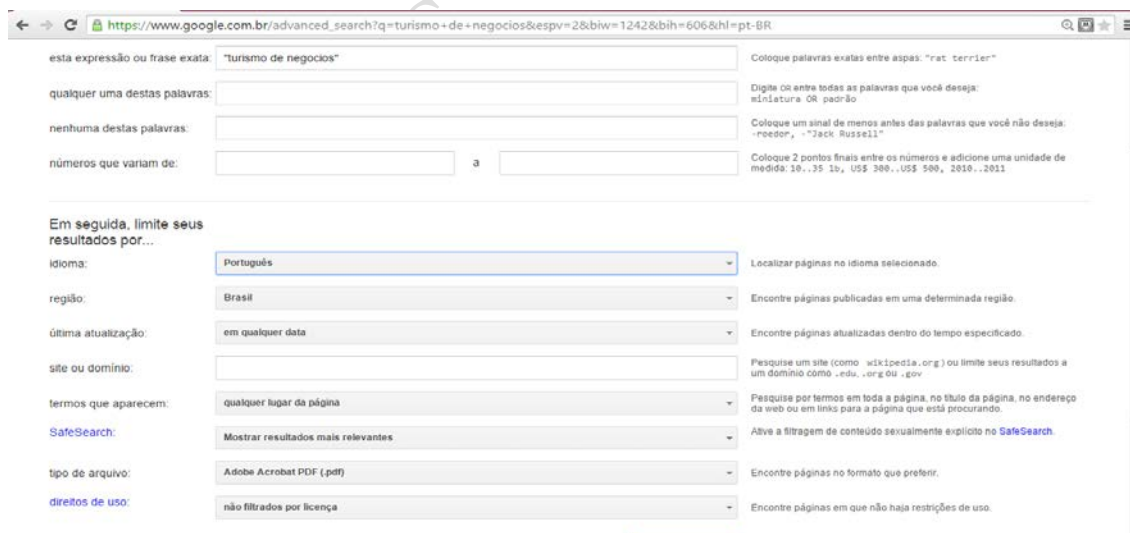


Figura 3 – Outros campos a serem preenchidos para a busca avançada

Fonte: https://www.google.com.br/advanced_search?q=turismo+de+negocios&espv=2&biw=1242&bih=606&hl=pt-BR Acesso em: 02 Feb 2015

Para a coleta e formação do *corpus* em inglês os mesmos procedimentos foram realizados.

3.3 A Organização do *Corpus* de Estudo

BERBER SARDINHA (2004, p.72) propõe que:

Uma vez que os textos tenham sido coletados e limpos, a tarefa seguinte é a organização dos arquivos em uma estrutura coerente. Não há regras gerais para isso. Alguns corpora vêm organizados em pastas simples, outros em pastas hierarquizadas (subpastas), outros em textos salvos em arquivos separados, outros em arquivos que contêm mais de um texto.

Realizamos a organização dos *corpora*, primeiramente, a partir de uma divisão simples em pastas: *Corpus* Inglês e *Corpus* Português (Figura 4):

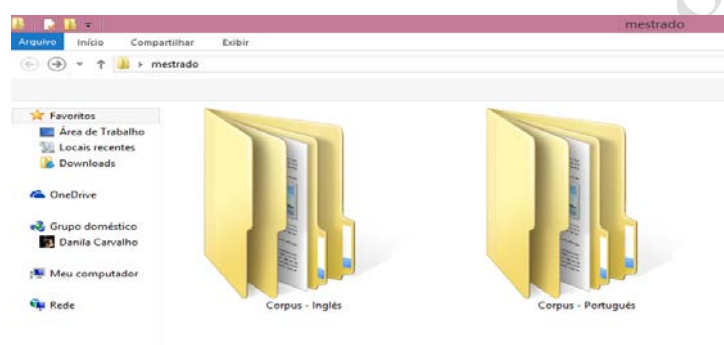


Figura 4– Organização dos *corpora* Turismo de Negócios e Eventos
Fonte: Elaboração própria.

Cada uma destas pastas possui subpastas que estão relacionadas às subáreas do Turismo de Negócios e Eventos. Cada subárea apresenta outras duas subpastas: Acadêmico e *Sites* Diversos (Figura 5):

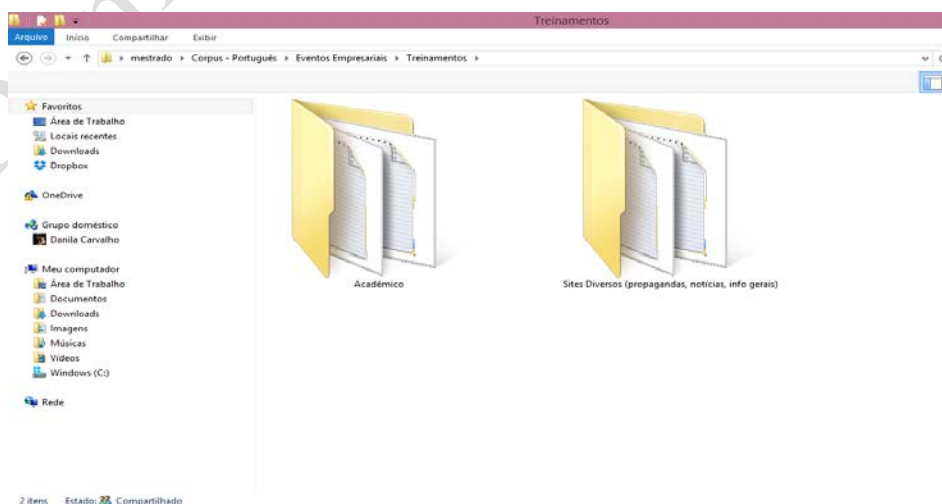


Figura 5 – Organização dos *corpora* dentro da subárea “Treinamentos”
Fonte: Elaboração própria.

Salvamos os arquivos nas pastas às quais pertencem de acordo com os nomes dos títulos principais, os registros de fonte e as datas, que foram colocados entre colchetes (<>).

Os muitos *links* ativados para a observação e composição dos *corpora* renderam os seguintes valores numéricos: 409 textos coletados em português e 441 textos coletados em inglês (que possuía mais textos acadêmicos de algumas subáreas e por isso a diferença em relação à tentativa de balanceamento quando comparado ao português), num total de 850 textos.

3.4 A Análise do *Corpus* de Estudo: Ferramentas utilizadas

Após a compilação do *corpus*, os dados são tratados quantitativamente e qualitativamente. Para iniciar o tratamento quantitativo dos dados, utilizamos a suíte *Wordsmith Tools* (SCOTT, 2011). As ferramentas do *Wordsmith Tools* contribuíram para este trabalho via metodologia da Linguística de *Corpus*, visto que esse programa, em sua arquitetura e recursos, possibilita que sejam seguidos os passos metodológicos básicos da LC. Seus principais recursos estão a seguir descritos.

3.4.1 *Wordlist*

O recurso *Wordlist do Wordsmith Tools* cria uma listagem geral de palavras de um *corpus*, de um texto ou de vários textos e informa o seu número de repetições. Esse é nosso primeiro contato com dados numéricos mais precisos do *corpus*. Para Viana (2010, p. 47):

Uma lista de frequência pode representar o primeiro contato do pesquisador com o seu *corpus* de estudo, já que ela permite a visualização das palavras mais e menos recorrentes. Esse procedimento revela inúmeros aspectos sobre o material incluído no *corpus* e igualmente gera hipóteses investigativas, que precisarão ser complementadas com análises mais detalhadas.

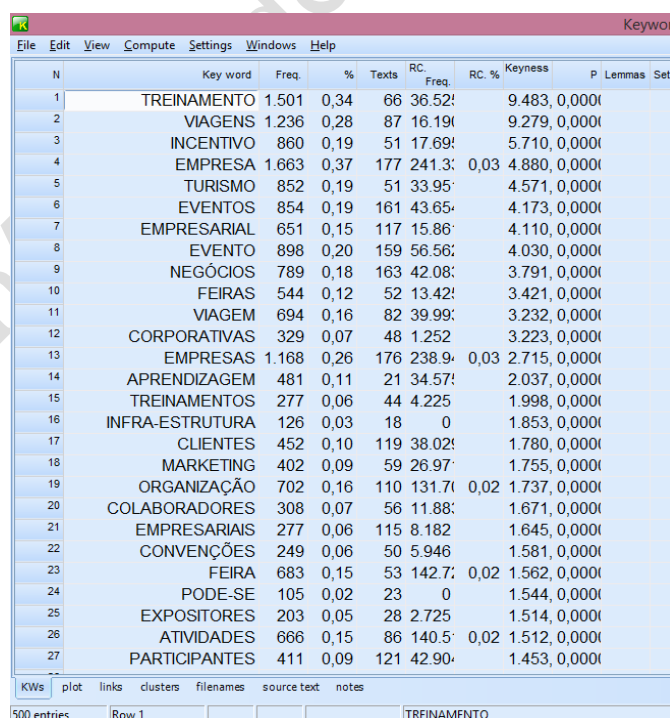
Para este trabalho, nosso *corpus* total apresenta número total de 1.586.426 *tokens*² nas duas línguas (português e inglês); sendo 711.703 *tokens* na língua portuguesa (base para levantamento dos candidatos a termo) e 874. 723 *tokens* língua inglesa.

² *Tokens*: palavras corridas dos textos.

3.4.2 Keywords

O recurso *Keywords* permite reconhecer as palavras-chave de um texto/*corpus*, sendo essas palavras específicas de um dado conjunto. Ao buscarem-se palavras ou expressões típicas ou específicas de um dado *corpus*, abre-se, em tese, um boa via para a identificação de que itens que são candidatos a termos. A especificidade é de sua ocorrência em um dados *corpus* – o *corpus* de estudos - comparativamente a outros conjuntos tomados como *corpora* de referência. No nosso caso, foram tomadas, como comparação de referência, as listas de palavras do COCA³ (Inglês) e do Banco do Português⁴ (Português).

Assim, em síntese, como ilustra a Figura 6 a seguir: “O extrator de palavras-chave permite o levantamento de índices linguísticos que caracterizam um *corpus* de estudo quando esse é comparado a outro de referência”. (VIANA, 2010, p.61). Como se pode observar, na busca por prováveis termos em meio a essas *Keywords*, itens gramaticais específicos ou verbos são desprezados, utilizando-se o recurso de filtragem denominado *Stopwords*.



N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	Lemmas	Set
1	TREINAMENTO	1.501	0,34	66	36.52		9.483, 0,000			
2	VIAGENS	1.236	0,28	87	16.19		9.279, 0,000			
3	INCENTIVO	860	0,19	51	17.69		5.710, 0,000			
4	EMPRESA	1.663	0,37	177	241.3	0,03	4.880, 0,000			
5	TURISMO	852	0,19	51	33.95		4.571, 0,000			
6	EVENTOS	854	0,19	161	43.65		4.173, 0,000			
7	EMPRESARIAL	651	0,15	117	15.86		4.110, 0,000			
8	EVENO	898	0,20	159	56.56		4.030, 0,000			
9	NEGÓCIOS	789	0,18	163	42.08		3.791, 0,000			
10	FEIRAS	544	0,12	52	13.42		3.421, 0,000			
11	VIAGEM	694	0,16	82	39.99		3.232, 0,000			
12	CORPORATIVAS	329	0,07	48	1.252		3.223, 0,000			
13	EMPRESAS	1.168	0,26	176	238.9	0,03	2.715, 0,000			
14	APRENDIZAGEM	481	0,11	21	34.57		2.037, 0,000			
15	TREINAMENTOS	277	0,06	44	4.225		1.998, 0,000			
16	INFRA-ESTRUTURA	126	0,03	18	0		1.853, 0,000			
17	CLIENTES	452	0,10	119	38.02		1.780, 0,000			
18	MARKETING	402	0,09	59	26.97		1.755, 0,000			
19	ORGANIZAÇÃO	702	0,16	110	131.7	0,02	1.737, 0,000			
20	COLABORADORES	308	0,07	56	11.88		1.671, 0,000			
21	EMPRESARIAIS	277	0,06	115	8.182		1.645, 0,000			
22	CONVENÇÕES	249	0,06	50	5.946		1.581, 0,000			
23	FEIRA	683	0,15	53	142.7	0,02	1.562, 0,000			
24	PODE-SE	105	0,02	23	0		1.544, 0,000			
25	EXPOSITORES	203	0,05	28	2.725		1.514, 0,000			
26	ATIVIDADES	666	0,15	86	140.5	0,02	1.512, 0,000			
27	PARTICIPANTES	411	0,09	121	42.90		1.453, 0,000			

Figura 6 - Exemplo de tela do *Keywords* do Corpus Turismo de Negócios

Fonte: *WordSmith Tools*

³ COCA: *Corpus of Contemporary American English*; o maior *corpus* monitor (sofre atualizações) do inglês, criado por Mark Davis.

⁴ Banco do Português: *corpus* monitor (sofre atualizações) do português do Brasil, criado por Tony Berber Sardinha.

Então, após eleitos os melhores candidatos a termo nas duas línguas (português e inglês) a partir do *Keywords*, partimos para a análise das linhas de concordância. Essas *concordâncias*, geradas pelo recurso *Concord*, são as listagens de contextos de cada item - que nos ajudam na identificação de segmentos úteis para a criação da definição terminológica.

3.4.3 Concord

Depois de escolhidos, os candidatos a termos são definidos a partir das informações potencialmente explicativas ou definitórias depreendidas das linhas de concordância (Figura 7) disponibilizadas no *Concord*. Assim, diferentes frases com predicções sobre um dado item/termo auxiliam para a composição da porção definitiva da microestrutura do verbete do dicionário. Lembramos que, conforme Biderman (2006), “As concordâncias de texto permitiriam identificar os significados (conceitos) e os usos dos vocábulos selecionados para a nomenclatura do dicionário, que é a espinha dorsal do dicionário”.

N	Concordance	Set	Tag	Word #	Sen	Sen	Parz	Parz	Hea	Sec	Sec	File	Date	%
27	Corporativos: Netas Adventure (Treinamento corporativo) Treinamento	117	3	28	0	72	0	72	0	72	0	Treinamentos	??	64%
28	conceitos que podem nos ajudar • Treinamento - é o processo intencional	125	8	25	0	13	0	13	0	13	0	O que é trein	??	14%
29	exclusivo em Brotas, a PS Alaya Treinamento Empresarial também	930	29	37	0	92	0	92	0	92	0	Treinamento	??	91%
30	chamados, por usuário, que fez algum treinamento 1,01 0,91 -11% 4-Número	30.79	1.4	58	0	43	0	43	0	43	0	CONTRIBUIÇ	??	43%
31	e outros custos relacionados ao treinamento, garantiu um rápido	1.857	12	70	0	3%	0	3%	0	3%	0	CONTRIBUIÇ	??	3%
32	os diversos aspectos relacionados ao treinamento, como local, alimentação,	1.254	49	78	0	74	0	74	0	74	0	Como desen	??	73%
33	processos de negócio relacionados ao treinamento. É importante conhecer os	43.051	2.1	10	0	60	0	60	0	60	0	CONTRIBUIÇ	??	61%
34	não só os diretamente relacionados ao treinamento em si, para que não	2.255	13	74	0	3%	0	3%	0	3%	0	CONTRIBUIÇ	??	4%
35	conceitos de educação a distancia ao treinamento: Um estudo de caso em	10.56	46	53	0	94	0	94	0	94	0	IMPORTANC	??	95%
36	ou diretamente relacionados ao treinamento e, portanto, a	70.04	3.4	85	0	98	0	98	0	98	0	CONTRIBUIÇ	??	98%
37	o grau de importância que você dá ao treinamento no que diz respeito à	11.151	51	64	0	99	0	99	0	99	0	IMPORTANC	??	99%
38	o grau de importância que você dá ao treinamento no que diz respeito à	10.881	50	64	0	96	0	96	0	96	0	IMPORTANC	??	97%
39	de empreendimento voltado ao treinamento e à capacitação de	258	6	64	0	16	0	16	0	16	0	TREINAMEN	??	16%
40	que os colaboradores dar ao treinamento organizacional.	1.727	11	94	0	15	0	15	0	15	0	IMPORTANC	??	19%
41	do cargo. Isso se aplica tanto ao treinamento inicial de um colaborador	1.830	11	37	0	16	0	16	0	16	0	IMPORTANC	??	20%
42	esforço terá sido em vão. O acesso ao treinamento deve ser confiável, não	20.49	1.1	23	0	29	0	29	0	29	0	CONTRIBUIÇ	??	29%
43	que os colaboradores deram ao treinamento. 46 Gráfico 07 - Grau de	8.633	34	10	0	77	0	77	0	77	0	IMPORTANC	??	79%
44	para quebrar a resistência ao treinamento on-line. Se o material de	62.87	3.1	97	0	88	0	88	0	88	0	CONTRIBUIÇ	??	88%
45	seja aprovado, ele pode dar início ao treinamento on-line, para depois fazer	62.011	3.1	54	0	87	0	87	0	87	0	CONTRIBUIÇ	??	87%
46	quebradas e deu-se um novo fôlego ao treinamento que utiliza recursos	18.691	94	85	0	26	0	26	0	26	0	CONTRIBUIÇ	??	26%
47	O usual é estarem habituadas ao treinamento presencial, com um	37.481	1.1	36	0	53	0	53	0	53	0	CONTRIBUIÇ	??	51%
48	e respectivos planos de aplicação. Treinamento Vivencial com refeição	636	20	5%	0	63	0	63	0	63	0	Treinamento	??	62%
49	trabalho, verificar quais aplicam treinamento com seus colaboradores,	736	22	30	0	34	0	34	0	34	0	A IMPORTAN	??	33%
50	os custos estimados para a aquisição, treinamento, implementação (quando	49.90	2.4	23	0	70	0	70	0	70	0	CONTRIBUIÇ	??	70%
51	. 1. Breve Relato da Atividade TREINAMENTO EMPRESARIAL	491	17	58	0	30	0	30	0	30	0	TREINAMEN	??	30%
52	Vivencial com atividades de aventura treinamento vivencial com atividades	148	5	20	0	15	0	15	0	15	0	Treinamento	??	15%
53	, agrônomos, advogados, etc.) b) Treinamento dos atendentes Após a	5.290	19	65	0	67	0	67	0	67	0	F undaçã	??	60%

Figura 7 - Exemplo de tela do *Concord* a partir do candidato a termo treinamento

Fonte: *WordSmith Tools*

A linha de concordância selecionada leva o pesquisador a uma análise mais

detalhada do contexto (Figura 8), o que vai além do cotexto ⁵ da informação:

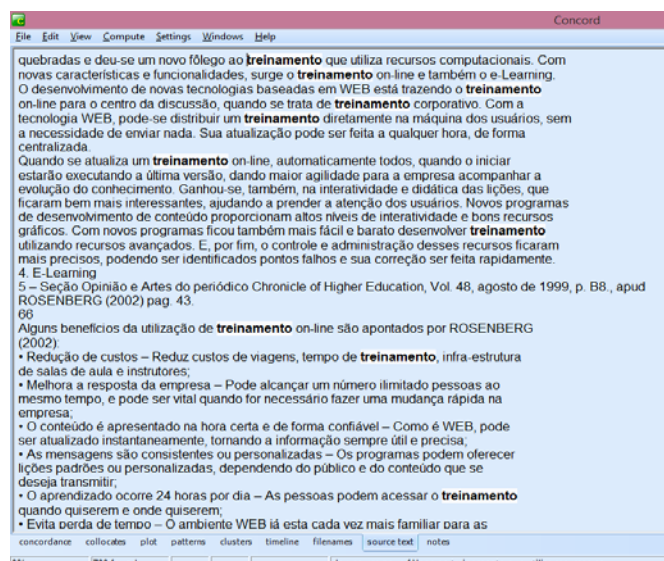


Figura 8 - Exemplo de tela de tela do *Concord* – Análise de linhas de concordância do candidato a termo “treinamento”

Fonte: *WordSmith Tools*

3.5 A Ficha Terminológica e a Criação de definições

Para chegarmos à definição final de um verbete, as fichas terminológicas são necessárias; são elas que organizam a informação coletada e como ela será trabalhada para a composição da definição de um verbete.

Duas etapas do processo terminográfico serão aqui discutidas: a definição e a ficha terminológica. Para Krieger e Finatto (2004, p. 92)

Há vários tipos de enunciados definitórios. A definição terminológica (doravante DT) é reconhecida como aquela que mais se ocupa de termos técnico-científicos. A definição lexicográfica, por sua vez, é conhecida como aquela que mais se ocupa das palavras. A definição lógica, de um outro modo, estabelece um valor proposicional de verdade, enquanto as definições explicativas ou enciclopédicas contêm informações variadas sobre um dado objeto da realidade.

Por tratar-se de um vocabulário especializado, a definição utilizada neste trabalho é a DT. Esta definição, para melhor atender o usuário, deve ser construída com base em duas características fundamentais: o gênero próximo e a diferença específica. Ainda para

⁵LINGUÍSTICA: conjunto de sequências linguísticas que precedem ou que se seguem a uma palavra ou um enunciado na linearidade textual. Fonte: <http://www.infopedia.pt/dicionarios/lingua-portuguesa/cotexto?homografia=0> Acesso em: 20 Maio 2015.

Krieger e Finatto (2004, p. 93)

Gênero próximo é a porção da definição que expressa a categoria ou classe geral a que pertence o ente definido. A diferença específica é a indicação da (s) particularidade (s) que distingue (em) esse ente em relação a outros da mesma classe.

Assim, partimos da ideia de hiperônimo (mais abrangente) para o hipônimo (mais específico) para criar a definição do termo. Então, a GPDE (Gênero Próximo Definição Específica), também chamada de definição aristotélica (ou clássica), é proposta pelo VoTec; como o VocTur é um trabalho que foi criado dentro da ferramenta VoTec, a definição aristotélica é a que está presente neste trabalho.

As Figuras 9, 10 e 11 ilustram partes das fichas terminológicas dispostas no Votec – Vocabulário Técnico Online (FROMM, 2007) e como é realizado o registro de dados para a posterior criação de definições. A Figura 9 mostra o cadastro dos dados de um conceito, exemplos e fonte, além da data de coleta:

Vocabulário Técnico Online Tela Cheia | English

Cadastro de Contextos para Treinamento

Novo Contexto

Passo 2

Dados do Contexto

Exemplo*:

Conceito*:

Fonte:

Data de Coleta*: (dia/mês/ano ex.: 18/03/2007)

Contextos Cadastrados

Exemplo	Conceito	Fonte	Ações

Figura 9 – Tela Cadastro de Contextos

Fonte: <http://www.pos.voteconline.com.br/>. Acesso restrito.

Após registrarmos os contextos, os dados que devem compor a microestrutura do verbete para posterior criação de uma definição são explorados. Preenchemos campos relacionados aos traços distintivos, semântica, termo equivalente, termos remissivos e informações enciclopédicas, como mostra a Figura 10:



Termo: Treinamento

Passo Anterior Salvar Sair sem salvar

Contextos **Conceito Final / Definição**

Contextos

Exemplo	Conceito	Fonte
1 o treinamento é visto como uma forma de treinar e adaptar o colaborador na função que exerce dentro da organização, não apenas em sua admissão, mas também no decorrer de sua função	forma de treinar e adaptar o colaborador na função que exerce dentro da organização	PDF 08/12/-1
2 o treinamento é uma maneira eficaz de delegar valor às pessoas, à organização e aos clientes	maneira eficaz de delegar valor às pessoas, à organização e aos clientes	PDF 08/12/-1
3 A função do treinamento é desenvolver o potencial de cada empregado para que eles	tem a função de desenvolver o potencial de cada empregado para que eles tenham	PDF 08/12/-1

Dados Traços Distintivos Semântica Termo Equivalente Termos Remissivos Informações Enciclopédicas

Dados

Ontologia: Turismo > Turismo de Negócios e Eventos > Eventos Empresariais > Treinamentos

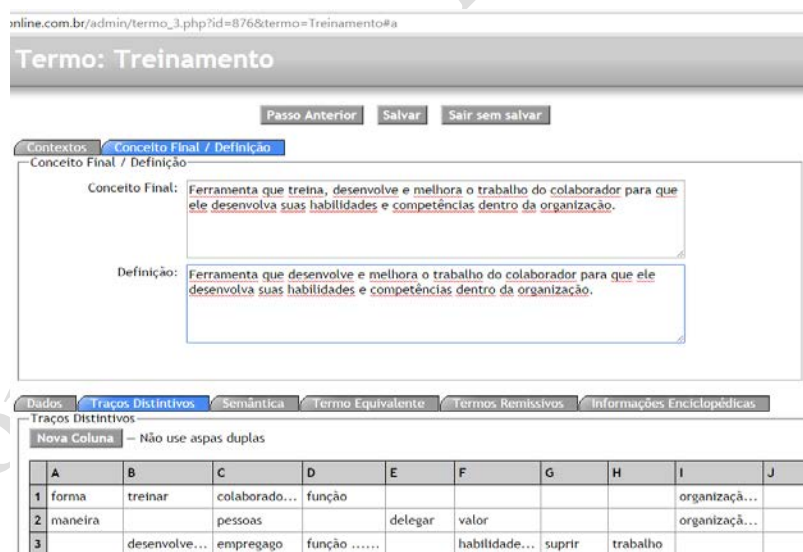
Categoria Gramatical: Escolha uma Categoria Número: Singular

Gênero: Masculino Sigla/Acrônimo:

Figura 10 – Tela que apresenta as abas que serão exploradas para a criação dos conceitos

Fonte: <http://www.pos.voteonline.com.br/>

Após o preenchimento dos paradigmas que formam a definição da microestrutura do verbete, conceito final e definição são criados pelo autor. Vejamos a Figura 11:



online.com.br/admin/termo_3.php?id=876&termo=Treinamento#a

Termo: Treinamento

Passo Anterior Salvar Sair sem salvar

Contextos **Conceito Final / Definição**

Conceito Final / Definição

Conceito Final: Ferramenta que treina, desenvolve e melhora o trabalho do colaborador para que ele desenvolva suas habilidades e competências dentro da organização.

Definição: Ferramenta que desenvolve e melhora o trabalho do colaborador para que ele desenvolva suas habilidades e competências dentro da organização.

Dados Traços Distintivos Semântica Termo Equivalente Termos Remissivos Informações Enciclopédicas

Traços Distintivos

Nova Coluna — Não use aspas duplas

	A	B	C	D	E	F	G	H	I	J
1	forma	treinar	colaborado...	função					organizaçã...	
2		maneira	pessoas	função	delegar	valor			organizaçã...	
3		desenvolve...	empregado	função		habilidade...	suprir	trabalho		

Figura 11 – Tela que apresenta as abas que serão exploradas para a criação dos conceitos – destaque para a tela de Conceito Final e Definição

Fonte: <http://www.pos.voteonline.com.br/>

Esta definição deve ser concisa. Todas as informações necessárias para a construção do conceito final são inseridos em uma oração que define o conceito.

4. Resultados

Abaixo (Figura 12) observamos a tela de um verbete já pronto e a maneira que ele é visualizado no VoTec. Esta disposição da informação permite que os consulentes (neste caso estudantes dos cursos de Tradução e Turismo) possam explorar os termos que buscam de várias maneiras.

Observamos que o tipo de exibição selecionado é o Normal e o tipo de consulta selecionada é a Total, onde visualizamos todos os campos disponíveis no banco de dados para que compreendamos o texto.

The screenshot shows the VoTec online dictionary interface. The search results are displayed in two columns. The left column contains navigation options: 'Tipos de Exibição' (Normal, Descritiva), 'Tipos de Consulta' (Total, Tradutor, Modular), and 'Consultas Externas' (Corpus NILC, Google, Answers.com, Wikipedia, CORTEC). The right column shows the search results for 'Treinamento' in Portuguese and English. The Portuguese section includes a definition: 'Treinamento. *Treinamentos*. s.m.s. ferramenta que desenvolve e melhora o trabalho do colaborador para que ele desenvolva suas habilidades e competências dentro da organização.. Ex.: o treinamento é visto como uma forma de treinar e adaptar o colaborador na função que exerce dentro da organização, não apenas em sua admissão, mas também no decorrer de sua função. *Hipônimo de:* Ferramenta. *Hiperônimo de:* Forma. *Corpus: Posição na Ordem de Frequência:* (5); *Nº de Ocorrências do termo:* (1523). *Informações Enciclopédicas:* A expressão *treino* (português europeu) ou *treinamento* (português brasileiro) refere-se à aquisição de conhecimento, habilidades e competências como resultado de formação profissional ou do ensino de habilidades práticas relacionadas à competências úteis e Em: *Treinamento* - [Wikipedia](#). The English section includes a definition: 'Training. *Training*. n.m/f.s. a program that ensures that the organization and its employees can improve and succeed by learning. Ex.: While skills gaps (we call it the "supply chain of skills") continue to challenge companies, an increased investment in training is good for everyone: employees, businesses, and job seekers.. *Hyponym of:* Program. *Corpus: Frequency order position:* (45); *Term number of occurrences:* (1222). *Encyclopedic Information:* Training is teaching, or developing in oneself or others, any skills and knowledge that relate to specific useful competencies. Training has specific goals of improving one's capability, capacity, productivity and performance. It forms the core of apparent em: *Training* - [Wikipedia](#).

Figura 12 – Tela que apresenta o termo treinamento com o tipo de exibição Normal e Tipo de Consulta Total

Fonte: <http://www.pos.voteconline.com.br/>

5. Considerações Finais

Este artigo teve o objetivo de expor os passos que foram fundamentais rumo à proposta de construção de um vocabulário bilíngue português/inglês para o Turismo de Negócios e Eventos, ao esclarecer, primeiramente sobre a importância das reflexões teóricas relacionadas à Terminologia, Terminografia e Linguística de *Corpus*. Posteriormente, caracterizamos as etapas metodológicas para a criação do vocabulário: a criação da árvore de domínio como forma de compreensão abrangente da área a ser trabalhada e estrutura que organiza a informação coletada (os termos dentro das áreas); coleta e tratamento dos dados com base na abordagem e metodologia da Linguística de *Corpus*, além do preenchimento das fichas terminológicas utilizadas para a criação de definições, fichas estas disponibilizadas pela plataforma Votec.

Alertamos, entretanto, que o processo de coleta e de identificação dos dados realmente úteis para o nosso usuário do VocTur não é tão simples. Já no primeiro passo, o da elaboração da Árvore de Domínio, observamos dificuldades na sua construção, pois

trata-se de uma estrutura conceitual complexa e para a qual convergem elementos de outras subáreas do Turismo. Outro aspecto importante é a coleta criteriosa dos textos que compõem os *corpora* Turismo de Negócios e Eventos, pois há chamadas para o tema, mas muitos dos textos, quando acessados, não apresentam informações relevantes ou significativas para a composição dos dados da coleta.

O trabalho apontou o caminho percorrido rumo à busca por informações que objetivam compor uma proposta de vocabulário bilíngue português/inglês para o Turismo de Negócios e Eventos direcionada pelo *corpus*. A clareza desta trajetória e as bases teóricas sugeridas como fundamentais possibilitam que esta sequência seja, inclusive, utilizada como um guia para a produção de outros trabalhos terminológicos. Também enfatizamos a importância das descobertas e do aprendizado adquirido quando um trabalho é guiado pelo *corpus*, visto que essa orientação nos mostra caminhos e também nos surpreende ao revelar termos, números e dados. Outro ponto positivo do relato de pesquisa posto neste artigo está no fato de reconhecer que há contribuição científica que se converte em benefício social compartilhado à medida que outras pessoas possam beneficiar-se deste percurso, replicando escolhas e procedimentos que adotamos.

Referências

ALMEIDA, G. M. B. **A teoria Comunicativa da Terminologia e a sua prática**. Alfa, São Paulo, v.50, n.2, p.85-101, 2006.

BARBOSA, M. A.. **Lexicologia, lexicografia, terminologia, terminografia, identidade científica, objeto, métodos, campo de atuação**. Disponível em http://www.ufrgs.br/riterm/por/txt_simposios_anteriores_1990.html Acesso em: 02 Fev.2015

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri S.P.: Editora Manole, 2004, 410 p.

BEVILACQUA, C. R. Por que e para que a Linguística de *Corpus* na Terminologia. In: **Corpora na Terminologia**. TAGNIN, S.; BEVILACQUA, C.. (orgs). São Paulo: Hub Editorial, 2013. p.11-27

BIDERMAN, M. T. C. O conhecimento, a terminologia e o dicionário. **Ciência e Cultura**, São Paulo, v. 58, n. 2, Junho, 2006. Disponível em: http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S00096725200600020014&lng=en&nrm=iso Acesso em: 10 Mai 2014

CABRÉ, M. T. **La terminología hoy: concepciones, tendencias y aplicaciones**. Ciência da Informação, Brasília, DF, Brasil, v.24,n 3 dez. 1995.

FROMM, G. **VoTec**: a construção de vocabulários eletrônicos para aprendizes de tradução. 2008. Tese (Doutorado em Estudos Lingüísticos e Literários em Inglês) - Faculdade de Filosofia, Letras e Ciências Humanas, University of São Paulo, São Paulo, 2008. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008-150855> Acesso em: 21 Agosto 2012

KRIEGER, M. da G. Do reconhecimento de terminologias: entre o linguístico e o textual. In: **As Ciências do Léxico**: lexicologia, lexicografia, terminologia. ISQUERDO, A. N.; KRIEGER, M. da G. (orgs). Campo Grande, MS: Editora UFMS, 2004. 381 p.

KRIEGER, M. da G.; FINATTO, M. J. B.. **Introdução à Terminologia**: teoria e prática. São Paulo: Contexto, 2004. 223 p.

OLIVEIRA, L. H. M. de; MULLER, A. F. A Terminologia e a utilização de ferramentas computacionais de análise de corpus. In: **Corpora na Terminologia**. TAGNIN, S.; BEVILACQUA, C. (orgs). São Paulo: Hub Editorial, 2013. p. 47-61.

PORTAL BRASIL. Disponível em:
<http://www.brasil.gov.br/turismo/2013/11/turismo-de-negocios-movimenta-setor-no-pais> Acesso em 24 Fev. 2014

SCOTT, M., 2011, **WordSmith Tools** version 6, Liverpool: Lexical Analysis Software.

TAGNIN, S. E. O. Glossário de Linguística de Corpus. In: VIANNA, V.; TAGNIN, S. O. E. **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010. 375p.

UBERLÂNDIA CONVENTION & VISITORS BUREAU (UC&VB). Disponível em:
<http://uberlandiacvb.com.br/portal/sobre-uberlandia/#> Acessado em: 25 Fev. 2015

VIANA, V. Linguística de Corpus: conceitos, técnicas e análises. In: **Corpora no ensino de línguas estrangeiras**. VIANA, V.; TAGNIN, S. E. O. São Paulo: Hub Editorial, 2010. 375p.

Artigo recebido em: 01.03.2015

Artigo aprovado em: 02.08.2015