

MÁRCIO ISSAMU YAMAMOTO

LINGUÍSTICA HISTÓRICA E LINGUÍSTICA DE *CORPUS*-
CAMINHOS QUE SE CRUZAM PARA DESVELAR A HISTÓRIA DA
LINGUAGEM:
um vocabulário bilíngue português - inglês

UBERLÂNDIA-MG

2015

MÁRCIO ISSAMU YAMAMOTO

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Letras e Linguística da Universidade Federal de Uberlândia, como requisito parcial para obtenção de Título de Mestre em Estudos Linguísticos.

Área de concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: (i) Teoria, descrição e análise linguística

Orientador: Prof. Dr. Guilherme Fromm

UBERLÂNDIA-MG

2015

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

- Y19L
2015 Yamamoto, Márcio Issamu, 1974-
Linguística Histórica e Linguística de Corpus - caminhos que se cruzam para desvelar a história da linguagem: um vocabulário bilíngue português-inglês / Márcio Issamu Yamamoto. - 2015.
116 f.
- Orientador: Guilherme Fromm.
Dissertação (mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Estudos Linguísticos.
Inclui bibliografia.
1. Linguística - Teses. 2. Linguística histórica - Teses. 3. Linguística de corpus - Teses. 4. Vocabulário - Teses. I. Fromm, Guilherme. II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Estudos Linguísticos. III. Título.

LINGUÍSTICA HISTÓRICA E LINGUÍSTICA DE *CORPUS*-
CAMINHOS QUE SE CRUZAM PARA DESVELAR A HISTÓRIA DA
LINGUAGEM:
um vocabulário bilíngue português - inglês

Dissertação aprovada para a obtenção do
título de Mestre no Programa de Pós-
Graduação em Estudos Linguísticos da
Universidade Federal de Uberlândia
(MG) pela banca examinadora formada
por:

Uberlândia, 13 de julho de 2015.

Prof. Dr. Guilherme Fromm, UFU/MG

Prof^a Dr^a Maria José Bocorny Finatto, UFRGS/RS

Prof. Dr. Ariel Novodvorski, UFU/MG

Dedico este trabalho ao Oráculo, quem faz acreditar
que sonhos podem se tornar realidade
e esteve sempre presente.

A minha esposa e filhos, Wânia, Marcel e Emily.

Aos meus pais e irmãos Samuel (*in memoriam*) e

Laudice, Mércia e Kenji.

Aos primos Ismael Cordeiro Sobral Filho, Mônica,

Emmanuelle e Ana Clara.

AGRADECIMENTOS

Não se trilha um caminho sem que haja quem vá adiante. Agradeço imensamente ao professor Guilherme Fromm, orientador extraordinário que me incentivou, instruiu, direcionou, apoiou e sempre me surpreendeu com sua sabedoria, conhecimento, entendimento e humanidade ímpares.

Agradeço ao Programa de Pós-Graduação em Estudos Linguísticos – PPGEL - pela acolhida durante estes anos de Mestrado, como aluno especial e regular. À professora Dilma Mello que em sua simplicidade motivou-me. À Maria Virgínia e Lorena, colegas sempre presentes.

Sou grato às professoras Ana Maria Donnard, Zeina Khoury e Benice Naves que me apresentaram o maravilhoso mundo da Lexicografia e Terminologia bilíngues.

Agradeço a minha família que sempre abriu mão da minha presença e dedicação para que pudesse fazer as aulas, as pesquisas, apresentar em congressos e tudo mais. Sou imensamente grato a vocês: Wânia, Marcel e Emily.

Agradeço aos colegas da UFG/Jataí que, incondicionalmente, me apoiaram nesta caminhada, ministrando aulas e cobrindo meus horários para que eu pudesse fazer as aulas e concluir o trabalho final. Meu muito obrigado, do fundo do coração, para o Fabiano Ramos, a Natasha Costa e a Neuda Lago.

Agradeço o apoio da família, nas pessoas dos tios Kiyoshi, Mitsue, Uruo e Hissae; os primos Rute, Beto, Ismael e Mônica, Bete e Yuji, à minha mãe e irmãos. Sem a ajuda de vocês os congressos teriam sido diferentes, sem uma hospedagem calorosa como a que vocês me ofereceram, além do amor e carinho incondicionais.

Um agradecimento especial a Samuel Yamamoto, meu pai, *in memoriam*, o responsável por plantar a semente do amor às Letras, Língua e Linguagem, por meio de seus estudos em Hebraico, Grego e Português, Lexicologia e Terminologia.

Não há como deixar de agradecer os colegas de Graduação e Pós-Graduação, constantes nesse caminhar rumo ao conhecimento. Lucas (todos), Vanesssa, Raphael, Neubiana, Solange (uma mãe), Vilma, Valeska, Thiago e tantos outros que o espaço de uma folha não permite nomear.

Finalmente agradeço à Prof^a Maria José Bocorny Finatto, ao Prof. Ariel Novodvorsky, a Prof^a Eliana Dias que me acompanharam desde o início até o fim, apoiaram direta ou indiretamente meu trabalho e estudos, e se dispuseram a fazer parte da banca avaliadora. Admiro-os muito!

RESUMO

Esta dissertação trata da terminologia da área da Linguística Histórica, adotando a perspectiva teórico-metodológica da Linguística de *Corpus* (BERBER SARDINHA, 2004), e partindo de pressupostos teóricos da Teoria Comunicativa da Terminologia – TCT – (CABRÉ, 1999), do conceito de vocabulário (BARBOSA, 1990), do conceito de terminologia (KRIEGER; FINATTO, 2004) e dos contextos definitórios (AUBERT, 1996). O objetivo principal foi criar uma obra terminológica bilíngue, português-inglês, português do Brasil e de Portugal, cujas definições em contraste das unidades terminológicas foram elaboradas a partir de um banco de dados terminológico, na área de Linguística Histórica, cujo público alvo são os tradutores, estudantes e profissionais da área de Letras e Linguística. Os objetivos específicos da pesquisa foram: (I) elaborar uma árvore de domínio (BARROS, 2004) das áreas relacionadas à Linguística Histórica, (II) compilar *corpora* bilíngue, (III) levantar os candidatos a termos e buscar posterior confirmação com especialistas, (IV) disponibilizar dados oriundos de fontes confiáveis aos usuários, (V) ordenar os dados em fichas de registro, como entradas em dicionários, (VI) elaborar definições para os termos selecionados, e (VII) disponibilizar os resultados no VoTec <www.pos.voteconline.com.br>, plataforma de gerenciamento terminológico. Para fazer a análise dos *corpora*, o programa *WordSmith Tools 6.0* (WST, SCOTT, 2012), foi utilizado. Os resultados obtidos quanto à elaboração da Árvore de Domínio foi que a área da Linguística Histórica é difícil de ser definida como uma área estanque. Isto se deve ao fato de que ela se alimenta de outras subáreas da Linguística, principalmente como *corpus*, e de outras, como a Fonologia, para fazer sua análise de dados (MATTOS E SILVA, 2008). Isto é, tratar de Linguística Histórica significa envolver a Etimologia, a Filologia e a Linguística Diacrônica. O *corpus* do português compõe-se de teses, dissertações e artigos científicos. O de inglês é composto de anais de eventos, artigos, teses e dissertações. Quanto às Unidades Terminológicas, observamos que nos *corpora* de português e inglês houve unidades terminológicas correspondentes que servem de candidatos a termos. As definições foram redigidas usando-se o padrão GPDE, gênero próximo, diferença específica.

Palavras-chave: Linguística de *Corpus*; Linguística Histórica; Terminologia; Terminografia; Vocabulário bilíngue.

ABSTRACT

This dissertation addresses Historical Linguistics terminology, based on theoretical and methodological perspective of *Corpus* Linguistics (BERBER SARDINHA, 2004). It is also based on theoretical assumptions of the Communicative Theory of Terminology (CABRÉ, 1999), on the concept of Vocabulary (BARBOSA, 1990), the concept of Terminology (KRIEGER; FINATTO, 2004), and the defining contexts (AUBERT, 1996). The aims of this research are to: (I) build a bilingual terminological dictionary, English-Portuguese, comprising Brazilian and European Portuguese, which definitions, by contrast, will be elaborated from a Historical Linguistics terminological database whose users are translators and Language and Linguistics professionals and students; (II) build a Historical Linguistics domain tree (BARROS, 2004); (III) compile bilingual *corpora*; (IV) select the probable dictionary terms and, afterwards, seek for specialists' confirmation; (V) make trustworthy data available for dictionary users; (VI) Organize data by using register tabs as dictionary entries; (VII) build up definitions for the selected terms; (VIII) make the outcome available at VoTec, a terminological management platform. For data analysis, WordSmith Tools 6.0 (SCOTT, 2012) is used. As a research outcome, it was observed that defining Historical Linguistics, as a linguistic domain itself, was quite a challenging task. Historical Linguistics is a domain that profits from other subdomains, especially as *corpus*, and from others to make its data analysis. Therefore, when it comes to Historical Linguistics, it comprises data from Etymology, Philology and Diachronic Linguistics. The Portuguese *corpus* comprises thesis, dissertations and academic articles, while the English one comprises conference proceedings, academic articles, thesis, and dissertations. Considering the terms for including in the dictionary, we have observed that corresponding terms are found in both *corpora*, and their definitions were written using proximate genus and specific difference patterns.

Keywords: *Corpus* Linguistics; Historical Linguistics; Terminology; Terminography; Bilingual Vocabulary.

Lista de Figuras

Figura 1. Macro e microestrutura – termo: semiótica	15
Figura 2. Árvore do Campo da Linguística com as subáreas de estudo, reformulada a partir da árvore de 2013.	47
Figura 3. Microestrutura conforme proposta de Hartmann e James (2002).....	51
Figura 4. Recorte da árvore de domínio da Linguística	57
Figura 5. Árvore de domínio inicial - proposta para a LH.....	58
Figura 6. Visão parcial dos arquivos do <i>corpus</i> em português de LH.	60
Figura 7. Visão parcial dos arquivos de <i>corpus</i> em inglês de LH.....	61
Figura 8. Lista de palavras em ordem de frequência do <i>corpus</i> de LH – inglês (visão parcial).....	64
Figura 9 . Lista de palavras em ordem de frequência do <i>corpus</i> de LH – português (visão parcial).....	65
Figura 10. Procedimento de palavras-chave (VIANA, 2010, p.61).	67
Figura 11. Lista de palavras-chave do <i>corpus</i> de LH (visão parcial).....	68
Figura 12. Vista da seleção do colocado <i>é</i> do termo Linguística.	71
Figura 13. Linha de concordância para o termo <i>linguística</i> seguida do colocado <i>é</i> (vista parcial).....	72
Figura 14 . Exemplo do termo linguística , acompanhado do colocado <i>é</i> na primeira posição à direita (R1).....	72
Figura 15. VoTec – visualização normal – termo verbo	80
Figura 16. VoTec – visualização descritiva parcial – termo: verbo (português).....	80
Figura 17. VoTec – visualização descritiva parcial – termo : verb	81
Figura 18. VoTec – visualização normal tradutor – termo : verb	82
Figura 19. VoTec – visualização normal modular – termo: verb	83
Figura 20. Visualização do passo 1 para cadastro de termos no VoTec.	84
Figura 21. Visualização da página Cadastro de contextos para <i>etymology</i> (visão parcial).	84
Figura 22. Visualização parcial do <i>corpus</i> salvo em txt com dados de coleta.	85
Figura 23. Visualização da página Cadastro de contextos para o termo <i>etymology</i> (visão parcial).....	86
Figura 24. Visualização de página da elaboração da microestrutura de um verbete no VoTec.	87
Figura 25. Visualização da aba Dados da página do VoTec - <i>verb</i>	88
Figura 26. Visualização da aba Traços Distintivos da página do VoTec.....	89
Figura 27. Visualização da aba Semântica da página do VoTec.	89
Figura 28. Visualização da aba Termo Equivalente, para o termo <i>verb</i> , página do VoTec.	90
Figura 29. Visualização da aba Termos Remissivos da página do VoTec do termo Verbo.	90
Figura 30. Visualização da aba Informações Enciclopédicas da página do VoTec para o termo <i>verb</i>	91
Figura 31. Visualização da aba Conceito Final, termo <i>verb</i> , na página do VoTec.	91
Figura 32. VoTec - Visualização do termo <i>gramática</i> na área de LH.	92

Lista de Quadros

Quadro 1. Quadro comparativo das áreas de Etimologia, Filologia e LH.	29
Quadro 2. Taxonomia do <i>corpus</i> de LH (português-inglês)	49
Quadro 3. Tamanho dos <i>corpora</i> de LH.	62
Quadro 4. Candidatos a termos da LH em português e inglês.	73
Quadro 5. Legenda usada para classificação dos candidatos a termos.	75
Quadro 6. VoTec - Termos do Vocabulário bilíngue de LH.	101

SUMÁRIO

1. INTRODUÇÃO	10
1.1 Contextualização do problema de pesquisa e apresentação dos objetivos.....	11
1.2 Importância do estudo	20
1.3 Organização da dissertação	21
2. REVISÃO DA LITERATURA/FUNDAMENTAÇÃO TEÓRICA.....	23
2.1 Linguística Histórica e/ou Diacrônica.....	25
2.1.1 Etimologia, Filologia Românica, Linguística Diacrônica e/ou Linguística	
Histórica	30
2.1.2 Etimologia	31
2.1.3 Filologia e Filologia Românica	37
2.2 Terminologia e Teoria Comunicativa da Terminologia (TCT)	40
2.2.1 Terminologia Wüsteriana, a TCT de Cabré e a TST de Temmerman	41
2.3 Árvore de Domínio	46
2.4 Linguística de <i>Corpus</i>	49
2.5 Macroestrutura e microestrutura	50
3. METODOLOGIA	57
3.1 Árvore de domínio.....	57
3.2 Compilação dos <i>corpora</i>	60
3.3 WST - Lista de Palavras e Lista de palavras-chave	63
3.4 Concordanciador	70
3.5 Candidatos a termos.....	73
3.5.1 Validação de termos	77
3.6 VoTec.....	79
4. RESULTADOS.....	94
5. CONSIDERAÇÕES FINAIS	102
REFERÊNCIAS BIBLIOGRÁFICAS	104
APÊNDICES.....	109
Apêndice 1. Lista de palavras-chave em português.	109
Apêndice 2. Lista de palavras-chave em inglês.....	113

1. INTRODUÇÃO

No meio do caminho havia um dicionário – assim começa o início de uma centelha lançada que posteriormente redundaria em um trabalho científico na área de Terminologia e Terminografia. Vamos às explicações: em geral, sabemos que a maioria das pessoas não é simpática com obras lexicográficas ou mesmo terminográficas. Quando se trata de definições, sua estruturação e sua conclusão, o público é menor ainda. No meu caso, o que provocou a disposição e ímpeto para a condução deste trabalho foi uma semente plantada quando ainda era pequeno. Quando de minha concepção, meu progenitor fora acometido de um segundo descolamento de retina, comprometendo a segunda visão que ainda lhe restava. Logo, a partir de então, ele necessitava de ajuda para leitura de textos. Profissional formado em Teologia, precisava manter suas leituras filosóficas para o exercício de sua profissão. É nesse momento da história que entro em cena. Recém-alfabetizado, fora escolhido por meu pai para o ofício de leitor. Como meu pai era filho de imigrantes nipônicos, tendo cursado o ensino Fundamental em instituições de ensino de língua japonesa, seu conhecimento da língua portuguesa era limitado. Ademais, preparar falas para direcionar-se a um público, cuja língua primeira era o português, era uma tarefa um tanto quanto desafiadora e exigente. Como o menino de então, eu era quem fazia as leituras de obras teológicas que envolviam dicionários, enciclopédias e manuais. Como meu conhecimento era de um recém-alfabetizado, o conhecimento de mundo de meu pai serviu como instrução informal, assim como a leitura que eu fazia das obras lexicográficas e enciclopédicas serviam a meu pai. Assim a semente do se maravilhar com o conhecimento disposto em estruturas de dicionários e enciclopédias foi lançada na minha alma. Nesse contexto, os dicionários tinham o papel de amigo, uma entidade familiar que não lhe causava estranhamentos.

Avançando na linha do tempo, em uma atitude reflexiva, pude entender a razão da escolha de minha pesquisa. Era o continuar de um processo iniciado na minha tenra infância, que amadurecera em minha fase adulta. Já havia crescido, passado da fase de mero consulente, para uma pessoa que se interessaria pelo fazer terminográfico. O processo de produção de sentidos me encantaria e o reconhecimento e organização dos traços conceituais, que levariam a definição, consumiriam o meu âmagô como desenvolvedor deste trabalho.

Para chegar a certa maturidade, foi necessário passar pela experiência com obras monolíngues, em seguida bilíngues, monolíngues em língua estrangeira e, por fim, o contato com obras plurilíngues, que sejam de cunho onomasiológico ou semasiológico. Conhecimento coroado com o domínio do alfabeto românico e nipônico, aquele de caráter mais arbitrário, este de caráter mais onomasiológico e motivado.

1.1 Contextualização do problema de pesquisa e apresentação dos objetivos

No início de uma pesquisa de cunho científico, é necessário estabelecer as definições e conceitos que norteiam o arcabouço teórico do trabalho que se busca desenvolver, para que o trabalho seja delimitado e dimensionado. As pesquisas e estudos aqui realizados buscam diferenciar o que é um dicionário, um glossário e um vocabulário. Posteriormente, objetivamos descrever o processo de produção de uma obra terminográfica bilíngue na área de Linguística Histórica, doravante LH. Ao compararmos as definições desses tipos de obras dentro da Linguística, observamos que cada uma dessas obras lexicográficas compartilha estatutos diferentes. Logo questionamos, o que nos é apresentado como dicionário é realmente um dicionário, um glossário ou um vocabulário? Se for um dicionário, a qual grupo de dicionários ele(s) pertence(m)? É um dicionário onomasiológico, semasiológico, alfabético, não alfabético, conceitual, ideográfico, ideológico, nocional, semântico, sistemático, temático etc.? Os glossários ou vocabulários são monolíngues, bilíngues, impressos, digitais, entre outros.

Para executar esta análise, partimos do conceito de **Vocabulário** de Barbosa (1990), no qual a autora explica que o léxico de uma língua se subdivide em vocabulários, manifestos em atos de fala. O vocabulário tem como unidade padrão o vocábulo que pode ser tratado numa obra lexicográfica. Visto como um conjunto, há o **Conjunto Vocabulário** ou **Vocabulários** das ciências, de linguagens de especialidades, conjunto terminológico ou terminologia. Os vocabulários contêm as noções extraídas de um contexto a partir das palavras-ocorrências e de unidades que podem compor uma obra lexicográfica ou terminográfica. O objetivo da obra terminológica é o de compilar, registrar e descrever a unidade lexical que representa uma noção específica - a unidade terminológica. Isto é, na Lexicologia/Lexicografia, temos a unidade lexical como objeto de estudo. Na Terminologia/Terminografia, há a unidade terminológica (BARROS, 2004).

Os dicionários são obras lexicográficas que disponibilizam definições para o vocabulário de língua geral ou de especialidade para os consulentes. Béjoint (2010), ao definir o termo **dicionário**, afirma que ele é proveniente das formas latinas *dictionary* ou *dictionary*, cuja raiz é *dictio* - ato de dizer ou palavra, do verbo *dicere* – dizer. No contexto anglófono, de acordo com o *Oxford Dictionary of English*, foi usado pela primeira vez em 1222 d.C. pelo poeta e gramático Joannes de Garlandia como título de sua lista de vocábulos, máximas e provérbios latinos. Esse glossário era organizado alfabeticamente, com mais de 3000 palavras da Bíblia para uso de alunos de Teologia.

Béjoint (2010) afirma que os dicionários possuem uma estrutura dupla: a macroestrutura e a microestrutura. A primeira é a organização das listas de palavras “o conjunto completo de entradas organizado em certa ordem, parte da qual é sempre usada para uma leitura vertical, quando o consulente procura uma informação específica”. A macroestrutura é comumente chamada ‘nomenclatura’ (REY-DEBOVE 1971, p. 21). A segunda é composta por certo número de informações ou campos que variam de acordo com a natureza do dicionário, podendo ser pronúncia, etimologia, sinônimos, ortografia etc. Isto é, no dicionário a macroestrutura é feita de palavras de uma língua, enquanto que a microestrutura traz informações sobre essas palavras, não aos elementos aos quais elas se referem (BÉJOINT, 2010, p. 13).

Ao tratar da distinção entre Lexicologia e Lexicografia, Parreira (2006) classifica o dicionário como hipônimo de obra lexicográfica. Segundo a autora, uma obra lexicográfica é “comumente definida como uma lista ordenada de palavras seguidas de um texto que informa sobre cada uma delas.” Quanto à sua serventia, comenta que “sua principal função é informar.” (PARREIRA, 2006, p. 18, 19). Um segundo questionamento levantado é: os dicionários especializados da Linguística trazem uma microestrutura de dicionário ou de enciclopédia? Em geral, os dicionários de Linguística optam pela segunda opção. Foram analisados cinco dicionários, sendo quatro da área de Linguística e um de termos literários: o Dicionário de Linguística, coordenado por Izidoro Blikstein (2004); o Dicionário de Linguística e Gramática de Mattoso Câmara Jr. (1986); o Dicionário de Linguística e Fonética de Davis Crystal, traduzido e adaptado por Maria Carmelita Pádua Dias (2000); o Dicionário de Linguística da Enunciação de Valdir do Nascimento Flores et. al.(2009) e o Dicionário de termos literários de Massaud Moisés (1995). Essas obras

disponibilizam conceitos e definições das entradas, bem como uma microestrutura com definições de abordagem histórica e explicativa (DUBOIS, 1999).

Com relação à construção da definição, Finatto (1998) propõe alguns parâmetros para a análise do paradigma definicional, que pode ser de cunho lexicográfico, enciclopédico ou terminográfico. Contudo a autora afirma que esta é uma área na qual não há clareza suficiente por parte dos pesquisadores em muitos aspectos.

Finatto (1998) conceitua o **definir** como: “estabelecer um vínculo entre um termo, um conceito e um significado” (1998, p.1). O que fizemos neste trabalho terminológico foi identificar os termos por meio da lista de palavras-chave, identificar os traços conceituais com o uso do concordanciador. Após a organização dos traços na ficha terminográfica do VoTec (FROMM, 2007), eles foram sistematizados para a construção do significado e do paradigma definicional, exposto na definição final para o usuário da obra. Para definirmos essa plataforma, dizemos que o “VoTec é uma ferramenta que se vale de *corpora* técnicos para a construção de seus verbetes e de um banco de dados (ambos exaustivamente descritos) para o seu funcionamento” (FROMM, 2007, p. 8).

Ademais, para a formulação de uma definição, entram em cena valores distintos e “potencialidades de conhecimento e significação”, o que se constitui como um trabalho metalinguístico. Isto é, traços conceituais são identificados, analisados e extraídos de contextos para a construção do significado e da definição final.

Quanto aos tipos de definição, (i) a definição de cunho lexicográfico é aquela predominantemente linguística, “tratando mais de palavras”; (ii) a definição enciclopédica se ocupa “mais de referentes e de descrição de coisas”; (iii) e a definição terminológica traz “conhecimentos formais sobre coisas e fenômenos. (FINATTO, 1998, p. 2).

Considerando os aspectos acima expostos, podemos descrever o perfil das definições encontradas nos dicionários analisados para servir de base comparativa para o trabalho que desenvolvemos. O Dicionário de Linguística, coordenado e traduzido por Izidoro Blikstein (2004), apresenta paradigmas definicionais de cunho lexicográfico, enciclopédico e terminológico. Apesar de trazer definições que poderiam ser enquadradas no perfil lexicográfico, traz elementos como a classificação gramatical e etimologia somente para alguns verbetes, não havendo uma padronização. Portanto, nota-se a prevalência da definição de caráter enciclopédico.

Quanto ao Dicionário de Linguística e Gramática de Mattoso Câmara Jr. (1986) observamos que há uma forte tendência ao uso de remissivas, e não há a classificação verbal dos termos. O autor busca trazer definições sucintas e exemplificar o termo definido ou mesmo o conceito. Para os verbetes, há a presença de um paradigma definicional de cunho parcialmente lexicográfico, enciclopédico e terminológico.

Os paradigmas definicionais do Dicionário de Linguística e Fonética de Davis Crystal (2000) apresentam uma estrutura mais padronizada, na qual a maioria dos termos serão definidos da seguinte forma: termo usado/utilizado na/por etc. Essa é uma estrutura mista com nuances dos padrões lexicográfico, enciclopédico e terminológico. A microestrutura traz diagramas e esquemas para explicar os conceitos, além de trazer várias remissivas indicando autores e obras e os termos contidos na própria obra. Quando os termos contidos nas definições fazem parte da obra, eles vêm transcritos em letras maiúsculas, indicando ao leitor a possibilidade de explorá-los na obra, caso deseje. Outra estrutura usada pelo autor para marcar as remissivas são as informações em parênteses, na qual encontramos a abreviação de conferir em parênteses (cf. REMISSIVA). O paradigma pragmático é introduzido pelo verbo Ver. Em seguida, o autor cita o autor e ano, dois pontos e o capítulo indicado para leitura. Ex.: Ver Robins 1980: Cap. 1.

O Dicionário de termos literários de Massaud Moisés (1995) apresenta uma estrutura definicional predominante de uma obra lexicográfica no início. O autor traz a etimologia dos termos, quer seja da origem clássica, quer seja proveniente de outras línguas ou o sentido literal do termo para auxiliar o consulente na compreensão do sentido. O autor optou por iniciar o paradigma definicional usando verbos: designava, diz-se, utilizado, designa; ou substantivos: problema, figura de linguagem, expressão latina etc. Analisado o conteúdo, detectamos a presença de características de cunho lexicográfico, enciclopédico e terminológico.

Somente a título de exemplificação, trazemos abaixo, na Figura 1, o padrão de macro e microestrutura, extraído do **Dicionário de Linguística da Enunciação**, para o termo **semiótico**.

semiótico *s. m.* Benveniste

Outras denominações: domínio do semiótico, modo semiótico, ordem semiótica.

Definição: sistema de signos linguísticos, cuja significação se estabelece intrassistema, mediante distinção.

Fonte da definição: EBE06.

Nota explicativa: Benveniste utiliza o termo *semiótico* em oposição a *semântico*, para traçar uma divisão entre dois domínios da língua. O campo do semiótico corresponde à língua na acepção saussuriana, sendo o signo sua unidade. Embora esse linguista assim delimite o sistema semiótico, este não é independente do semântico, e vice-versa, pois ambos se superpõem, constituindo a língua tal como é utilizada.

Fonte da nota: EBE06.

Leitura recomendada: BEN89A; BEN95D.

Termos relacionados: forma, língua (2), semântico.

Figura 1. Macro e microestrutura – termo: semiótica
Fonte: Flores *et alii*. (2009)

A estrutura do dicionário da Figura 1 é mais atual, com campos diferentes para as estruturas: classificação gramatical, a qual autor pertence, outras denominações (*clusters*), definição clara e objetiva (oração única, padrão GPDE), fonte da definição, nota explicativa, fonte da nota, recomendação de leitura e termos relacionados/remissivos. Observamos que nessa obra há, de uma forma mais evidente, características de padrões definitórios lexicográficos, enciclopédico nas notas explicativas e terminológico no campo definição. Nesse campo, observamos que há a predominância de uma oração que inicia e finaliza com a definição terminológica do verbete. Outra característica que fica evidente na obra é a preocupação em homogeneizar a obra para que se tornasse legível para os não especialistas (p.9). Essa preocupação com o usuário/consulente é um diferencial relevante para uma obra terminográfica tanto no que tange ao processo de elaboração, quanto ao processo de uso após a publicação.

Quanto à classificação de obras lexicográficas, os dicionários, segundo Landau (2001, p.7-42), podem ser monolíngues, bilíngues, plurilíngues, onomasiológicos, semasiológicos, para falantes nativos ou estrangeiros; alfabético ou temático; escolar, acadêmico ou comercial; para adultos ou infantil; geral ou de períodos específicos (do séc. XVIII, por exemplo); conciso, de bolso ou mini; geral ou especializado; dialetal, etimológico, de pronúncia, de sinônimos; de abreviações, de expressões ou fraseológico; prescritivo ou descritivo.

Considerado o fato de que os estudos europeus, na área de Filologia Românica, precedem os estudos de LH da academia brasileira e que eles seriam relevantes para o público lusófono, um questionamento surge: por que não há dicionários bilíngues? Além disso, por que não há dicionários especializados de LH disponíveis no mercado? Será que, com o desenvolvimento e difusão de livros em formato digital em um ritmo crescente, os dicionários impressos continuarão a ter a mesma procura que os dicionários digitais ou em formato eletrônico? Os estudiosos da Linguística brasileira, como Mattoso Câmara, Basseto, Mattos e Silva, parecem ter lançado mão de uma bibliografia estrangeira monolíngue e, a partir delas e de suas experiências, produziram as obras que serviram ao estabelecimento dessa ciência em território nacional.

Além desses questionamentos, tratamos da delimitação das subáreas da Linguística que contribuem para a formação e constituição da árvore de domínio da Linguística Histórica, tais quais a Etimologia, a Filologia Românica e a Linguística Diacrônica e/ou LH. Explicitar a função da árvore de domínio na construção de um banco de dados terminológico foi um dos objetivos específicos deste trabalho. A árvore de domínio é uma forma sistematizada de se dispor as informações, em forma de diagrama, para que o leitor possa ver o conjunto como um todo, bem como suas especificidades. Consequentemente propusemos um olhar diacrônico sobre o histórico da LH, retomando-o a partir da Etimologia e Filologia, nos primórdios da civilização grega, até os dias atuais, envolvendo a Linguística Diacrônica. Esta delimitação serviu para estabelecer os limites conceituais entre essas subáreas. Com relação à nomenclatura, a Filologia Românica, ao mencionarmos Filologia de língua portuguesa será a que esteve em foco. Ao tratarmos do *corpus* de língua inglesa, a Filologia Germânica será a referida quanto aos estudos de LH do inglês.

Reconhece-se a Etimologia como a ciência mais antiga, no que se trata do estudo linguístico, datada sua existência de há pelo menos 25 séculos (VIARO, 2011, p.29). Seu objetivo, um tanto quanto filosófico, era o de resgatar a essência, a originalidade do objeto, ligado ao processo de denominação. Esse processo implicava numa conexão entre a descrição do mundo real e a nominalização por meio de um processo psicológico, subjetivo e abstrato da sociedade. A Filologia, como ciência primordial, é subdividida em Clássica, Românica, Germânica, entre outros, dependendo de seu objeto de estudo. A partir de estudos comparativos dos filólogos e neogramáticos, a Filologia Românica rapidamente se

fortaleceu, tendo as leis fonéticas confirmadas por meio de *corpus* abundante, disponível em língua latina, o que pode ser observado na gramática comparada de Friedrich Christian Diez (1836-1844). O advento do Curso de Linguística, compilado pelos discípulos de Saussure, ajudou a consolidar uma mudança de perspectiva nos estudos das línguas, consolidando o foco no estudo sincrônico das línguas. É a partir da segmentação do olhar diacrônico, característico dos filólogos, e sincrônico, dos futuros linguistas, que se subdivide a Linguística em sincrônica ou diacrônica. Em geral, Linguística Diacrônica é sinônimo de Linguística Histórica. A busca de *corpora* para o desenvolvimento deste projeto mostra que esta concepção ou denominação é comum na academia brasileira, bem como na de fala inglesa.

A área de LH foi a subárea da Linguística escolhida para que o objetivo geral desta pesquisa fosse cumprido: criar uma obra terminológica bilíngue. As línguas escolhidas para o vocabulário foram português-inglês e as definições, em contraste, das unidades terminológicas foram elaboradas a partir de um banco de dados terminológico, na área de LH. A **definição em contraste**, em obras bilíngues, é aquela cujo enunciado não é uma tradução equivalente nas duas línguas. Isto significa que as definições das unidades terminológicas em português e inglês trouxeram traços conceituais provenientes dos *corpora*, logo, os enunciados construídos para as definições foram diferentes, apesar de serem semelhantes quanto ao conceito. A este tipo de definição dizemos que ela é proveniente de um estudo baseado em *corpus*. O banco de dados ou *corpora* foi constituído de textos científicos das subáreas previamente listadas, sendo composto por artigos científicos, anais de eventos, dissertações, teses e, excepcionalmente, um dicionário enciclopédico de Etimologia, o que será explicado, adiante, na metodologia.

A metodologia e abordagem escolhidas para o tratamento dos *corpora* foi a Linguística de *Corpus* – LC -. Isso se deu por vários motivos: (i) possibilitar a descrição da língua em uso, (ii) possibilitar a análise quantitativa e qualitativa de dados linguísticos, (iii) fornecer resultados estatisticamente confiáveis e precisos dos termos usados pelos especialistas. Contudo, é importante elucidar as concepções do termo *corpus* nesta pesquisa.

Corpus é um termo frequente nas duas áreas principais deste projeto. Entretanto, as definições dadas para o termo nas duas áreas diferem entre si. Na LH, *corpus* é o material que serve de fonte para a investigação linguística, independentemente de seu tamanho ou

gênero. Isto é, os trabalhos desenvolvidos por esta disciplina, partem de *corpora* para fazerem a investigação científica da língua e das mudanças linguísticas e buscam “representar e investigar estágios passados de uma língua e/ou estudar a mudança linguística” (MATTOS E SILVA, 2008; CLARIDGE, 2008, p. 242 *apud* KYTÖ, 2011, p. 418).

Já na LC, a definição dada para *corpus* (VIANA, 2010) é a de uma compilação de textos de ocorrência natural que representa uma certa língua ou seus aspectos específicos, possibilitando uma análise linguística pré-estabelecida. Tagnin (2015) define *corpora* (*corpus* no plural) da seguinte forma: “bancos de textos de linguagem autêntica, criteriosamente construídos, destinados à pesquisa e legíveis por computador.” (2015, p.20). Observa-se que para a LC, o fato de um *corpus* ser representativo da língua em estudo é de alta relevância devido ao seu aspecto probabilístico.

Como abordagem, a LC foi escolhida pelos seguintes aspectos: (i) ela permite que os dados a serem usados sejam provenientes dos *corpora*, ou seja, as definições foram construídas a partir dos dados provenientes dos *corpora* de LH em inglês e português; (ii) a partir desses dados, podemos observar a regularidade dos termos que aparece nas linhas de concordância, o que mostra a repetição de itens que coocorrem; (iii) a partir da localização dos candidatos a termos nas linhas de concordância, pudemos buscar os contextos definitórios, por meio dos traços semânticos neles contidos. Já a LH foi escolhida, pois é uma área que pode ser útil ao ensino de língua portuguesa e para o ensino de línguas estrangeiras, mais especificamente com sua contribuição com os metaplasmos¹. Além dessa vantagem, há aspectos das áreas da Fonologia e Ortografia que permeiam o ensino da língua portuguesa e estrangeira, tais quais inglês, francês e espanhol que podem ser elucidados com o auxílio da LH. Além da LH, as Filologias Clássica, Românica e Germânica foram relevantes para a pesquisa, já que seus estudos servem de *corpus* para a constituição da LH em ambas as línguas.

Para que esse objetivo fosse atingido, outros específicos se fizeram necessários:

¹ Metaplasmos: mudança fonética que consiste na alteração de uma palavra pela supressão, adição ou permuta de fonemas: a elisão e a síncope são exemplos de metaplasmos. (BLIKSTEIN, 2004, p. 412)

- (I) Reelaborar a árvore de domínio (BARROS, 2004), a partir daquela proposta por Fromm e Yamamoto (2013), das áreas relacionadas à LH, para que traços conceituais de cada uma delas fossem delineados;
- (II) Compilar *corpora* bilíngue, que serviram como banco de dados de onde os traços conceituais foram extraídos a partir dos contextos definitórios e explicativos;
- (III) Extrair candidatos a termos, provenientes das listas de palavras-chave, após processamento dos *corpora* pelo *WordSmith Tools* 6.0 (SCOTT, 2012), doravante WST. Os candidatos foram escolhidos, desde que estivessem presentes nas listas provenientes dos *corpora* do português e do inglês. Busca posterior confirmação com especialistas da área de LH será o segundo passo deste processo, cumprindo requisito metodológico do trabalho terminológico; Veja observações sobre este passo na seção de resultados.
- (IV) Disponibilizar dados oriundos de fontes confiáveis aos usuários, já que as definições não foram construídas aleatoriamente, nem de forma totalmente subjetiva, pelo fato de serem construídas a partir dos *corpora* acadêmicos;
- (V) Ordenar os dados, de forma que atendam padrões terminográficos², em fichas de registro, como entradas em vocabulários. Este passo objetivou disponibilizar aos leitores uma obra de leitura acessível, organizada e padronizada;
- (VI) Elaborar definições para os termos selecionados, baseadas no padrão GPDE, gênero próximo e diferença específica, adotado nas obras terminológicas em nível nacional. Esta definição partiu das noções mais amplas, da relação de hiperonímia, para aquelas mais específicas, as relações de hiponímia (ILARI, 2002, p.39);
- (VII) Disponibilizar os resultados no VoTec, ambiente de gestão terminológica < www.pos.voteconline.com.br >, disponível *on-line*, de acesso público gratuito. Esta plataforma oferece subsídios aos tradutores, escritores, estudiosos da Linguística, oferecendo modos distintos de visualização, facilitando o acesso dos consulentes.

² Alguns desses padrões são: para cada conceito, tivemos um entrada diferente. Ex.: *Language* e *Language2*; informação mais recortada, delimitada (FINATTO, 2006)

1.2 Importância do estudo

Desde o surgimento da Etimologia, da Filologia, da Linguística e suas subáreas, observa-se que o estudo da linguagem tem sido relevante para a humanidade. Com a invenção do computador, a administração, o gerenciamento e a preservação de documentos históricos de relevância linguística têm se tornado mais acessíveis e factíveis. Consequentemente, o acesso a informações publicadas na atualidade é desejável aos pesquisadores que buscam desvendar as mudanças linguísticas às quais as línguas são submetidas. Contudo, muitas obras são publicadas em inglês, já que esta língua tem ocupado o lugar da *lingua franca* no século XXI. Em contrapartida, o português do Brasil tem sido objeto de interesse de vários países devido à sua importância comercial no cenário mundial. Além dessa razão, os intercâmbios dos Estudos da Linguagem do Brasil com outros países são relevantes devido à identidade que o Brasil tem construído no cenário internacional. Diante disso, um dicionário bilíngue na área de LH seria de grande valia para impulsionar e facilitar o acesso à informação para o público lusófono e para os usuários da língua inglesa em geral.

O fato de que a Filologia Românica teve seu florescimento em países da Europa, como Alemanha, Inglaterra e França, e como um número considerável de obras desse cunho foram publicadas primeiramente em alemão e francês, foi outro motivo considerado para a condução deste projeto (IORDAN, 1982; VIDOS, 1996; BASSETO, 2001; FARACO, 2005). Devido à difusão do inglês, muitas obras basilares da Filologia foram publicadas ou traduzidas para a língua inglesa. A Filologia Românica é, até os dias atuais, estudada nos centros acadêmicos da Europa e continua a atrair mais estudiosos para o estudo de suas origens e desenvolvimento a partir do império romano. Essas obras seriam inacessíveis aos pesquisadores que não dominam a língua inglesa e muito do que foi construído até então seria deixado de lado. Ao disponibilizar um dicionário terminológico na internet, os termos específicos de LH seriam disponibilizados em nível mundial, bem como contribui para o desenvolvimento da disciplina em vários países onde é estudada. Considerando-se o fato de que o português brasileiro tem sido estudado nos quatro cantos do mundo, este trabalho mostra sua relevância, pois o estudo de LH já não se limita ao Brasil e Portugal.

O uso de *corpora* acadêmicos e autênticos possibilitou a construção de uma obra robusta, já que a fonte dos contextos foi original e não traduzida de uma língua a outra.

Ademais, o fato de os *corpora* serem sincrônicos os conteúdos desenvolvidos neles permitem o acompanhamento da terminologia atual, em consonância com as pesquisas desenvolvidas a nível mundial.

Com relação ao usuário/consulente, como são textos de especialistas da área, a probabilidade de a obra atender, de forma satisfatória, a necessidade deste mesmo público é considerável. O perfil dos gêneros selecionados também possibilita uma maior funcionalidade e utilidade para a obra, já que os consulentes de um dicionário terminológico serão, em sua grande maioria, profissionais, acadêmicos e estudantes da Linguística que necessitam se comunicar com especialistas da área, mas limitados pelo fator língua estrangeira.

O perfil do usuário ponto importante quando planejam uma obra terminográfica e consideramos esse fator durante todo o trajeto do fazer terminográfico. O usuário almejado para esta obra são os alunos veteranos do Curso de Letras, aquele que cursa a disciplina de LH, estudantes de programas de pós-graduação em Linguística, pesquisadores dessa área. Isto é, em geral, são estudantes ou profissionais que já atingiram uma certa maturidade acadêmico-científica para lidarem com os estudos históricos de língua portuguesa e inglesa.

1.3 Organização da dissertação

Esta dissertação foi organizada em 5 capítulos. Neste capítulo introdutório, contextualizamos nosso estudo no cenário atual, apresentamos nossos objetivos e destacamos a importância desta pesquisa. No capítulo 2 fazemos a revisão da literatura expondo a razão da escolha da LC, da teoria da Terminologia adotada – TCT -, explicamos a árvore de domínio da Linguística, e as subáreas que compõem a LH: a Etimologia, a Filologia Românica e a Linguística Histórica e/ou Diacrônica.

No terceiro capítulo, apresentamos a metodologia por meio do uso do WST para produção das listas de palavras, listas de palavras-chaves e do concordanciador, responsável pelos contextos, fontes dos traços conceituais das unidades terminológicas. O VoTec foi detalhado passo a passo, por meio de texto e imagens para melhor compreensão do leitor. Explicamos os procedimentos desde o registro até a disponibilização do conteúdo que está exposto *on-line* aos consulentes.

O quarto capítulo é o de resultados, no qual apresentamos alguns exemplos de verbetes desenvolvidos e tecemos algumas considerações sobre a elaboração dos mesmos.

O quinto capítulo apresenta as considerações finais, mostrando os resultados gerais atingidos e as repercussões futuras desta pesquisa.

2. REVISÃO DA LITERATURA/FUNDAMENTAÇÃO TEÓRICA

O objetivo deste capítulo é apresentar os fundamentos teóricos que embasaram este trabalho de pesquisa e a produção do Vocabulário na área de LH concebido como um produto terminológico, vocabulário este que é fruto de um fazer terminográfico e de uma pesquisa terminológica. Primeiramente, tratamos aqui da questão denominativa da área da Linguística Histórica, que hora é chamada dessa forma, hora é denominada Linguística Diacrônica, uma vez que esta discussão ainda não foi levada a termo na academia brasileira. Para este trabalho, elegemos o termo Linguística Histórica (LH) ao tratarmos dessa subárea da Linguística - e reservaremos o termo Linguística Diacrônica para a abordagem de estudo da Língua.

A LH se constitui como disciplina independente dentro da Linguística, mas para se constituir, ela se vale de dados provenientes de *corpus* das outras disciplinas da Linguística: a Filologia Clássica, Românica e Germânica e da Etimologia em seu contexto lusófono e anglófono. Portanto, é justo que dediquemos parte deste capítulo para tratar dessas disciplinas que contribuem de forma vital para a consolidação da LH.

Por ser um trabalho na área da Terminologia e Terminografia, fizemos um apanhado das correntes existentes dentro da Terminologia, caracterizando a Teoria Geral da Terminologia de Wüster, a Teoria Comunicativa da Terminologia de Cabré – adotada como fundamento deste trabalho -, e a Teoria Sociocognitiva da Terminologia de Temmerman.

Como parte primordial/constituente do fazer terminológico, a Árvore de Domínio é um procedimento essencial. Assim sendo, dedicamos parte deste capítulo a este conteúdo para defini-la, explicar sua contextualização e justificar a proposta da área da Linguística proposta por Fromm e Yamamoto (2013). Este trabalho de esquematizar uma área tão vasta quanto a Linguística ainda está em processo de amadurecimento, mas as subáreas que se fundem, contribuindo para a consolidação da LH, já se encontram delineadas. Explicaremos este percurso em detalhes.

Por configurarem subáreas da Linguística Teórica, abaixo da grande área da Linguística, enquadramos a Filologia - subdividida em Clássica, Românica e Germânica como parte do objeto de estudo deste trabalho -, a Etimologia, e finalmente a Linguística Histórica ou Diacrônica, essas duas serão contempladas neste capítulo.

A metodologia, bem como a abordagem escolhida para esta pesquisa foi a Linguística de *Corpus*, logo é importante que reservemos a ela um espaço para apresentá-la e justificarmos sua escolha ou adoção. Para tal, nos valeremos de Sardinha (2004) e Viana e Tagnin (2010) e suas contribuições para a área de Terminografia e Lexicografia no Brasil e de seus trabalhos de cunho bilíngue já desenvolvidos na academia.

Quanto à organização das informações usadas nas definições, temos a macro e microestrutura de uma obra terminográfica, que diferem daquelas de uma obra lexicográfica. Contudo, a tradição de composição dessas estruturas na Terminografia se estabelece a partir do que já fora posto na escola lexicográfica. Portanto, dedicamos um espaço para diferenciar as duas práticas, que possuem estilos parcialmente distintos, contudo, ao mesmo tempo semelhantes, a partir de parâmetros propostos por Bevilacqua e Finatto (2006).

Especificando mais sobre este trabalho, entendemos que este estudo se insere na área da Terminologia e Terminografia, no intuito de buscar as unidades terminológicas, conceituá-las e defini-las em uma obra terminográfica. Para atender ao rigor teórico-metodológico da Terminografia, adotamos a Linguística de *Corpus*, observando-se os princípios de frequência, representatividade terminológica de natureza probabilística (BERBER SARDINHA, 2004; VIANA; TAGNIN, 2010). Como teoria da Terminologia, ancoramos este trabalho na Teoria Comunicativa da Terminologia de Cabré (1999), que valoriza os aspectos comunicativos da Terminologia e concebe as unidades terminológicas como parte da linguagem natural e da gramática das línguas. A ferramenta utilizada para processamento dos *corpora* e a extração dos candidatos a termos para o VoTec (FROMM, 2007; YAMAMOTO, 2013) é o console do WST. Usaremos as ferramentas lista de palavras, a de palavras-chave e o concordanciador. Os *corpora* são de cunho científico da área de Etimologia, Filologia Românica e LH.

Quanto à diferenciação do estatuto das disciplinas que interagem entre si para constituição da LH, fizemos uma pesquisa bibliográfica e, a partir dos princípios dos contextos definitórios (AUBERT, 1996), buscamos entender em que ponto elas diferem entre si. O contexto definitório é aquele que traz traços semânticos que são utilizados para a representação conceitual da unidade terminológica a partir de seu uso. É o contexto que traz informações constitutivas do signo; essas informações podem ser processadas e usadas para a construção da definição de uma obra terminográfica.

Feita esta diferenciação, partimos para o próximo passo, o de construção da árvore de domínio da Linguística. Nesse estudo, buscamos pesquisar a Etimologia (VIARO, 2011; DURKIN, 2009), Filologia Românica (IORDAN, 1982; VIDOS, 1996; BASSETO, 2001), Linguística Diacrônica (VIARO, 2014) e LH, (TARALLO, 1990; MUSSALIM e BENTES, 2012; WEEDWOOD, 2002; FARACO, 2005).

2.1 Linguística Histórica e/ou Diacrônica

Para iniciar a discussão sobre a LH, trazemos a contribuição de Faraco (2005), que propõe ao leitor a diferença entre os princípios de Diacronia e Sincronia colocados no *Curso de Linguística Geral* de Saussure. Baseado nesses princípios, o autor chama a atenção para as duas dimensões dos estudos linguísticos: a diacrônica ou histórica e a sincrônica ou estática, juntamente com os pressupostos de mutabilidade, para a primeira, e de imutabilidade, para a segunda. Considerando-se os princípios metodológicos, a linguística sincrônica ou descritiva se ocupará da “investigação dos diferentes estados da língua” enquanto a histórica tratará das mudanças linguística no tempo. Considerados esses fatores, o autor explica que os estudos linguísticos, no Brasil do século XX, vão privilegiar os estudos sincrônicos aos diacrônicos. Para tratar da origem da LH, o autor faz um apanhado dos estudos da linguagem que a precederam, retomando os estudos linguísticos dos hindus no século IV a.C., a filosofia nos gregos, os cuidados filológicos dos alexandrinos, as gramáticas latinas, os filósofos da Idade Média, a filosofia árabe, os estudos renascentistas e a gramática de Port-Royal do século XVII. Faraco (2005) aborda, então, o assunto das mudanças das línguas nos estudos filológicos em várias sociedades humanas. Nesse momento, ele conceitua a Filologia como “o estudo de textos antigos com o objetivo de estabelecer e fixar sua forma original” (p. 131), ao que tomamos por exemplo, os sábios hindus do século IV a.C., os alexandrinos em II a.C. e os intérpretes do Corão na Idade Média.

O autor data o início da LH, conduzida dentro dos pressupostos da cientificidade moderna (fundamentação empírica e modelos teóricos), do final do século XVIII e explica que os estudos filológicos que a precederam são indispensáveis a ela. Essa crítica textual contribui para a reflexão sobre as línguas e sua variabilidade no tempo. O percurso da LH pode ser subdividido em aproximadamente dois momentos: (i) o estabelecimento e consolidação do método comparativo, de 1786 a 1878, quando há o manifesto dos

neogramáticos, e (ii) o período da coexistência das linhas interpretativa e a do gerativismo. A interpretativa é de caráter imanentista, herdeira dos neogramáticos, adepta do estruturalismo, fundada na dialetologia e na sociolinguística, na qual a mudança linguística é concebida a partir do contexto social dos falantes (fatores internos e externos); o gerativismo concebe a mudança de uma forma imanentista, condicionada somente pelos fatores internos da língua.

Para explicar a gênese da LH, Faraco (2005) retoma os estudos do sânscrito pelos intelectuais europeus. Em 1786, William Jones ressaltou as semelhanças existentes entre o sânscrito, o latim e o grego à Sociedade Asiática de Bengala por meio de uma comunicação. Gramáticas e dicionários da língua clássica dos hindus foram publicadas e, em 1795, em Paris, a Escola de Estudos Orientais se transformou em um centro de investigação que abrigaria intelectuais como Friedrich Schlegel e Franz Bopp. Schlegel publicou o texto *Über Sprache und die Weisheit der Inder* (Sobre a língua e sabedoria dos hindus), obra marco para os estudos comparativos alemães, onde, além das línguas analisadas por Jones, ele incluiu o persa e o germânico. Nessa obra, analisaram-se semelhanças lexicais e gramaticais entre essas línguas, interpretadas como originárias de uma língua comum, posteriormente nomeada de indo-europeu. O método comparativo, procedimento fundamental na LH, consolidou-se com Bopp, em seu livro *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen, und germanischen Sprache* (Sobre o sistema de conjugação da língua sânscrita em comparação com o da língua grega, latina, persa e germânica), onde esse autor fez a comparação morfológica verbal dessas línguas em 1816. Contudo, o trabalho de Bopp era só comparativo e não considerou uma análise diacrônica das línguas estudadas. Esse estudo viria com Jacob Grimm em seu livro *Deutsche Grammatik* (Gramática alemã) de 1822 (2ª ed.), no qual ele interpretou correspondências fonéticas entre as línguas, num período de quatorze séculos, como consequência das mudanças no tempo. Nesse estudo foram considerados o sânscrito (1000 a.C.), o grego (séc. V a.C.), o germânico (IV d.C.), o eslavo (IX d.C.) e o persa moderno.

Nos anos seguintes, a Filologia ou Linguística Românica se desenvolveria caracterizada pelo estudo histórico-comparativo das línguas originadas do latim, cujo pioneiro seria Friedrich Diez, autor de uma gramática histórico-comparativa e um dicionário etimológico das línguas românicas, publicados entre 1836 e 1854. Devido ao

número abundante de obras preservadas em latim, foi possível haver um refinamento metodológico nos estudos histórico-comparativos, já que formas ascendentes puderam ser atestadas.

Sobre os neogramáticos, Faraco (2005) explica que:

A última metade do século XIX ficou caracterizada como a época dos neogramáticos, uma nova geração de linguistas relacionados com a Universidade de Leipzig (Alemanha) que, questionando certos pressupostos tradicionais da prática histórico-comparativa, estabeleceu uma orientação metodológica diferente em um conjunto de postulados teóricos para a interpretação da mudança linguística (FARACO, 2005, p.139).

A partir dos neogramáticos, uma concepção psicológica da língua é iniciada, preocupada em investigar os mecanismos da mudança das línguas. As mudanças fonéticas eram abordadas tendo em vista que “afetavam a mesma unidade fônica em todas as suas ocorrências, no mesmo ambiente, em todas as palavras, não admitindo exceções” (FARACO, 2005, p.141).

Depois desse breve recuo no tempo, voltaremos a abordar a conceituação da disciplina LH por Faraco (2005, p.13). Segundo o autor, a LH ocupa-se de estudar as “mudanças que ocorrem nas línguas humanas à medida que o tempo passa, atividade específica dos estudiosos de linguística histórica”. Como parte das mudanças, há as mudanças fonético-fonológicas, as morfológicas, as sintáticas, as semânticas, as pragmáticas e as lexicais. A mudança é de caráter lento, gradual e regular, explicada por leis fonéticas, pela analogia e pelo encaixamento estrutural e social.

Rumo à conclusão desta reflexão, apresentamos algumas definições e diferenciações propostas por Mattos e Silva (2008, p.14) para as áreas de Filologia e LH. A autora afirma que a Filologia é “uma das formas de se abordar a documentação escrita, tanto literária como documental em sentido amplo, enriquecida pelas vias da crítica textual, tanto de textos antigos como modernos” (p. 14). Ela explica que, sem o trabalho predecessor de um filólogo, os estudos da mudança linguística, mesmo a gerativa, seria impraticável. No que tange à definição da LH, Mattos e Silva (2008, p.8) a apresenta com um campo da linguística que busca “interpretar mudanças – fônicas, mórficas, sintáticas e semântico-lexicais – ao longo do tempo histórico, em que uma língua [...] é utilizada por seus utentes em determinável espaço geográfico [...]” (p. 8).

Mattos e Silva (2008) subdivide a LH em duas vertentes: a *lato sensu* e a *stricto sensu*. A primeira, como os estudos linguísticos baseados em *corpora*, trabalha com “dados datados e localizados [...] tal como os estudos descritivos, sobretudo do estruturalismo americano, que teve seguidores no Brasil [...] (MATTOS E SILVA, 2008, p. 9), além de incluir as “teorias do texto, do discurso e da conversação” (p. 9) baseados em *corpora*. A segunda estuda as mudanças nas línguas no tempo, à medida que são usadas. Ela pode ser caracterizada por duas orientações: (i) a LH sócio histórica - que leva em consideração fatores intra e extralinguísticos como em Labov, e a sócio histórica em S. Romaine; e (ii) a diacrônica associal, que se vale somente de fatores intralinguísticos, presente nos estruturalistas diacrônicos, exemplificada por A. Martinet, e nos gerativistas diacrônicos, como em D. Lightfoot.

Nessa busca da diferenciação conceitual da Filologia e da LH, para que possamos dar a elas seu lugar devido na árvore de domínio, entendemos que:

a. tanto Faraco (2005) quanto Basseto (2001) são unânimes na definição da Filologia. Os dois estudiosos a definem como a área da ciência que busca estudar, analisar, e explicar os textos a partir de seu contexto linguístico, histórico, político, e social de produção e os explica, num dado momento da história humana, numa perspectiva sincrônica;

b. Faraco (2005) diferencia a Filologia da LH, pois esta tem como objeto de estudo as mudanças que ocorrem numa língua numa perspectiva diacrônica, que se vale de um método comparativo-histórico.

Apesar de compartilharem pontos de vista em comum, os dois autores abordam aspectos das disciplinas de formas distintas. Ao apresentar Bopp, Basseto (2001) o faz dizendo que era reconhecido como filólogo, enquanto Faraco (2005) o apresenta como parte dos intelectuais alemães da época, não o nomeando como filólogo. Ao comentar sobre o advento da obra de Saussure, Basseto afirma que, naquela época, a concepção de linguista e filólogo era indissociável. Já Faraco, ao apresentar os neogramáticos, os apresenta como linguistas, enquanto Basseto se restringe ao termo neogramático para nomear alguns estudiosos da linguagem em sua obra.

Para concluir esta discussão teórica sobre o domínio em foco, intentamos apresentar adiante o quadro comparativo das subáreas (Quadro 1) que compõem a LH nos dias de hoje, baseado na pesquisa bibliográfica.

Quadro 1. Quadro comparativo das áreas de Etimologia, Filologia e LH.

ÁREA	OBJETO	MÉTODO	OBJETIVO
Etimologia	Étimo	Gramático-histórico, etimológico (corpus, datação)	estudo da origem das palavras
Filologia	Texto (Língua)	Crítica textual, histórico-comparativo	Estudo do texto, contexto de produção e a língua
Linguística Histórica/ Diacrônica	Língua no tempo	Analogia, histórico-comparativo	Estudo das mudanças linguísticas

Fonte: Yamamoto (2015).

A partir do Quadro 1, é possível ver os *status* diferente de cada disciplina em contexto de língua portuguesa, e do ponto de vista da Linguística no Brasil. Não nos ativemos à escola anglófona por não fazer parte de nossos objetivos. As informações contidas no quadro são uma síntese da pesquisa bibliográfica feita com os autores referência das áreas da Etimologia (VIARO, 2011; DURKIN, 2009), Filologia Românica (IORDAN, 1982; VIDOS, 1996; BASSETO, 2001), Linguística Diacrônica (VIARO, 2014) e LH (TARALLO, 1990; MUSSALIM e BENTES, 2012; WEEDWOOD, 2002; FARACO, 2005). Cada uma dessas disciplinas adota objeto de estudo, método e objetivos diferentes, o que as torna autônomas em relação à LH. Apesar de os autores de cada disciplina apresentarem alguns pontos distintos quanto ao estatuto de suas áreas, a discussão sobre objeto, método e objetivo delas são coincidentes.

A conclusão a que chegamos é que, devido ao fato de a Filologia e a Etimologia contribuírem para a LH, às vezes, para um leigo, elas podem ser confundidas, devido à perspectiva de estudo diacrônico da língua. É importante salientar que, apesar de o Quadro 1 dar a impressão de que as fronteiras são bem delineadas, na verdade, elas só foram representadas dessa forma no intuito de facilitar a visualização para o leitor e servir como uma síntese do conteúdo investigado.

A partir da pesquisa bibliográfica, constatamos que os estudos filológicos e etimológicos servem de *corpus* para as análises linguísticas da LH. A partir desses estudos,

que abrangem aspectos sociológicos, históricos, geográficos, culturais e linguísticos, as mudanças e variações linguísticas são identificadas e os processos de alterações refletidos na língua são estudados, teorizados e registrados pela LH.

Se considerarmos a Filologia como subárea à parte nos dias atuais, podemos dizer que, futuramente, essa subárea da Linguística não estará limitada aos textos escritos somente. Ao contrário, poderá incluir os textos em forma audiovisual que a tecnologia permite que existam hoje e que tenham uma duração considerável, dentro da perspectiva de Mattos e Silva (2008), ao usar os termos textos modernos para inferi-los.

Entendemos com essa reflexão que tanto a Filologia quanto a LH são áreas de conhecimento que estudam línguas humanas, porém com objetos de estudo diferentes. A Filologia tem caráter mais abrangente no estudo de textos e serve como provedora de *corpus* de estudo para a LH. A LH como disciplina, por sua vez, aborda um aspecto linguístico analítico mais pontual nas línguas, as mudanças linguísticas, explicitadas por Mattos e Silva (2008). Além de conceituar a LH, a autora também propõe as subdivisões dessa disciplina que Faraco (2005) não menciona na obra analisada.

Nas subseções seguintes, o leitor poderá ver detalhadamente como esses autores delineam essas áreas e como elas se constituem dentro dos Estudos da Linguagem e da árvore de domínio da Linguística. Nelas apresentaremos informações relacionadas às subáreas da Etimologia, a primeira das disciplinas a surgir na linha do tempo, da Filologia Clássica e Românica, desde seus primórdios na Grécia até os estudos europeus de mudança linguística e a LH, posterior aos estudos da Filologia Clássica.

2.1.1 Etimologia, Filologia Românica, Linguística Diacrônica e/ou Linguística Histórica

Nesta seção buscamos expor os resultados finais da pesquisa bibliográfica sobre os estatutos das disciplinas que contribuem para a composição da LH como disciplina científica. Foram consideradas as áreas de: Etimologia, Filologia, Linguística Diacrônica ou Histórica.

Quanto à última nomenclatura, é importante salientar que há autores que preferem o uso da Linguística Histórica como disciplina, enquanto que, outros usam a Linguística Diacrônica como disciplina (VIARO, 2011, 2014) ou como abordagem de se tratar a língua (SAUSSURE, 2006, p. 106), em contraste com a sincrônica. Ou seja, há autores que veem

a LH como Diacrônica, sendo as duas sinônimas, enquanto que, para outros, a nomenclatura diacrônica será somente uma abordagem.

2.1.2 Etimologia

Nesta seção, apresentaremos a Etimologia a partir das perspectivas de Viaro, (2011) e Durkin, (2009), sendo o primeiro de um contexto de língua portuguesa e o segundo de língua inglesa.

Viaro (2011, p. 24) explica que a Etimologia se ocupa do código linguístico e sua origem, ou seja, as unidades linguísticas ou étimo. Nessa área da Linguística também se faz necessário diferenciar a Etimologia, disciplina ou ciência, da unidade terminológica etimologia, usada para significar o “estudo etimológico de uma palavra ou de um elemento de formação”.

Ainda segundo Viaro (2011), a Etimologia existe há aproximadamente 2.500 anos, com registros de Heráclito e os questionamentos sobre a semelhança lexical e as modificações sofridas pelo léxico; isso seria um embrião da noção de diacronia aplicada à linguagem. Platão também se preocupa com a etimologia, como em *Protágoras* 312c, *Fedro* 237a, *República* 396c, e em *Crátilo* de forma mais marcante.

Em *Crátilo*, descrito por Viaro (2011), questiona-se se o signo linguístico é fruto de uma convenção social, ou se o processo denominativo faz parte da identidade da palavra, como se trouxesse traços genéticos de sua essência impressos na forma verbal. Nessa discussão, para Sócrates, o processo de nominalização era a representação do essencial do objeto descrito, logo o aspecto imagético do signo poderia permanecer, independentemente da presença de todos os traços conceituais. Nessa obra, Sócrates se vale do método analítico dos étimos, por meio do qual a interpretação era resultado de uma retomada dos nomes primitivos e suas origens. Essa origem, o étimo, era buscada em signos foneticamente idênticos, unidos em uma composição hipotética. Neste grupo incluíam-se vários verbos e adjetivos que seriam a representação da essência do objeto. Para exemplificar essa hipótese, mencionamos o termo *sôma*, cuja origem teria sido *sêma*, que significa “túmulos” ou “sinal”. O problema deste fazer teórico residiu no fato de que o método não foi problematizado, resultando em uma postura mais dogmática quanto ao estudo do étimo.

O autor também explica que, já na Alta Idade Média, em idos do século VII d.C., na Espanha, Isidoro de Sevilha (c560-636) compõe as *Etymologiae*, obra enciclopédica de

vinte volumes, com o objetivo de informar o significado das palavras, incorrendo no erro entre o étimo e o significado. Isso se deu porque, naqueles dias, os conceitos de significado e étimo caminhavam juntos, nuance posteriormente elucidada por Thomás de Aquino (1225-1274). A conceituação de Etimologia segundo Isidoro era: “a origem dos vocábulos, deduzida dos verbos ou dos nomes por sua interpretação” (VIARO, 2011, p. 36). Em outros casos, o autor acreditava que havia palavras sem etimologia, pois algumas teriam sido fruto do processo de nominalização decorrente da vontade humana.

Na sua obra, *Summa Theologiae*, Aquino propõe que o significado de uma palavra e o termo usado para nominalização nem sempre são os mesmos. Consequentemente, encontramos nesse autor a dicotomia **significado** versus **etimologia** como noções distintas (VIARO, 2011).

Viaro (2011) explica que tanto Platão quanto Isidoro consideram a equivalência sonora parcial entre o étimo e as etimologias como diretriz em suas etimologias. Quase sempre, a Etimologia se reduzirá a uma questão de aproximação nesse período, com a aplicação das regras de metaplasmo (*latine transformatio*), regras de adição, subtração, transposição e transformação. Apesar da falha metodológica, Isidoro cumpre um relevante papel para a Filologia Românica, graças ao registro da língua falada da época, útil como *corpus* para estudos diacrônicos e filológicos.

A Etimologia atual parte dos *corpora*, do *terminus a quo*, como método etimológico para reconstruções dos étimos. Como exemplo, há a palavra **açúcar**, de étimo árabe, e de origem indiana. Essa palavra é proveniente do árabe *as-sukkar*, cuja origem é do sânscrito *çarkarā*. Partindo desta perspectiva, Etimologia é o “percurso entre o étimo, ou a origem, e a palavra investigada” (VIARO, 2011, p.106). Para que o étimo se estabeleça como tal, é necessário que haja *corpus* datado, método semelhante ao da Paleontologia e da Arqueologia. Em suma, a Etimologia moderna estuda as dicotomias oral vs. escrito e popular vs. culto; ela não é de caráter prescritivo, mas considera a fragmentação linguística em seus estudos, já que as línguas sofrem mudanças sociolinguísticas no espaço e no tempo. Isso também acontece na concepção da protolíngua, já que ela seria um construto e não um fato estabelecido.

Trazendo a contribuição de língua inglesa, Durkin (2009) propõe que a Etimologia traz à luz diferenças em sentido e como chegar a essas diferenças é o objeto desta ciência, ou seja, é a ciência que estuda a história das palavras. A Etimologia faz parte de um campo

mais amplo da pesquisa histórica linguística, com a intenção de explicar o como e o porquê de as línguas mudarem. Contudo, seu estudo não se limita a áreas específicas da língua como a fonologia, a morfologia, semântica ou sintaxe históricas; ela se vale desses aspectos, em conjunto ou em partes, para explicar o sentido das palavras. De fato, a Etimologia pode ser definida como a aplicação, no nível específico da palavra, de métodos e interpretações provenientes de diferentes áreas da LH para produzir uma justificativa coerente para a história dessa palavra.

Além de nomear esta subárea da LH, ela também pode significar, como um substantivo abstrato, a história de uma palavra. Para se considerar a etimologia de uma palavra, é importante levar em conta alguns conceitos, tais quais: investigação da história linear de uma palavra, a mudança lexical (forma), a mudança semântica, empréstimos, relações genéticas entre línguas, cognatos, reconstrução comparativa, e mudança de sons. Isto é, assim como podemos nos referir à Terminologia como ciência que estuda os termos ou como o grupo de termos de uma área específica, a Etimologia pode ser a subárea da Linguística Teórica ou Linguística Histórica, que estuda a origem do sentido da palavra, ou a origem de uma palavra em si, neste caso, etimologia grafada com letra minúscula (DURKIN, 2009).

Para exemplificar a metodologia da Etimologia, Durkin (2009) usa o exemplo da palavra *friar*, originada do latim *frāter* que transforma-se no francês antigo *frere* (no francês moderno *frère*), emprestado ao inglês médio como *frere*, que, por fim, se desenvolve em *friar* no inglês moderno.

O autor, então, explica que se pode traçar a história do som e forma das palavras por meio de regularidades, mudanças recorrentes, às quais foram submetidos os mesmos sons ou combinação de sons. Busca-se fazer paralelos e comparações, a partir do irregular ou do inesperado para serem explicados. O sentido da palavra também pode ser traçado historicamente. O autor exemplifica a mudança semântica na diferença de uso para *frair* e *brother*. Sendo o primeiro: membro da ordem religiosa que defende a pobreza, como os franciscanos, augustinos, dominicanos e carmelitas; em oposição ao segundo, usado para irmãos biológicos e religiosos de ordens não mendicantes.

Observamos que no fazer etimológico há uma identidade marcante no processo de reconstrução dos antecedentes linguísticos. Durkin (2009) menciona a possível reconstrução da língua inglesa do inglês atual ao inglês antigo (*Old English*) até às

inscrições rúnicas. O processo de reconstrução mapeia as grandes variantes do inglês como o inglês americano, o britânico e o sul-africano. Posteriormente, estes podem ser subdivididos em áreas administrativas ou geográficas de Londres, ou à língua de diferentes classes sociais em uma cidade ou grupos de faixas etárias, entre outros. O autor afirma que esta nuance é presente em toda a história do inglês. Assim como o latim sofria variações em si mesmo, antes de constituir as línguas neolatinas, assim também o germânico sofria variações internas antes de originar o inglês e outras línguas germânicas. Da mesma forma, diferentes grupos germânicos, em diferentes regiões geográficas desenvolveram diferentes comunidades linguísticas. Diante disso, algumas variantes linguísticas teriam predominado em certas comunidades específicas, enquanto que outras surgiam em cada comunidade de falantes.

Durkin (2009) propõe que muitas variações linguísticas europeias devem-se à política e aos espaços geográficos distintos nos quais elas existiram. Nos séculos XVI e XVII os escoceses já estavam a caminho de padronizar uma forma oficial, diferente do inglês da Inglaterra. Hoje, o neerlandês e o alemão são línguas bem definidas, o que limita a inteligibilidade entre falantes dessas línguas. Diferentemente, os moradores de regiões fronteiriças entre esses dois países, falantes de dialetos, podem entender-se mutuamente com facilidade, apesar de serem membros de comunidades linguísticas distintas: do alemão e do neerlandês. O autor explica que há um *continuum* dialetal entre a Alemanha e a Holanda, assim como há na fronteira da França e a Itália, e em outras regiões do mundo onde línguas distintas se desenvolvem a partir de uma mesma origem em territórios adjacentes. Quando duas comunidades linguísticas divergem, cada uma leva consigo características da língua mãe.

Além das línguas germânicas, o autor traz o conjunto de línguas originárias do indo-europeu, justificando que os traços da mesma origem ancestral dessas línguas residem nas evidências de correspondências fonéticas regulares e também nas similaridades de sistemas gramaticais.

Em seguida, o autor faz um resgate histórico da palavra *sad* do proto-germânico ao proto-indoeuropeu e os procedimentos etimológicos necessários pra tal. O autor parte da origem deste termo do protoindo-europeu ao grego, do latim ao germânico, ao inglês atual, explicando as mudanças linguísticas por meio das leis de mudanças fonéticas de Grimm e de Verner. A reconstrução comparativa dessas mudanças, segundo o autor,

depende da regularidade das relações e mudanças fonéticas, o que cria uma base sólida para a pesquisa etimológica comparativa.

Conforme Durkin (2009), a Etimologia é uma ferramenta essencial para reconstrução da história da língua, já que um *corpus* de histórias das palavras supre a base necessária para vários aspectos do trabalho de LH. Da mesma forma, a história individual de cada palavra depende de plausibilidade no trabalho realizado em várias subáreas da LH. Por exemplo, um estudioso da semântica histórica terá interesse no sentido individual das palavras como fruto de uma pesquisa etimológica. Cada subárea informa e enriquece as outras em uma relação mútua benéfica.

Tradicionalmente, a Etimologia tem sido mais proximadamente associada à construção de gramáticas históricas e comparada. Uma gramática histórica traça os desenvolvimentos em formas lexicais encontradas na história da língua, estendendo-se até sua pré-história. A gramática comparativa relaciona as mudanças encontradas em uma língua com aquelas ocorridas em línguas semelhantes, para explicar o desenvolvimento de duas ou mais línguas de origem comum com o uso da técnica da reconstrução comparativa (DURKIN, 2009).

Segundo o autor, às vezes “etimologia” é vista quase como sinônimo de “reconstrução comparativa”, ou ao menos, supõe-se tudo o que um etimólogo deve considerar é de importância secundária, se comparada com a reconstrução de formas lexicais antecedentes e sua identificação com mudanças fonéticas históricas. A reconstrução comparativa tem uma metodologia irmã, conhecida como reconstrução interna, na qual a reconstrução é baseada em dados provenientes de uma única língua. Ela é menos confiável que a reconstrução comparativa, apesar de ter contribuído para áreas como a linguística indo-europeia, na qual dados comparativos existem em abundância. Ela tende a ser mais eficaz na busca da origem de relações morfofonêmicas como no inglês *mouse* e *mice*, ou o contraste entre consoantes sonoras e surdas no alemão *Rad* e *Rades*.

A função chave da etimologia é a de elucidar as relações formais e semânticas entre as palavras de uma língua. Isso interessa aos falantes de línguas nas quais há uma grande quantidade de empréstimos e cujas relações semânticas como em *hand* e *manual* são obscurecidas pela ausência de qualquer relação formal entre as palavras. Nesse caso específico, a palavra *manual* é derivada da palavra que significa mão, porém a palavra em questão é do latim *manus* mais o sufixo -*ālis*, que forma o adjetivo *manualis*, emprestado

ao inglês no século XV. Por um tempo, aquele competiu com o termo *handy*, contudo hoje, esse termo tem o sentido especializado de “prático para uso”. Este exemplo ilustra a tendência que existe no estudo histórico das palavras, o que os etimologistas devem sempre ter em mente. Sempre haverá uma relação formal entre palavras que compartilham conexão semântica entre si (DURKIN, 2009).

Frequentemente, a história de um termo que surge de processos como o citado no parágrafo anterior, envolverá o trabalho de mais de um pesquisador. O trabalho do investigador atual sempre se volta para o que já foi investigado, analisa-se o que foi posto, e observa-se novos dados e descobertas, hipóteses sugeridas pela pesquisa anterior são reforçadas e confirmadas tais hipóteses, ou mesmo se propõe um novo olhar sobre elas (DURKIN, 2009).

Finalmente, o autor afirma que palavras fazem parte de um sistema, o léxico de uma língua, com várias relações com sua gramática. Qualquer mudança da nossa compreensão de uma parte desse sistema em outra, em partes que foram removidas há tempos, componentes de um sistema maior, e sempre devemos levar em conta tais implicações em nosso trabalho e no de outrem. Às vezes, uma mudança etimológica do presente pode abrir caminho para inúmeras soluções para problemas de outrora. Durkin propõe que se deve levar em conta a declaração de Meillet (1929, p. 26) de que a língua é um sistema no qual tudo é conectado. Ainda segundo o autor, não devemos esquecer que palavras e línguas são faladas por pessoas reais, vivendo em uma sociedade particular, em um determinado ponto na história, e que é no uso individual dos falantes que mudanças lexicais e semânticas surgem e se desenvolvem. Para a compreensão das palavras do passado, é necessária a imersão nesse material e na cultura intelectual, para que tracemos as conexões entre as palavras e conceitos que parecem pouco relacionados com o uso atual. Também devemos considerar os registros e estilos de língua, o vocabulário especializado de diferentes comunidades linguísticas. Quando tais fatores forem considerados, melhores etimologias serão produzidas e serão feitas descobertas sobre a história cultural e social de uma comunidade.

Comparados os dois autores, observamos a preocupação de Viaro (2011) em definir a área e fazer o levantamento do histórico da disciplina. Durkin (2009) também define a disciplina, mas tenta detalhar um pouco mais o método científico usado pela Etimologia. Viaro trata também do método etimológico, mas seu foco é mais centrado nos

metaplasmos, enquanto que Durkin, foca mais em explicar e exemplificar a reconstrução comparativa.

Na seção 2.1.3, apresentaremos a Filologia Clássica e a Filologia Românica, seus históricos e desenvolvimentos, bem como sua constituição.

2.1.3 Filologia e Filologia Românica

Basseto (2001) propõe que uma pesquisa histórico-bibliográfica extensa mostra que os termos **filólogo** e **filologar** precedem o termo **filologia** na historiografia greco-romana. A análise do termo na escrita grega mostra que há variações semânticas do termo. O autor propõe que não há univocidade nas obras fonte, apesar da contemporaneidade de autores. Posteriormente, o termo aparece em obras romanas e, no século VI, desaparece na literatura ocidental devido à tradição cristã que eliminava aquilo que não conseguia cristianizar. O termo reaparece posterior à Reforma Carolíngia e volta a se evidenciar a partir do séc. XV e XVI com José Justo Escalígero (1540-1653), Cláudio de Saumaise (1588-1653) e Isaac Casuabon (1559-1614).

Inicialmente na Grécia, explica Basseto (2001), em V a.C., o termo **filólogo** era presente na oralidade precedendo Platão e Aristóteles. O **filólogo** personificava-se como um falante e ouvinte, não como um profissional da língua escrita. O sentido do termo era: “aquele que ama e apreende as palavras, e delas extrai sabedoria”. Nesse contexto, o **filólogo** era mais que uma linguista, era aquele que dominava várias áreas do conhecimento tal qual um sábio, detentor de um conhecimento “enciclopédico”. Exemplos desses usos aparecem na Arte da Retórica (1398b), na qual Quílon (séc. V a.C.) é citado como sábio/**filólogo**.

A partir da escrita, o termo abrange a ideia de “amigo da palavra tanto falada e ouvida como escrita”, tendo uma variação posterior para “aquele que gosta de falar ou de aprender ouvindo”. Em Isócrates (436-338 a.C.), o termo **filologia** denota o “gosto pelo estudo da palavra” (Antidosis, XV, 296). Cícero (104-43 a.C.) usa o termo **filólogo** em grego e classifica suas últimas obras como “mais filológicas” que as primeiras. Em suas obras *Ad Familiares*, XVI, 21 e *Ad Atticum*, 11, 17, ele diz que, apesar de serem nobres, alguns homens eram faltos de intelectualidade para o ambiente acadêmico, não eram **filólogos** (BASSETO, 2001, p. 17-20).

Suetônio (c.69 – c.126) menciona Erastóstenes (275-194 a. C.) e Ateius Praetextatus como sábios, conhecedores de “todos os gêneros”, e considerava que o adjetivo que melhor os descrevia era “filólogos”. Posteriormente Sêneca mostra a distinção entre **filólogo** e **gramático** que era feita na sua época: o **gramático** tratava de questões tais quais problemas específicos da língua e de literatura enquanto o **filólogo** se ocupava de análises, interpretações de fatos, conhecimento histórico registrados em livros, tal qual faziam Ateius e Erastóstenes, explica Basseto (2001).

Já Sextus Empiricus (cerca de 200 d.C.) registra em *Contra os Matemáticos*, I, 235, a acepção que o termo **filólogo** indica “algo refinado, culto e estilizado no campo da linguagem como em Cícero” (*op. cit.*, p.23).

Apesar dessas acepções serem próximas em seus aspectos semânticos, elas não são unívocas. Basseto (2001), ao tratar do termo **filólogo**, diz que:

...a partir do significado etimológico de “amigo da palavra”, “amante do falar”, seu campo semântico se amplia bastante, passando a abranger tudo o que se refere ao ato da comunicação pela linguagem sob qualquer de suas formas. Nessa acepção abrangente se acomodam todas as variantes semânticas, até a atribuição do qualificativo aos sábios, “de múltipla e variada doutrina”, na expressão de Suetônio, para os quais a língua é mais um meio do que o objeto de estudo (o que é próprio do gramático) [...] (BASSETO, 2001, p, 24).

Suetônio comenta a obra de Cassius Longinus (205-269/70 d.C.) e como esse fizera a análise literária de Platão, então era **filólogo**. Nesse período, considerava-se **filólogo** o autor de análises e de críticas literárias, ação que pertencia ao campo de atuação dos “sábios” ou do “erudito”. Até o momento, a análise historiográfica não traz a acepção de **filólogo** como o profissional que faz a análise etimológica, semântica ou formal do léxico em um texto.

Contemporâneo ao estabelecimento e crescimento do Cristianismo, o termo torna-se mais raro e não é encontrado em Santo Agostinho (354 a 430), Anicius Manlius Severinus Boethius (480-583) e nem mesmo em Isidoro de Sevilha (602-634), com a obra *Etymologiae*, o termo **filólogo** ou **filologia** pode ser encontrado.

A **filologia** seria retomada nos séculos XV e XVI com os humanistas, envolvidos na pesquisa dos antigos, numa busca para compreender seus textos. O humanista Jácio César Escalígero é um exemplo de “sábio” ou “filólogo” de acordo com a visão grega ou latina. Observa-se, então, que **filólogo** volta a ser sinônimo dos intelectuais despontados

daquela época. Essa realidade pode ser observada em Guillaume Budé, conhecido como o Erasmus da França, ao redigir a obra *Philologia Libri II* em 1532. Além dessa obra, o autor redigiu obras em grego e latim. Durante esses séculos, as línguas neolatinas se consolidam e muitos estudiosos se dedicam aos estudos linguísticos. O termo **filólogo** passa, então, a ter a acepção de “pesquisador da ciência da linguagem e da literatura a partir de textos”, especialmente os antigos. Ele já não é sinônimo do profissional cujo perfil era de “múltiplos e variados conhecimentos” assim como Erastóstenes, Ateius e Longinus.

Apesar dos séculos XVII e XVIII serem prolíferos no que tange às obras linguísticas, explica Basseto (2001), principalmente na criação de gramáticas como a de Port-Royal, referências à Filologia são bastante escassas. O século XIX, com o conhecimento aprofundado do sânscrito, como nas cartas de Sasseti do século XVI, contribui para o crescimento dos estudos da linguagem e da **filologia**. Paris sediaria o centro de investigação do sânscrito em 1806, no Colégio da França, dirigido por Silvestre de Sacy. De lá, partem Humboldt e Franz Bopp, que aplicariam o método comparativo para analisar, comparar, classificar e estabelecer o parentesco entre as línguas por meio da tradução e comentário de textos. A esses estudiosos, dá-se o nome de **filólogos** naquela época.

Desse momento em diante, **filólogo** era o estudioso que associava os estudos histórico-comparativos das línguas à **filologia** no estudo da gramática e literatura, principalmente das línguas clássicas e das indo-europeias. Os que estudavam as línguas românicas também eram considerados **filólogos** como August Schlegel, com a obra *Observations sur la langue et la littérature provençales*, de 1818, similar ao trabalho desenvolvido por Grimm sobre os poemas medievais alemães. Friedrich Diez (1794-1876), **filólogo** alemão, fez uso do método histórico-comparativo às línguas românicas, da mesma forma como Bopp o usara com as línguas indo-europeias e Grimm com as línguas germânicas. Diez estudou obras castelhanas e do provençal e posteriormente dedicou-se às outras línguas românicas e, entre 1836 e 1843, publicou a *Grammatik der romanischen Sprachen* (Gramática das línguas românicas), onde mostra que as línguas românicas haviam se originado do latim falado e não do escrito; e, em 1854, o *Dicionário etimológico das línguas românicas*. Devido aos estudos desenvolvidos, ele é considerado o pai da Filologia Românica (BASSETO, 2001).

Sobre o século XIX e a distinção entre linguística e **filologia**, Basseto (2001) afirma que,

Em outros movimentos, correntes e teorias relativas à linguagem, que surgiram no fim do século XIX e início do XX, como a Teoria das Ondas de Johannes Schmidt (1843-1901), e a escola Idealista e Estética, de Karl Vossler (1872-1949), não se faz distinção entre filologia e linguística. Como “estudo científico da linguagem”, a linguística tomou grande impulso depois de Ferdinand de Saussure (1857-1913), considerado o pai da linguística moderna (BASSETO 2001, p. 33)

No século XX, no Curso de Linguística Geral (CLG, p. 7 e 8), ainda segundo Basseto, Saussure define o termo Filologia como área que busca “fixar, interpretar, comentar os textos”, que se ocupa “da história literária, dos costumes, das instituições, etc.” A língua e a história literária são listadas como **objeto** da Filologia, cujo método usado era a crítica. O CLG menciona que a pesquisa filológica prepararia o terreno para a LH.

2.2 Terminologia e Teoria Comunicativa da Terminologia (TCT)

A proposta de uma obra terminográfica demanda que as unidades terminológicas de uma área, objeto de estudo do terminólogo, sejam identificadas e delimitadas a partir de um *corpus* de referência. Estabelecida a área específica de estudo, o terminólogo pode extrair unidades terminológicas pertencentes àquela subárea específica, construir as definições e preocupar-se em como disponibilizar esses dados para os consulentes.

A Terminologia constitui-se da acepção dada à unidade terminológica no estudo científico do termo técnico-científico. É uma disciplina da Linguística que se dedica à pesquisa, desenvolvimento e tratamento dos termos como ramo da ciência. Outrossim, terminologia é o vocabulário especializado de um grupo específico de profissionais, pesquisadores que têm a necessidade de manter uma comunicação específica, atendendo ao princípio da univocidade.

Optamos por desenvolver este trabalho a partir da Teoria Comunicativa da Terminologia de Cabré (1999) por atender melhor o que acreditamos ser a concepção de um termo, qual seja, uma unidade lexical que pertence ao léxico de uma língua. Devido a aspectos pragmáticos, a unidade lexical passa a ter traços mais específicos e delimitados quando usada em contextos específicos, objetivando uma comunicação mais técnica. Quando isto ocorre, temos a unidade terminológica. Esta é uma visão que a TCT permite,

diferentemente da Teoria Geral da Terminologia, doravante TGT, de Wüster, na qual o termo só pode atender a uma comunicação específica, não servindo à comunicação fora de um contexto de uso da linguagem especializada.

Explicitando melhor a perspectiva da TCT, segundo Dornelles (2015), **unidade terminológica** é “uma unidade lexical da língua natural que adquire valor especializado dentro de um contexto especializado, segundo critérios semânticos, discursivos e pragmáticos” (p.13). Ou seja, dentro da perspectiva da TCT, a unidade lexical pode ser, ora uma unidade lexical, ora uma unidade terminológica. O que distingue um status do outro é o contexto no qual ela se insere, é o que Cabré (2011, p.149) denomina como caráter poliédrico do termo. Isto se dá pelo fato de o termo poder ser ao mesmo tempo uma unidade linguística, cognitiva e sociocomunicativa; linguística por ser proveniente do contexto real de uso da língua, a partir de qual a unidade lexical adquire um valor especial em relação às outras; cognitiva por promover a distinção entre o conhecimento especializado do geral ou não especializado; sociocomunicativa, pois consideram-se os aspectos pragmáticos que identificam a conformação e uso das unidades em determinadas situações comunicativas.

2.2.1 Terminologia Wüsteriana, a TCT de Cabré e a TST de Temmerman

Ao tratarmos da Terminologia, é necessário considerarmos seu histórico, bem como a concepção do que é seu objeto de estudo: a unidade terminológica. Os primórdios da Terminologia como disciplina científica são marcados pelos estudos de Wüster (1898-1977) na Áustria e pelo tratamento dado à unidade terminológica pelos russos como Lotte entre as décadas de 1930 e 1960. A partir do pensamento de Wüster, temos a Teoria Geral da Terminologia – TGT. Dentro da perspectiva deste autor, a unidade terminológica é concebida a partir da análise dos conceitos e é dissociada da “gramática, do contexto e do discurso” objetivando a univocidade técnico-científica internacional. Essa forma de conceber o signo terminológico difere da proposta de Saussure, que propõe que a forma e o conteúdo do signo linguístico são indissociáveis. De acordo com a proposta de Wüster, o signo assume um caráter metalinguístico, já que é concebido unicamente para referenciar os conceitos dentro de um domínio técnico e de caráter normativo (BARROS, p.55). A proposta de Wüster vê o termo como uma unidade denominativa e prescritiva, de cunho

onomasiológica, ou seja, criada a partir de um conceito, sem que se considere contextos de produção linguística ou mesmo a comunidade de usuários.

Com o desenvolvimento da Terminologia, observa-se que a unidade terminológica pode adquirir outro *status* além daquele de termo. É nesse contexto que surge a Teoria Comunicativa da Terminologia proposta por Cabré (1999) e pesquisadores da Universidade Pompeu Fabra, em Barcelona, em idos de 1990. A diferença entre esta teoria e aquela de Wüster é que o caráter comunicativo prevalece sobre o normativo e a abordagem do termo é mais linguística. Segundo a pesquisadora, a unidade terminológica não existe *per se*, mas se faz termo, considerado o contexto de discurso especializado no qual se insere. A definição proposta pela autora é que “o termo é uma unidade de forma e conteúdo indissociáveis entre si, que representam, no plano da verbalização ou expressão, um conceito” (CABRÉ, 2011, p. 129).³

A TCT é uma teoria terminológica descritiva, de base linguística, que valoriza o uso da unidade terminológica em seu uso real. É uma teoria que se insere nos Estudos do Léxico, logo o termo não perde seu status de signo linguístico, pertencente às línguas naturais. Dessa forma, as unidades terminológicas, por pertencerem a um sistema linguístico, fazem parte da gramática das línguas, mantendo suas propriedades de unidade linguísticas. O fato de conceber o termo dentro de uma sistema linguístico, implica em aceitar que o termo é passível de polissemia, homonímia e sinonímia.

Na dicotomia **conceito** (de valor universal, comum às línguas naturais) *versus* **significado** (atribuído a línguas particulares), a TCT concebe a existência de ambos no signo linguístico, diferentemente da TGT. Isto é, a unidade lexical pode ser tanto terminológica quanto lexical, pertencente ao discurso especializado ou ao de língua geral. O que define seu *status* é o contexto no qual se insere, dependendo de seu uso pragmático (KRIEGER; FINATTO, 2004). Um exemplo contextualizado desse princípio é o caso da unidade terminológica **língua** que, dependendo da subárea em que se insere, pode conter traços sêmicos diferentes dos de língua geral, distintos daqueles encontrados na Filologia, Etimologia, Linguística Diacrônica, Comparativa ou Histórica.

Além da TGT e a TCT, a Teoria Sociocognitiva da Terminologia de Rita Temmerman (2000), doravante TST, de base hermenêutica, surgiu também com a

³ “...el término es una unidad de forma y contenido indisociables entre sí que representan, en el plano de la verbalización o expresión, un concepto”.

concepção do termo como uma unidade linguístico-comunicacional, identificado pelo componente conceitual que mantém estreitas relações com a definição terminológica e a fraseologia especializada (neste caso o conceito não é o foco primordial). Esse enfoque faz parte da Terminologia linguístico-textual, adicionado-se a ele a abordagem cognitivista da ciência, baseada na análise da terminologia das ciências biológicas.

Essa escola, assim com a da TCT, questiona o objetivismo científico proposto pela TGT, na qual os termos eram unidades denominativas de um sistema de conceitos estruturado na lógica e numa ontologia fixas. Para a TST, os “termos são unidades de compreensão e de representação, funcionando em modelos cognitivos e culturais.” Para esta teoria, o conhecimento corresponde a um padrão modelado sócio-cognitivamente, cuja constituição baseia-se em módulos distintos que abrangem informações relativas a história, a categorias, até métodos e processos. (KRIEGER; FINATTO, 2004, p.).

O conceito do termo, a partir da perspectiva da TST, existe de forma subjetiva, pois a delimitação do conteúdo é embasada no texto que possui uma dimensão linguística, pragmática, discursiva e comunicativa. Consequentemente, o conceito e significado não são universais, mas passíveis de mudanças ocasionadas pela natureza linguística materializada em um texto multidimensional. Ou seja, fora de um texto não seria possível entender o conceito do termo, característica enfatizada pela linguística textual. Dentro dessa perspectiva, o termo, assim como na TCT, é concebido como uma unidade lexical inserido em um sistema linguístico, passível de homonímia, sinonímia e polissemia como uma palavra qualquer e também pelo fato de sempre estarem evoluindo. Ademais, o termo existe em um ambiente no qual se submete às condições morfológicas, fonológicas, sintáticas, semânticas ou pragmáticas das línguas. Além desses aspectos, Temmerman concebe como relevante o papel da metáfora para a constituição dos sentidos, o que para área da Terminologia é atípico, já que esta sempre privilegiou o sentido denotativo ao conotativo da linguagem.

Após análise das características da TCT expostas por Cabré (1999), nossa escolha para o desenvolvimento deste trabalho foi pela TCT, o que se deu pelas seguintes razões:

1. Poliedricidade do termo: a TCT postula que os termos integram aspectos linguísticos, cognitivos e sociais. Caso seja do interesse da pesquisa, é possível estudar somente um desses aspectos. **Linguístico** por pertencer a um sistema

de uma dada língua; **cognitivo** por representar conhecimentos específicos da ciência; e **social** por ser usado pela comunidade na interação social cotidiana.

2. Comunicação: a TCT concebe o termo como elemento de caráter comunicativo, que integra um sistema linguístico usado na interação social; ele não existe à parte, como elemento normatizador, restrito à comunicação especializada. Essa comunicação pode ser de dois tipos: **direta** ou **indireta**. **Direta** quando ocorre entre especialistas, num discurso didático, numa divulgação, ou **indireta** quando a comunicação é mediada pela tradução, por interpretação ou por meio de publicações especializadas. Essa comunicação representa o conhecimento especializado denominando conceitos pertencentes a essas áreas.
3. Variação: denominar a variação de um mesmo conceito é o que gera a sinonímia (termos diferentes para o mesmo conceito), a polissemia (mesma forma, mas conceitos diferentes). Cabré (1999) diz que, a variação pode ocorrer em três graus diferentes: máximo - na divulgação técnica e científica dos termos das áreas mais popularizadas; médio - na comunicação entre especialistas; e mínimo - quando se tratar da terminologia normatizada.
4. Linguagem natural: a Linguística reconhece a terminologia como parte do léxico da gramática de uma língua. A linguagem é especializada quando se consideram fatores temáticos, pragmáticos e semânticos. Desse ponto de vista, o termo é um constituinte da língua e do sistema no qual ele se insere, sendo usado em contextos diversos, cumprindo funções e transmitindo conceitos distintos. Da perspectiva da tradução, os termos são conjuntos de unidades de comunicação, avaliados por equivalência, adequação e economia. Do ponto de vista da linguagem de especialidades, os termos servem como meio de expressão e comunicação profissional e sistema de representação da estrutura do conhecimento das áreas especializadas.

O aspecto poliédrico do termo pode ser identificado neste trabalho nas diferentes formas de como o termo **língua** é definido em diferentes subáreas da Linguística dentro do

projeto VoTec⁴ (FROMM, 2007). Na área da LH, **língua** foi definida da seguinte forma: “produto social de uma comunidade linguística, que sofre mudanças devido a fatores sócio-histórico-culturais, usado para comunicação.” Na subárea de Aprendizagem de Língua Materna, a acepção para este termo é “sistema de signos que constitui instrumento de comunicação e de interação entre os falantes.” Apesar de o signo permanecer o mesmo em termos morfofonológicos, a acepção dada a ele é diferente; o aspecto comunicativo prevalece em ambas as áreas, mas as mudanças devidas a fatores sócio-histórico-culturais são restritas à área de LH.

O aspecto comunicativo dos termos é identificado nas comunicações diversas que usam a terminologia da Linguística e da Linguística Histórica. Apesar de seu aspecto metalinguístico, a terminologia da LH pode ser usada entre especialistas e discentes aprendizes de uma língua estrangeira. Esse uso objetiva a explicação de mudanças linguísticas que explicam a variação de **água** e *eau*, em francês, o que facilita a aquisição do léxico francês por um público lusófono. Esta comunicação, como definida no item 2 acima, seria do tipo indireta.

Observamos a variação quando definimos os termos *linguistic change* e *language change* em inglês, que podem ser tratados como sinônimos; a polissemia, ao analisarmos os conceitos diversos para o termo *language*, que em português pode ser traduzido por **língua** ou **linguagem**. Nesse caso, temos o mesmo termo *language* em inglês, porém com conceitos diferentes, dependente dos fatores temático, pragmático e semântico.

Do ponto de vista da linguagem natural, os termos selecionados fazem parte do léxico das línguas e em seus sistemas eles se inserem. Assim, os termos da LH são termos pertencentes à língua portuguesa e inglesa, que originalmente têm seus conceitos já estabelecidos e que adotam nova significação para representar os conceitos específicos da LH. Exemplificando este princípio temos o termo **analogia**, que no léxico das línguas portuguesa e inglesa significam: “relação ou ponto de semelhança, criado mentalmente, entre coisas ou seres diferentes⁵” e *a comparison of two things based on their being alike in some way*⁶. Contudo, num contexto de LH, teremos as seguintes definições “princípio de regularização de padrões linguísticos no qual semelhanças entre diferentes formas são

⁴ Disponível em < <http://pos.voteconline.com.br/> >. Acesso em 05 ago. 2015.

⁵ Disponível em < <http://www.aulete.com.br/analogia> >. Acesso em 05 ago. 2015.

⁶ Disponível em < <http://www.merriam-webster.com/dictionary/analogy> >. Acesso em 05 ago. 2015.

identificadas e as formas irregulares menos frequentes são reduzidas” e *mechanism based on learners' tendency to establish associations between regularities and irregularities in grammatical paradigms, leading to language change, simplification or optimization*. A partir desses exemplos, observamos que os traços conceituais de semelhança e comparação se mantêm, contudo, os traços de **redução de aspectos linguísticos irregulares** e de **mudança linguística, simplificação e otimização** se agregam ao conceito da unidade lexical, tornando-a em uma unidade terminológica, específica da área de LH.

Na seção 2.3 trataremos da árvore de domínio, sua definição e como a contextualizamos para esta pesquisa.

2.3 Árvore de Domínio

Primeiramente, é importante entender o conceito que embasa a árvore de domínio. Segundo Krieger e Finatto (2004), a árvore de domínio, “diagrama hierárquico composto por termos-chaves de uma especialidade, semelhante a um organograma” (2004, p.134), permite uma visão geral e ampla do objeto de estudo de uma especialidade, auxilia na compreensão das hierarquias básicas e situa o recorte terminológico do projeto em desenvolvimento. Ou seja, é uma forma alternativa de síntese do nosso objeto de estudo: as subáreas da Linguística e suas relações quanto às grandes áreas de Linguística teórica ou pura e a de Linguística Aplicada.

Fromm (2007, p.39) propõe que informações coletadas para a formação de um banco de dados sejam organizadas em uma árvore de pesquisa de forma que seja hierarquicamente organizada. Logo, temos a proposta de organização de informações em “campo, área, domínio, subdomínios e outros”. Essa organização busca demarcar o lugar que cada conceito ocupa dentro do sistema de uma forma gráfica. A vantagem dessa abordagem é a possível configuração do sistema conceitual, que oferece uma visão global e clara do conjunto a ser explorado.

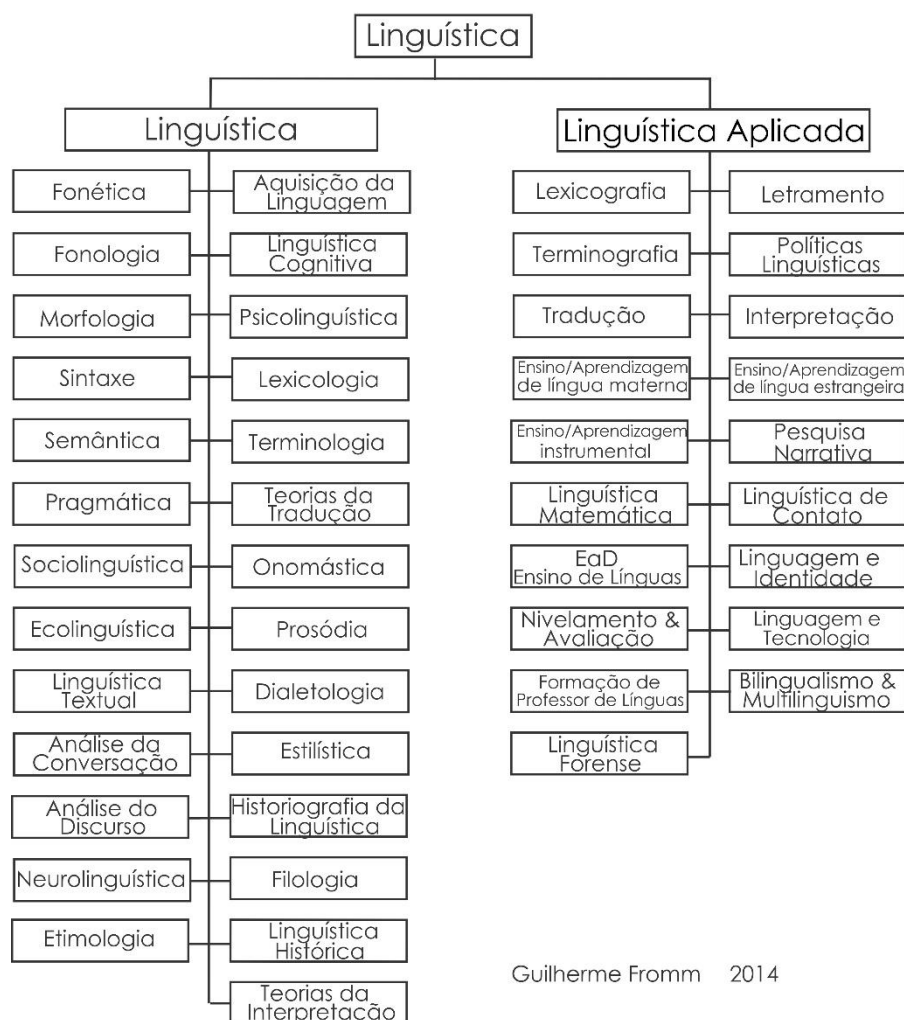
O desenvolvimento desta árvore se deu a partir de um projeto maior de Fromm (2013, 2015). Neste projeto, alunos da graduação e de mestrado desenvolveram a pesquisa das subáreas, coletaram *corpora* e organizaram os dados na plataforma do VoTec. Cada aluno ficou responsável por uma subárea em uma língua específica⁷. Essa árvore representa

⁷ Para maiores detalhes dos procedimentos teórico-metodológicos adotados para construção da Árvore de LLinguística, veja Fromm e Yamamoto (2013).

a Linguística no contexto brasileiro. Os alunos compilaram *corpora* especializados para cada área específica da Linguística, usados para a criação de um vocabulário bilíngue, português-inglês. Para todos esses trabalhos, a árvore da Linguística da academia brasileira e não a anglófona é utilizada para o desenvolvimento das subáreas. O público alvo desse vocabulário são os tradutores, os aprendizes de Tradução, os linguistas, docentes e discentes do curso de Letras e o público em geral que se interesse pela área.

A seguir, na Figura 2, temos a árvore da Linguística, subdividida em Linguística e Linguística Aplicada e suas subáreas como resultado parcial do projeto supracitado.

Árvore do Campo da Linguística



Guilherme Fromm 2014

Figura 2. Árvore do Campo da Linguística com as subáreas de estudo, reformulada a partir da árvore de 2013.

Fonte: Fromm; Yamamoto, 2013.

Observando a árvore da Figura 2, as áreas que serão abordadas neste trabalho pertencem a grande área da Linguística, subáreas da Filologia, Etimologia e Linguística Histórica. Na seção de Metodologia discutiremos como esses princípios serão aplicados em nosso trabalho.

Cremonese (2007) diz que a árvore de domínio é uma representação, um recurso metodológico que serve para auxiliar a equipe de elaboradores de uma obra terminográfica e seu usuário. Ela representa a área ou áreas que se busca representar por meio de diagramas, situa o campo de trabalho, a denominação e inter-relações entre as áreas ou subáreas de especialidade. O suporte que a árvore disponibiliza à equipe auxilia a concepção geral do trabalho e a elaboração de verbetes. Permite a concepção de uma visão geral do trabalho, promove a organização de dados, sistematiza as informações. A autora também menciona o fato de as áreas de especialidade se situarem em um espaço para depois transmitir.

A árvore possibilita a visão epistemológica de um campo, é uma forma de se descrever esquematicamente teorias e concepções de uma área, de um campo.

Considerando esses fatores, como parte do desenvolvimento deste trabalho, elencamos algumas áreas que contribuem para a formação do nosso banco de dados, úteis à construção de uma obra terminográfica na área de LH, a saber: a Filologia, Etimologia e a própria LH. Uma pergunta pertinente a ser feita é: se trabalharemos com a Linguística Histórica, como representaremos as subáreas de Filologia e Etimologia na árvore de domínios?

Refletindo sobre a representação da LH, da Filologia e da Etimologia na árvore de domínios da Linguística, segundo Mattos e Silva (2008), são a Filologia e a Etimologia as subáreas que suprirão a LH com os dados para aplicação de sua metodologia para, *a posteriori*, explicar os fatores que influenciaram a mudança Linguística no tempo e espaço. Logo, apesar de elas estarem separadas na árvore, há momentos em que a LH será bem próxima das duas. Na verdade, elas se misturam para subsidiar os resultados da LH. Isso se dá uma vez que o objeto de estudo da LH, as mudanças fônicas, mórficas, sintáticas e semântico-lexicais que ocorrem na língua, devem ser comprovadas pelo linguista histórico por meio de um *corpus* que comprove essas mudanças, *corpus* advindo da Filologia e Etimologia.

Na seção seguinte, discorreremos sobre a Linguística de Corpus, adotada neste trabalho como metodologia e abordagem para análise dos *corpora*, descreveremos o *corpus* do trabalho.

2.4 Linguística de *Corpus*

A LC tem sido reconhecida como uma ferramenta relevante para os estudos de LH, “dentro de espectro metodológico da linguística de *corpus*, a linguística de *corpus* histórica emergiu como um campo de investigação vibrante que tem adicionado interesse ao estudo da história e da mudança linguística.” (KYTÖ, 2011, p. 417).

O uso da LC neste trabalho terminográfico possibilita a descrição da língua em uso, isto é, a descrição das unidades terminológicas específicas, usadas por especialistas em contextos específicos, para comunicação específica de um dado grupo de informações. Consequentemente, a LC permite que nós o alinhemos à proposta da pesquisa e à identificação dos termos usados pelos especialistas em artigos, teses, dissertações e anais de congressos. A LC possibilita a análise quantitativa e qualitativa de dados linguísticos por meio do uso do WST com a criação da lista de palavras, a lista de palavras-chave e o concordanciador. O WST permite o processamento e análise dos dados que seria impossível se dependesse da leitura humana para o período desta pesquisa. Além disso, ela fornece resultados estatisticamente confiáveis e precisos dos termos usados pelos especialistas.

De acordo com a LC, os *corpora* podem ser classificados de diversas formas e a partir de perspectivas diferentes. Essa classificação ou tipologia muda com o tempo, já que a computação é uma área marcada pela dinamicidade e pela inovações frequentes. Abaixo segue a classificação do corpus de LH, escolhida para esta pesquisa, conforme proposta de Teixeira (2008).

Quadro 2. Taxonomia do *corpus* de LH (português-inglês)

Língua	Bilíngue (inglês e português)
Modo	Escrito (textos acadêmicos: artigos científicos, dissertações e teses)
Data de publicação	Sincrônico (levantamento realizado entre 2011 e 2013),
Seleção	Estático
Conteúdo	Especializado (Linguística Histórica, Filologia e Etimologia)
Autoria	Falantes nativos (inglês e português), individual/coletivo
Disposição Interna	Comparável
Uso na pesquisa	Estudo (análise terminológica/terminográfica)

Nível de Codificação	Com cabeçalhos, sem etiquetas
-----------------------------	-------------------------------

O Quadro 2 traz a classificação da tipologia do *corpus* que usamos neste trabalho, o que explicamos a seguir. Ele é um *corpus* **bilíngue**, nesse caso, por abarcar a língua portuguesa e inglesa; **escrito** contrastando com o oral; **sincrônico** - aquele que traz o registro de uma dada língua num momento específico da história, enquanto que o diacrônico cobre períodos diferentes dessa língua - os textos desses corpora abarcam o período de 1963 a 2011. **Estático** significa que o *corpus* não permite ser alterado, ou seja, a inclusão ou exclusão de textos não pode ser realizada. É **especializado**, já que objetiva a linguagem de especialidade da Etimologia, Filologia e Linguística Histórica, e contrasta com o de língua geral. A autoria é caracterizada por **falantes nativos** de língua portuguesa de português de Portugal ou do Brasil; e a autoria de língua inglesa são falantes de inglês britânico, australiano ou americano. É **comparável**, pois o *corpus* de português não é uma tradução do *corpus* de inglês e *vice-versa*. Tagnin (2015) o define da seguinte forma, “textos originais em duas (ou mais) línguas, numa determinada área de domínio.” (2015, p.26). Com relação ao aspecto de uso, o *corpus* é de **estudo**, pois será usado para estudo nesta pesquisa, quanto ao nível de codificação, **com cabeçalhos, sem etiquetas**.

Na próxima seção, trataremos da macro e microestrutura de uma obra terminográfica, traçando princípios que nortearão a construção de nossa obra.

2.5 Macroestrutura e microestrutura

Nesta seção será abordado o conceito de macro e microestrutura, explicaremos como estruturamos este dicionário, e também apresentaremos os princípios utilizados para construção da definição dos termos.

Para tratar da estrutura dupla de uma obra lexicográfica, Béjoint (2004) cita Rey-Debove (1971), que distingue a estrutura dupla do dicionário em macroestrutura e microestrutura. O autor afirma que a macroestrutura equivale ao inglês *word-list*, também chamada nomenclatura, e a define como o conjunto de entradas ordenadas, sempre submissas a uma leitura vertical parcial quando da localização do objeto da mensagem. Ainda segundo Béjoint, ela deve corresponder a um conjunto de unidades lexicais que existe, de certa forma, somente na mente do consulente.

Além de Béjoint (2004), Hartmann e James (2002) afirmam que a macroestrutura de um dicionário refere-se ao formato de acesso geral a um dicionário. Inicia-se com a

forma de organizar as entradas, em geral na forma alfabética. Contudo, os autores mencionam outras formas de organizar as entradas, tais quais: por temas ou tópicos, pela cronologia, pela frequência. A macroestrutura é constituída também pelo que o autor denomina de *outside matter*, como a *front matter*: prefácio, guia do usuário, página como dados bibliográficos, agradecimentos e dedicatória, lista de colaboradores, lista de abreviações e ilustrações, isto é, a parte introdutória de um dicionário; a *middle matter*: painéis, páginas ilustrativas, mapas, diagramas, lista de termos gramaticais ou campos semânticos, informações enciclopédicas e imagens; e a *back matter*: lista de nomes próprios, pesos e medidas, abreviações, hierarquia militar, tabela de elementos químicos, notas musicais, provérbios e ditados etc, ou seja, a parte final. No total, há o que pode ser chamado de **megaestrutura**.

No que tange à microestrutura, Hartmann e James (2002) abordam o formato dos verbetes, como sua informação é apresentada e quão apropriada a estrutura da definição é para o usuário. A microestrutura envolve dados tais quais ortografia, pronúncia, dados gramaticais, definição, etimologia. A microestrutura também é definida como a organização dos significados das entradas ou verbetes. Outra forma de defini-la é: a organização interna do verbete. Contrária à macroestrutura, a microestrutura provê informações minuciosas sobre o verbete, detalhando suas características semânticas e formais mencionadas anteriormente (ortografia, pronúncia etc). No caso de um verbete ter mais de uma acepção, definições são dadas para um deles de forma diferente.

Hartmann e James (2002) trazem um exemplo de como seria uma microestrutura, como mostra a Figura 3 abaixo.

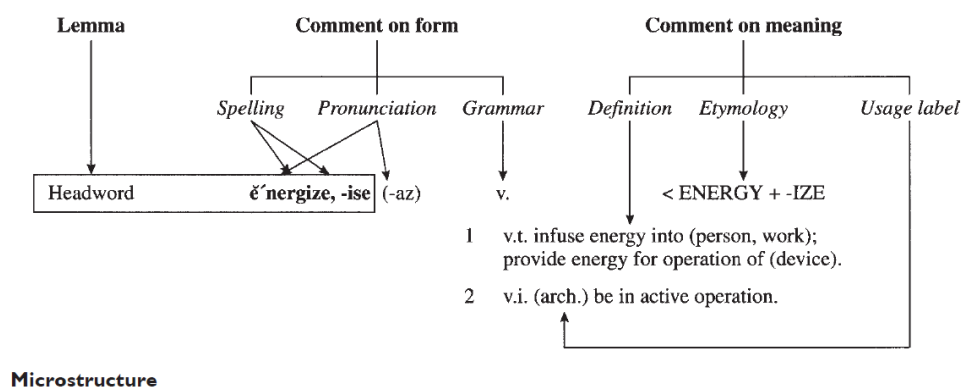


Figura 3. Microestrutura conforme proposta de Hartmann e James (2002).

Nesta estrutura, presente na Figura 3, observamos que existe o verbete, entrada ou lema (*headword/lemma*); há o paradigma que traz informações morfológicas, fonológicas e gramaticais (ortografia, pronúncia e gramática: verbo), e o paradigma definicional, no qual encontramos a definição, com acepções diferentes separadas por ponto e vírgula, a etimologia e o uso arcaico do termo.

Além da estrutura proposta acima por Hartmann e James (2002), os autores afirmam que em uma obra terminográfica, a microestrutura pode incluir a entrada ou verbete, o conceito designado pelo termo, o primeiro uso documentado, normalmente com a definição. Ademais pode haver as relações de conceito e termo.

A microestrutura, segundo Béjoint (2004), é a estrutura disposta na horizontal de uma obra lexicográfica ou terminográfica, *grosso modo*. Ela pode incluir a classificação gramatical do termo, a definição, informações enciclopédicas, exemplos. Este modelo, como proposto, também pode ser nomeado paradigma definicional. Em um dicionário ou vocabulário, procura-se manter um modelo recorrente dessa estrutura com o intuito de facilitar o manuseio pelo usuário, para que haja padronização, e também como traço identitário da obra. Essa estrutura é fixa, no sentido de que os dicionários modernos deveriam todos trazer a uniformidade de suas entradas, tanto em termos de conteúdo quanto de formatação.

Se compararmos as duas estruturas propostas por Béjoint (2004) e Hartmann e James (2002), observamos que a macroestrutura é mais flexível, ou seja, é possível adicionar ou subtrair uma entrada de um dicionário sem destruir a qualidade da macroestrutura. Da mesma forma, é impossível imaginar um dicionário sem a macroestrutura, de forma que se ela não for concebida, obviamente a obra não é um dicionário. Outras obras podem ter a estrutura dupla, como um guia ou manual e uma lista telefônica. O que distingue essas obras de um dicionário é a interação existente entre as duas estruturas: todas as entradas que fazem parte da macroestrutura recebem uma microestrutura e todas as palavras usadas na microestrutura devem, em geral, ser incluídas e tratadas na macroestrutura, ou seja, o dicionário é uma estrutura “fechada” (BÉJOINT 2004, p. 12,13).

Como parte da microestrutura, há a definição. Para que a definição seja redigida, é necessário que o terminólogo analise os contextos nos quais os traços semânticos ou distintivos se inserem. No início dos estudos terminológicos, o cientista partia dos

contextos para chegar ao termo, o referente, também chamado de **percurso onomasiológico** (BARROS, 2004). Com o advento do computador e com o uso de ferramentas computacionais para a análise de grandes *corpora*, é possível partir de uma lista de palavras-chave, verificar o contexto no qual elas são usadas e, a partir de então, criar-se as definições; no caso de o pesquisador partir da unidade terminológica para a identificação dos traços distintivos que o constituem, temos o **percurso semasiológico** (BARROS, 2004). Contextualizando o processo semasiológico neste trabalho, podemos dizer que as listas de palavra-chave produzidas pelo WST, permitiu que pudéssemos identificar os termos. Em seguida, pudemos identificar e analisar os contextos nos quais esses se inseriram para a identificação dos traços contextuais. A extração desses traços conceituais, organizados nas fichas terminológicas levam à construção da definição terminológica final.

O percurso onomasiológico pode ser observado quando tivemos que diferenciar o significado do termo *language* do inglês para o português e para selecionar os exemplos que se adequavam a cada definição. A diferenciação do inglês para o português, fez-se necessária, pois o termo *language* poderia ser traduzido por **língua** ou **linguagem**. Se a análise dos traços conceituais não equivalassem ao termo adequado, poderia haver uma equivalência de conceitos errônea, e consequentemente a uma denominação equivocada. Logo, foi necessário analisar primeiramente os conceitos para, *a posteriori*, chegarmos ao termo.

Quanto ao *layout* de uma obra terminográfica, para que haja a organização dos dados usados para a construção das definições, temos as estruturas denominadas macro e microestrutura. A categorização de macro e microestrutura pertence, originalmente, à Lexicografia. Como nosso trabalho situa-se na área de Terminografia, partes da macroestrutura, tais como prefácio, página com dados bibliográficos, agradecimentos, dedicatória, painéis, mapas ou diagramas, listas de nomes próprios, lista de pesos e medidas etc, comum a um trabalho lexicográfico, não se faz presente. Outra razão para a inexistência dessas estruturas é que elas não faziam parte da constituição da plataforma do VoTec, quando de sua concepção (FROMM, 2007). Como este projeto foi desenvolvido com uma extensão da obra do VoTec, ele também não traz essas partes como constituintes.

A diferença entre a macro e microestrutura existe por uma questão epistemológica entre as áreas da Lexicografia e Terminografia. Uma obra terminográfica objetiva fazer um

recorte do léxico de uma língua, focando uma área de especialidade. O usuário de uma obra terminográfica tem em mente o elucidar de dúvidas conceituais específicas, enquanto que o usuário de uma obra lexicográfica, em geral, busca descobrir o significado de uma unidade lexical, desconhecida para ele. Tratando-se da microestrutura, para cada conceito temos uma entrada diferente, pois definições diferentes implicam em “conceitos diferentes, individualizados, válidos em determinadas situações e não em outras” (FINATTO; 2006, p. 50)

O próximo passo da construção da microestrutura é a redação das definições. Segundo Auger e Rousseau (1987, p. 28-29), alguns princípios devem ser observados quando da sua construção:

1. Descrever o conteúdo nocional das unidades de maneira apropriada, precisa, clara, sucinta, dando relevo aos traços de significação próprios do termo;

Na proposta do VoTec (FROMM, 2007), ambiente de gestão terminológica, disponível *on-line*, de acesso público gratuito, os traços de significação são organizados em colunas na ficha terminológica virtual. Esses traços são extraídos dos contextos localizados a partir das linhas de concordância do *Concordancer*.

2. A definição pode ser anotada diretamente da ficha do terminólogo; pode-se anotar diversas definições de um termo, para selecionar, [...] , a mais completa e apropriada, dentro dos limites do campo de trabalho;
3. a nova definição a ser redigida será uma síntese e uma crítica àquela já existente

Tomamos este passo preenchendo as fichas de conceito final e definição, sendo esta última a definitiva.

4. A definição há de reunir todos os traços semânticos necessários para dar, com precisão, o sentido do termo. Assim a definição se apresenta sob a forma de indicações sobre a função de um objeto, a descrição de um procedimento, o funcionamento de uma máquina ou aparelho. Cabe, de qualquer modo, limitar a extensão da definição aos elementos necessários para a compreensão do sentido do termo por parte dos usuários [...].

Em nosso trabalho, esses princípios se aplicam parcialmente, pois a maioria das definições é de caráter abstrato, ligadas ao campo da LH.

5. Em terminologia a definição se situa obrigatoriamente dentro de um campo de atividades: será menor o referencial, isto é, corresponderá a uma única realidade, bem determinada. Na formulação da definição deve-se ter em conta o nível dos usuários: operários, técnicos ou especialistas.

O campo de atividades que serve como delimitador deste trabalho é a LH. Logo, os termos que foram definidos são aqueles dessa área, cujos consulentes serão pesquisadores, docentes, discentes e público interessado na área de LH.

O paradigma definicional adotado neste trabalho é o de gênero próximo e diferença específica - GPDE, adotado pelo projeto da plataforma do VoTec e pelo GETerm (ALMEIDA; SOUZA; PINO, 2007). Este padrão implica em partir de conceitos mais abrangentes - a hiperonímia, para conceitos mais específicos - o de hiponímia, no qual se faz a diferenciação conceitual de processos denominativos distintos. Este passo é desafiador, pois há conceitos que são presentes em mais de uma subárea da LH, como, por exemplo, o conceito de metaplasmos, presente na Etimologia, Filologia e LH. Outro exemplo é o conceito de língua, que além de ser distinto nessas subáreas, é distinto em toda a Linguística como ciência da linguagem.

Além desse padrão, outro padrão adotado foi o LIDE, proveniente da área do jornalismo. O LIDE, segundo Callado (2008), “é o primeiro parágrafo de uma notícia e deve narrar, resumidamente, o fato mais relevante da série de fatos que compõem a notícia.”(2008, p. 46). O lide deve responder as seguintes perguntas: o que, quem, como, quando, onde e por que. Obviamente, nem sempre é possível contruir uma definição que inclua todas essas informações. Contudo, ao organizar os traços conceituais, é possível identificarmos alguns desses elementos que ajudam a compor as informações básicas que serão disponibilizadas aos usuários. Para exemplificar a aplicação desse princípio, usaremos a unidade sintagmática **língua românica** com as perguntas em parênteses. **Língua românica** foi definida da seguinte forma: “língua (o quê?) proveniente (do quê?) do latim, que sofreu (o quê?) mudanças devido a (a quê?) fatores (quais?) geográficos, histórico-sociais e políticos, podendo ser estudadas (como?) de forma retrospectiva e prospectiva na língua escrita. Outro exemplo é o termo **linguagem**, definido assim: sistema (o quê?) de (de quê?) sinais, (como?) escrito ou falado, usado (por quem?) pela humanidade para (para quê) comunicação de ideias, marcado (pelo o quê?) pela variabilidade, considerado (o quê?) uma ciência moral e histórica.

Em inglês, os mesmos exemplos são: *Romance Language* e *Language2*. Sendo que a unidade sintagmática foi definida da seguinte forma: *Romance Language - language system* (o quê?) *developed from* (do quê?) *Late Latin in* (onde?) *communities related to* (a quê?) *politics and linguistic areas that were* (como?) *in contact*. O segundo exemplo foi: *Language2*⁸ - *system* (o quê) *used* (para quê?) *to produce meaning, reference, naming and used by* (por quem?) *different groups of people for* (para quê?) *communication*.

Ao escolher princípios atípicos à Terminologia, objetivamos levar em consideração a relação autor-leitor de uma obra terminográfica, no sentido de disponibilizar o máximo de informação possível ao usuário em um único paradigma definicional. O paradigma definicional seria composto de um oração e o mais completo possível, a partir dos padrões de LIDE, do *corpus* analisado e dos traços conceituais provenientes dos contextos definitórios e explicativos.

Na seção de Metodologia, buscaremos apresentar os passos adotados para a construção do Vocabulário de LH, detalhando a organização da Árvore de Domínio, a compilação e tratamento dos *corpora* pelo WST, a escolha e validação dos termos, e a inserção de dados na plataforma do VoTec.

⁸ O termo foi registrado como *Language2* para linguagem, pois o termo *Language*, sem numeral, fora usado para o termo língua, caso de polissemia neste trabalho.

3. METODOLOGIA

Nesta seção, serão apresentados os procedimentos da metodologia, adotados para a concretização deste trabalho, desde a definição da árvore de domínio até à disponibilização das definições finais no VoTec para os consulentes.

3.1 Árvore de domínio

Conforme proposto por Barros (2004), a árvore de domínio permite ao leitor uma visão mais heurística do objeto de estudo em questão. Por esta razão, trouxemos aqui um recorte da árvore de domínio da Linguística, subáreas da Filologia, Etimologia e Linguística Histórica para promover uma reflexão metodológica contextualizada para esta pesquisa conforme a Figura 4, abaixo.

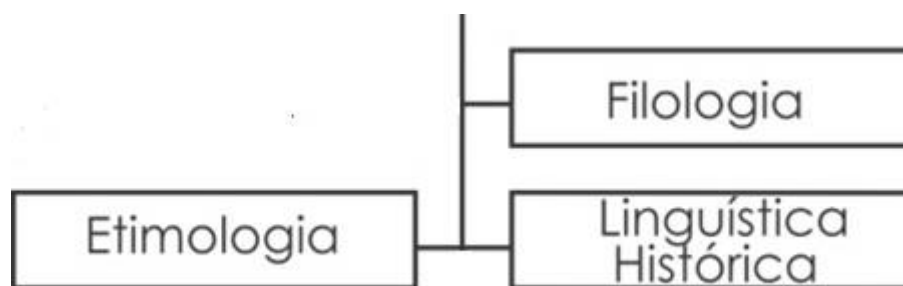


Figura 4. Recorte da árvore de domínio da Linguística
Fonte: Fromm; Yamamoto, 2013.

No contexto de língua portuguesa do Brasil, as áreas de Etimologia, Filologia e Linguística Histórica são subáreas da grande área da Linguística. É a partir desse contexto que desenvolveremos nosso trabalho de pesquisa. Ou seja, apesar de ser uma obra terminológica bilíngue, a base de nosso trabalho será de língua portuguesa, seguindo as propostas da academia brasileira.

Inicialmente, nosso objetivo foi o de desenvolver um trabalho terminológico na área de LH. Como passo necessário desse processo, após o estudo da área da LH, entendemos que essa área vale-se de outras para se estabelecer, isto é, na busca de estabelecer as mudanças de uma língua a LH vale-se das contribuições da Etimologia e da Filologia Clássica e Românica em contextos de língua portuguesa. O questionamento que buscamos responder foi: essas áreas se estabelecem como áreas ou estão incluídas na Linguística Histórica como um todo? Teria havido alguma mudança metodológica ou

epistemológica que uniria a Etimologia e as Filologias sob o escopo da LH, tornando aquelas áreas uma só com a LH?

Estudada a bibliografia da área da Etimologia com Viaro (2011) e Durkin (2009), entendemos que ela se estabelece como subárea da Linguística, apesar de estar, de certa forma, apagada nos dias atuais. Como exemplo desta realidade, podemos citar a falta de eventos na área de Etimologia, ou mesmo de uma bibliografia específica da área.

Além da Etimologia, pesquisamos a área da Filologia/Filologia Românica com Iordan (1982), Vidos (1996) e Basseto (2001). Essa área também se mostrou independente, apesar de que também já teve menos destaque; no momento tem crescido mais que a Etimologia, fato este visível na quantidade de congressos e simpósios dessa área. Não faz parte do objetivo desse trabalho essa mesma pesquisa no contexto anglófono.

Conforme apresentamos no Quadro 1, cada uma delas tem um objeto de estudo, método e objetivos distintos. Contudo, devido ao fato de a Etimologia estudar o étimo das palavras, ela o faz a partir de uma perspectiva diacrônica, o que implica em mudanças morfológicas e fonéticas, mudanças essas que interessam à LH.

Além do estudo do étimo, observa-se que estas mudanças de sentido e forma relacionadas aos étimos se dão no interior do texto. O estudo do texto, da cultura que o sustenta, o momento histórico e social no qual se insere, é objeto de estudo da Filologia. Como o texto e o léxico servem de base para as mudanças linguísticas, essas áreas contribuem para a LH, conforme explicitado por Mattos e Silva (2008).

A partir das primeiras pesquisas, a hipótese seria a de que a LH, pelo fato de ser uma subárea mais recente da Linguística, abarcaria a Filologia Clássica e Românica, bem como a Etimologia. Caso essa hipótese se comprovasse, teríamos a árvore da Linguística com a alteração apresentada na Figura 5.

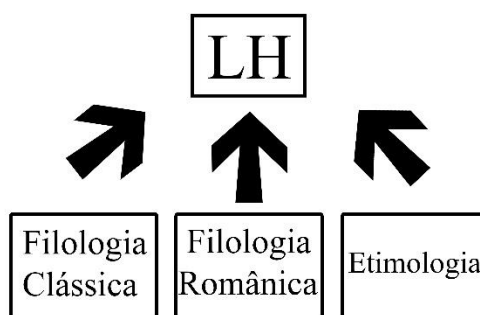


Figura 5. Árvore de domínio inicial - proposta para a LH
Fonte: autor.

Na proposta apresentada na Figura 5, trabalharíamos com a hipótese de que a Etimologia e as Filologias Clássica e Românica se aglutinariam para formar o que hoje temos como LH, sendo esta a mais recente dessas ciências. Mas após a pesquisa bibliográfica, o que tivemos foi outro resultado. Cada uma delas tem seus parâmetros e epistemologia bem definidos, apesar de serem disciplinas que estudam as línguas em uma perspectiva diacrônica.

Desse modo, apesar de não ser possível aglutinar a Filologia e a Etimologia sob o escopo da LH, tornando-as uma área única simplesmente pelo fato de elas se relacionarem, uma não substitui a outra. Então, propomos manter a árvore subdividida como já proposto por Fromm e Yamamoto (2013), sabendo que a Etimologia e Filologia se inserem na LH, mas o contrário não se dá da mesma maneira. O que queremos defender é que a Filologia e a Etimologia continuam independentes como subáreas da Linguística. Elas se estabelecem com seus objetos, métodos e objetivos distintos. O resultado do trabalho filológico e etimológico servem à LH, mas isso não quer dizer que elas se inserem na LH e se misturem a ela, no sentido de se tornarem uma “nova” subárea, juntamente com a LH.

Além da pesquisa bibliográfica, observamos no *corpus* de estudo que essas áreas também se mostram distintas, de forma que o caráter conteudístico de cada uma se difere, sendo que podemos agrupá-las sob a LH, no que tange a conceitos terminológicos, mas não como áreas que se juntam para tornar-se uma só. De uma forma mais detalhada, entendemos que os conceitos presentes nos estudos etimológicos: a mudança conceitual dos étimos, os princípios do método gramático-histórico usados na datação de *corpus*, são úteis e vitais à LH. Da mesma maneira, a crítica textual, com seu método histórico-comparativo, seu enfoque no contexto sócio-histórico e geográfico de produção da língua, contribuem para a compreensão das mudanças linguísticas em uma perspectiva diacrônica, estudadas pela LH. Finalmente, propomos a autonomia de cada subárea da Linguística, já que a árvore de domínio mantém-se igual a proposta anteriormente.

Concluimos nesta seção a proposta que acreditamos ser a melhor a se manter para os estudos da LH, conforme recorte da árvore da Linguística. Na seção seguinte, trataremos dos procedimentos metodológicos adotados para a coleta dos *corpora* de LH em português e inglês.

3.2 Compilação dos *corpora*

Os *corpora* utilizados para esta obra terminográfica foram compilados na área da Linguística Teórica, subáreas da Filologia, Etimologia e LH. O *corpus* do português compõe-se de teses, dissertações e artigos científicos, provenientes de *sites* de LH e Filologia, de acesso público. Dentre eles, citamos os *sites* da revista *Philologus*, disponível em <<http://filologia.org.br/>>, e *Filologia e Linguística Portuguesa*, disponível em <<http://www.fflch.usp.br/dlcv/lport/flp/>> para o *corpus* de língua portuguesa. Quanto às teses e dissertações, usamos a ferramenta de busca do Google, digitando a grande área, asterisco, teses ou dissertações, dois pontos, pdf. Dessa forma, focamos nossa busca em arquivos de cunho científico.

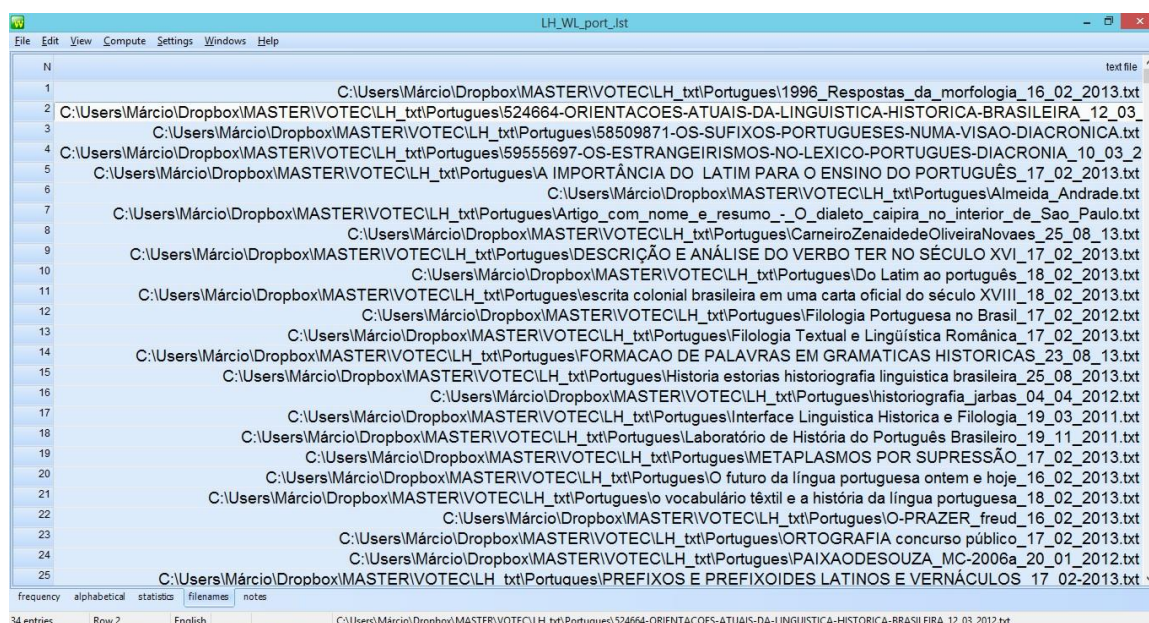


Figura 6. Visão parcial dos arquivos do *corpus* em português de LH.

Fonte: autor.

Analisando a visão parcial de arquivos da Figura 6, observamos a presença de vários documentos, sendo a maioria deles artigos, depois teses e dissertações. Esse *corpus* foi coletado do fim de 2011 até o segundo semestre de 2013. Vale lembrar que todos são de cunho científico, primeiramente coletados em formato PDF, posteriormente salvos em formato de texto (.txt), para serem lidos pelo WST. O formato PDF permite pré-selecionar arquivos de cunho científico, já que esse formato impede a alteração de dados e o bloqueio do arquivo, caso seja opção do autor.

O *corpus* de inglês é formado por anais de eventos, artigos, teses, dissertações e um dicionário enciclopédico de Etimologia (devido à escassez de artigos/dissertações e teses nessa área). Os arquivos são de acesso público, baixados em formato pdf e salvos em formato txt a partir de *sites* específicos da internet, assim como os de português. No caso da língua inglesa, usamos a base de dados *Jstor*, disponível em: < <http://www.jstor.org/> > e o site *Library Genesis*, disponível em < www.libgen.org >. Ressaltamos que o primeiro disponibiliza artigos ao ser acessado de um servidor público federal devido a acordos interinstitucionais. A desvantagem oferecida por esta base é que grande parte dos arquivos são obsoletos. Diferentemente do primeiro, o *site Library Genesis* é um *site* que oferece uma bibliografia acadêmica de acesso gratuito. Há obras completas como livro e anais de congressos, o que nos serviu para composição do *corpus* em inglês.

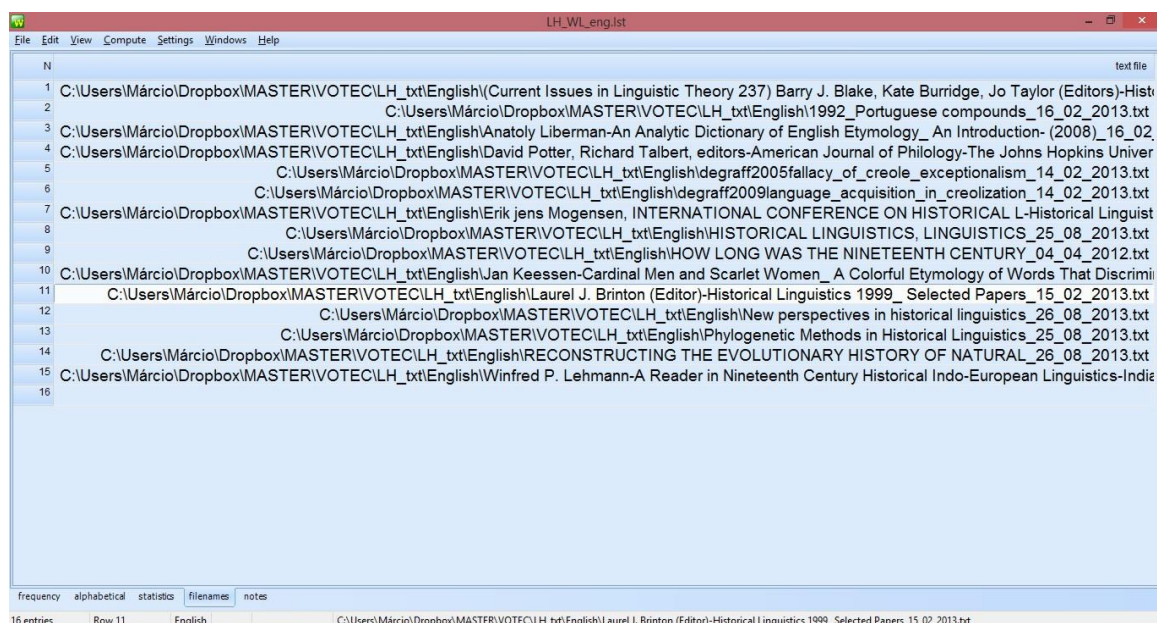


Figura 7. Visão parcial dos arquivos de *corpus* em inglês de LH.

Fonte: autor.

Como pode ser observado na Figura 7, temos um grupo de arquivos composto de anais de congressos que discutem as perspectivas atuais para a Linguística Histórica, além de teses e dissertações na área da LH. Devido à ausência de artigos científicos na área de Etimologia, que fossem de acesso gratuito, excepcionalmente, para que esta área não ficasse sem ser contemplada, incluímos um dicionário de Etimologia, *An Analytical Dictionary of English Etymology – an Introduction*. O dicionário traz informações enciclopédicas sobre termos diversos da língua inglesa, discute as várias etimologias e

propõe a que seria a mais pertinente, de acordo com a aspectos metodológicos da Etimologia⁹

Os arquivos foram salvos em formato txt para possibilitar a leitura pelo WST. De acordo com a classificação proposta por Teixeira (2008), esse *corpus* classifica-se como: bilíngue, escrito, sincrônico, estático, especializado, de falantes nativos, de autoria individual/coletiva, comparável, de estudo (uso na pesquisa) e com cabeçalhos.

Abaixo, no Quadro 3, temos o dimensionamento dos *corpora* e suas especificações.

Quadro 3. Tamanho dos *corpora* de LH.

Língua	Nº palavras	Nº textos
Português	536.330	33
Inglês	521.794	8

Fonte: autor.

Observando e analisando os dados do Quadro 3, quanto ao balanceamento dos *corpora*, temos um número de palavras equilibrado, contudo o mesmo não ocorre quanto ao número de textos. Isso se deu devido ao fato que, dentre os arquivos em inglês, houve anais de encontros internacionais em Linguística Histórica (1999, 2001, 2003), mais extensos que os outros gêneros, que agregavam vários artigos nesta área. Outra razão que justifica este baixo número de artigos é a dificuldade de acesso a artigos científicos em inglês, de acesso gratuito na Internet.

Em contraposição, temos o *corpus* do português, composto em sua grande maioria por artigos, teses e dissertações. Nesse *corpus*, os arquivos maiores são constituídos de teses enquanto que os outros eram compostos pelos gêneros restantes. Observamos que no Brasil há uma maior facilidade de acesso a artigos científicos gratuitos. Pelo que observamos, pode ser que isto se justifique pelo interesse governamental em popularizar o conhecimento, logo há uma maior quantidade de artigos disponíveis.

Após a escolha dos artigos, fez-se a leitura dos resumos para a certificação de que tais produções científicas realmente se adequavam ao objetivo do projeto: coletar arquivos

⁹ Para conhecer melhor a obra e o trabalho desenvolvido pelo autor, acesse < <https://muse.jhu.edu/books/9780816654086> > e < http://www.academia.edu/3191945/Review_of_Liberman_An_Analytical_Dictionary_of_English_Etymology_2008 >.

nas áreas de Filologia, Etimologia, Linguística Histórica que servissem de *corpora* para análise linguística e fonte de contextos para construção de definições terminológicas.

Para assegurar o tamanho dos *corpora*, fizemos a leitura com o WST para verificação do número de palavras no intuito de balancear os *corpora*, chegando ao número aproximado de 500 mil palavras em cada *subcorpus*.

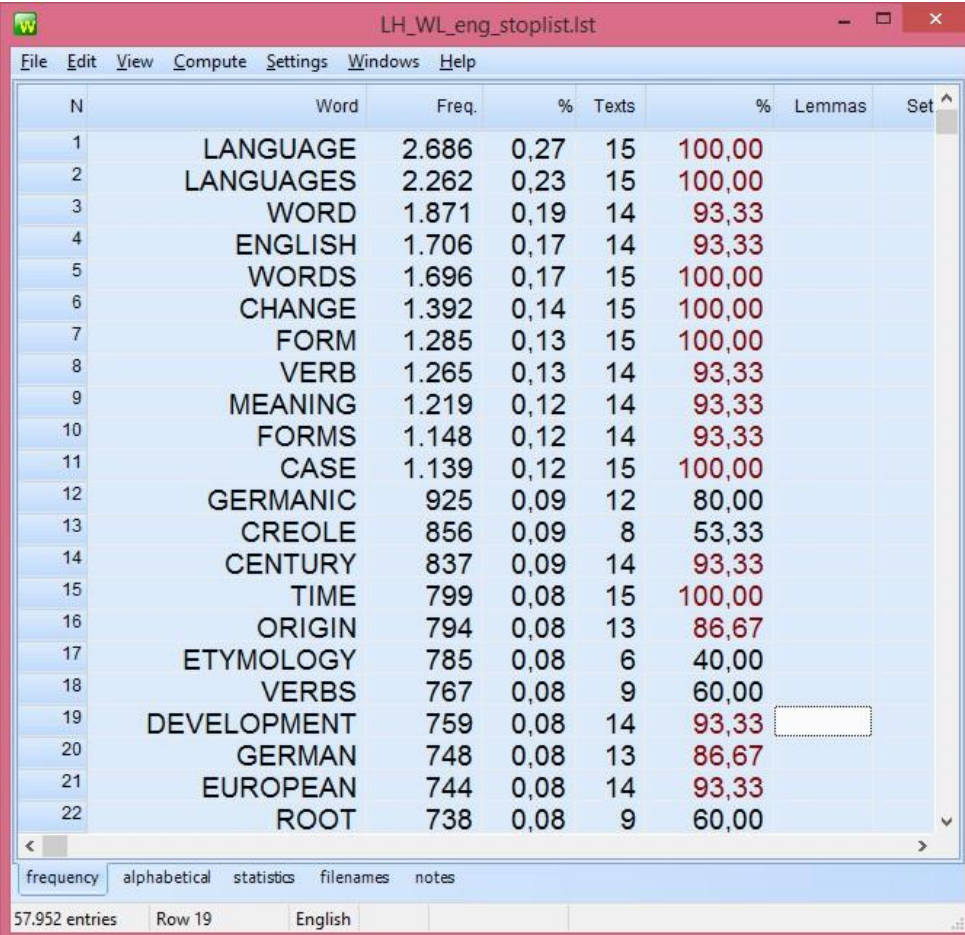
Na próxima seção, explicaremos como usamos o console do WST para fazermos as listas de palavras e as listas de palavras-chave dos *corpora* de português e inglês.

3.3 WST - Lista de Palavras e Lista de palavras-chave

Para o tratamento de grandes coleções de textos, a Linguística de *Corpus* vale-se de consoles tais quais o *AntConc* (ANTHONY, 2007) e o *WordSmith Tools* (SCOTT, 2012). Neste trabalho, optamos pelo programa WST, em virtude de uma maior familiaridade com o mesmo e pelo fato de ele permitir que os dados processados fossem salvos, o que o difere do *AntConc*.

Explicamos, então, que o segundo passo tomado para o tratamento dos *corpora* foi a utilização de três ferramentas do console do WST: o gerador de listas de palavras, o extrator de palavras-chave e o concordanciador (VIANA, 2010, p. 43). O gerador de lista de palavras provê o pesquisador com dados quantitativos de suas instâncias, isto é, produz uma lista das palavras diferentes existentes no texto, bem como suas frequências. A ordem desses dados pode ser escolhida pelo pesquisador a fim de adequar-se a seu projeto. Neste projeto a frequência das palavras, a recorrência dessas nos dois *corpora* e a presença dos contextos definitórios e explicativos foram itens considerados para a validação dos termos como objeto de pesquisa.

Na Figura 8 a seguir, mostramos a visão parcial da lista de palavras do *subcorpus* em inglês.



N	Word	Freq.	%	Texts	%	Lemmas	Set
1	LANGUAGE	2.686	0,27	15	100,00		
2	LANGUAGES	2.262	0,23	15	100,00		
3	WORD	1.871	0,19	14	93,33		
4	ENGLISH	1.706	0,17	14	93,33		
5	WORDS	1.696	0,17	15	100,00		
6	CHANGE	1.392	0,14	15	100,00		
7	FORM	1.285	0,13	15	100,00		
8	VERB	1.265	0,13	14	93,33		
9	MEANING	1.219	0,12	14	93,33		
10	FORMS	1.148	0,12	14	93,33		
11	CASE	1.139	0,12	15	100,00		
12	GERMANIC	925	0,09	12	80,00		
13	CREOLE	856	0,09	8	53,33		
14	CENTURY	837	0,09	14	93,33		
15	TIME	799	0,08	15	100,00		
16	ORIGIN	794	0,08	13	86,67		
17	ETYMOLOGY	785	0,08	6	40,00		
18	VERBS	767	0,08	9	60,00		
19	DEVELOPMENT	759	0,08	14	93,33		
20	GERMAN	748	0,08	13	86,67		
21	EUROPEAN	744	0,08	14	93,33		
22	ROOT	738	0,08	9	60,00		

Figura 8. Lista de palavras em ordem de frequência do *corpus* de LH – inglês (visão parcial).

A Figura 8 acima sofreu o processamento dos dados usando-se uma *stoplist*. A *stoplist* adotada neste trabalho é uma lista na qual há as palavras gramaticais que podem ser eliminadas dos *corpora* sem que afetem o objetivo do trabalho. A classe gramatical que faz parte da nossa busca, atendendo ao objetivo deste trabalho, são os substantivos. As palavras gramaticais, segundo Cavalcanti (2004), “abrange os pronomes/artigos e os conectivos, subdivididos em conjunções e preposições.” Segundo o autor, a função das palavras gramaticais é a de “localizar o ser no discurso, sem lhe acusar características” e para os conectivos é a de servir de ligação entre palavras ou a de articular o discurso.

Na Figura 9 a seguir podemos visualizar a lista de palavras em português do *corpus* de LH. Nesta lista também houve o tratamento do *corpus* com a *stoplist*; contudo, a *stoplist* do português não é tão completa quanto a do inglês. Consequentemente, há palavras lexicais, tais quais verbos, que restaram na lista de palavras. Esta lista é organizada na ordem de frequência dos itens no *corpus* de língua portuguesa.

N	Word	Freq.	%	Texts	%emmas	Set
1	É	5.589	0,54	33	100,0	
2	CARTA	2.541	0,25	15	45,45	
3	SÃO	2.373	0,23	33	100,0	
4	PORTUGUÊS	2.083	0,20	31	93,94	
5	LÍNGUA	1.960	0,19	33	100,0	
6	SER	1.892	0,18	33	100,0	
7	LAT	1.713	0,17	6	18,18	
8	PALAVRAS	1.632	0,16	30	90,91	
9	FORMA	1.456	0,14	33	100,0	
10	FOI	1.442	0,14	32	96,97	
11	RIO	1.321	0,13	20	60,61	
12	HÁ	1.277	0,12	30	90,91	
13	CF	1.240	0,12	19	57,58	
14	TEM	1.207	0,12	32	96,97	
15	QUANDO	1.172	0,11	31	93,94	
16	SÉCULO	1.159	0,11	32	96,97	
17	SUFIXO	1.154	0,11	11	33,33	
18	VERBO	1.109	0,11	19	57,58	
19	LATIM	962	0,09	24	72,73	
20	BAHIA	947	0,09	9	27,27	
21	EXEMPLO	938	0,09	32	96,97	
22	PODE	925	0,09	33	100,0	
23	DADOS	907	0,09	28	84,85	
24	FUTURO	888	0,09	16	48,48	

Figura 9 . Lista de palavras em ordem de frequência do *corpus* de LH – português (visão parcial).

A vantagem de se usar a *stoplist* para a confecção das duas listas de palavras é que a lista produzida será mais limpa, sem as palavras gramaticais. Ao elaborar uma lista de palavras-chave, as palavras gramaticais, se não retiradas, podem ser listadas. Como na Terminologia objetiva-se, primordialmente, os substantivos (palavras lexicais), a lista de palavras-chave sem a presença de palavras gramaticais é a mais desejável.

Além do gerador de lista de palavras, usamos também o extrator de palavras-chave, para levantar as unidades terminológicas que existem no *corpus* de estudo e compará-las com palavras de um *corpus* de referência. Neste caso, usamos o *corpus* de

referência COCA¹⁰, de 100 mil *types*, para a língua inglesa, e o Banco do Português¹¹, com um recorte também de 100 mil *types*, para o português. Em LC, *types* são as palavras diferentes que se encontram em um *corpus* (VIANA, 2010) ou palavras distintas ou formas, conforme Tagnin (2015). Em contraste com as formas, temos os *tokens*, que são as palavras presentes no *corpus*, independentemente de serem repetidas ou não, isto é, o número total de palavras presentes no texto ou ocorrências, segundo Tagnin (2015). O *corpus* de referência do COCA foi comprado e disponibilizado pelo orientador; o *corpus* de referência do português foi cedido por Tony Berber Sardinha.

Após essa comparação, o WST exibe as palavras-chaves positivas e negativas, aquelas que servem para identificar o *corpus* a ser descrito. As palavras positivas são aquelas que encontramos mais frequentemente no *corpus* de estudo, enquanto que as negativas são mais frequentes no *corpus* de referência. “A chavicidade reporta o resultado de um procedimento estatístico pelo qual a ferramenta levanta o quão importante cada palavra-chave positiva é para o *corpus* de pesquisa em relação ao de referência” (VIANA, 2010, p. 64). O autor explica que “as palavras-chave são empregadas com maior frequência **relativa** em um *corpus* de que em outro” (VIANA, 2010, p.62). Nesse caso, a chavicidade refere-se à frequência dessa palavra dentro de cada *corpus*. Além do aspecto estatístico, a chavicidade serve como referência para as unidades terminológicas que compuseram a obra. Elas foram as candidatas a termos que, quando confirmadas nos dois *corpora*, e trouxeram contextos definitórios aceitáveis, foram usadas na obra.

O extrator de palavras-chave, após comparar as palavras, permite o levantamento de índices linguísticos¹² do *corpus* de estudo. O dado levantado, considerado útil para este projeto foram as palavras-chave, aquelas pertencentes ao *corpus* de estudo que, quando comparadas ao *corpus* de referência, aparecem com mais frequência no primeiro do que no segundo, como ilustra a Figura 10.

¹⁰ DAVIES, M. COCA – **The Corpus of Contemporary American English** [*corpus*]. Provo, UT, USA: Brigham Young University, 2008.

¹¹ DAVIES, M.; FERREIRA, M. J. (2006-) **Corpus do Português: 45 million words, 1300s-1900s**.

¹² Neste contexto, os índices linguísticos são as palavras-chave positivas e negativas e as palavras em proporção semelhante ao *corpus* de referência.

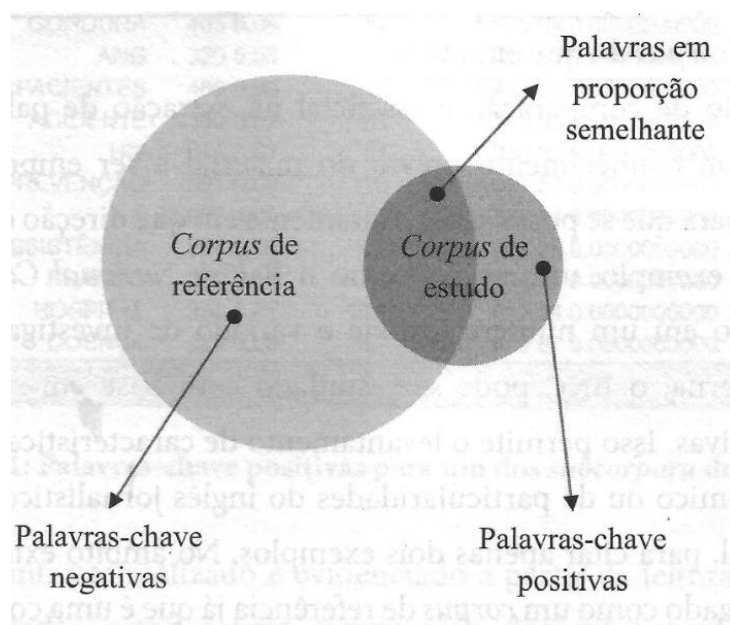


Figura 10. Procedimento de palavras-chave (VIANA, 2010, p.61).

A Figura 10 representa o processo de leitura e processamento de textos para a geração de palavras-chave, no qual temos três grupos de resultados: (i) empregadas na mesma proporção em ambos os *corpora*, (ii) frequentemente associadas ao *corpus* de estudo e (iii) mais recorrentes no *corpus* de referência. Neste trabalho, a lista que nos interessou é aquela cujos resultados estavam associados ao *corpus* de estudo, já que elas foram as que forneceram a lista de candidatos a termos.

Na Figura 11 abaixo, visualizamos, parcialmente, as duas listas de palavras-chave, do inglês e do português. Nessas listas já podemos verificar a presença de termos recorrentes que podem servir de candidatos a termos, tais quais: *língua/ language(s)*;

N	Key word	Freq.	%	Texts	RC Freq.	RC %	Keyness	P
1	LANGUAGE	2.518	0,26	15	60,1	0,01	9.764,2	0,00
2	LANGUAGES	2.170	0,23	15	7,55		16.086,0	0,00
3	WORD	1.799	0,19	14	78,7	0,02	4.993,9	0,00
4	ENGLISH	1.688	0,18	14	22,7		8.338,0	0,00
5	CHANGE	1.330	0,14	15	66,0	0,02	3.396,5	0,00
6	FORM	1.266	0,13	15	59,2	0,01	3.364,7	0,00
7	MEANING	1.213	0,13	14	5,10		8.587,7	0,00
8	VERB	1.150	0,12	13	1,54		10.312,0	0,00
9	FORMS	1.140	0,12	14	4,63		8.139,7	0,00
10	CASE	1.105	0,11	15	5,27		11.380,0	0,00
11	GERMANIC	855	0,09	12	3,72		8.888,3	0,00
12	ORIGIN	794	0,08	13	9,07		4.168,0	0,00
13	ETYMOLOGY	783	0,08	5	2,67		8.327,5	0,00
14	CREOLE	770	0,08	8	1,03		6.900,5	0,00
15	LINGUISTIC	762	0,08	14	3,95		5.106,6	0,00
16	DEVELOPMENT	757	0,08	14	96,0	0,02	789,23	0,00
17	GERMAN	739	0,08	13	4,10		4.864,4	0,00
18	ROOT	728	0,08	9	1,21		6.280,1	0,00
19	CENTURY	706	0,07	13	61,1	0,01	1.142,1	0,00
20	MODERN	697	0,07	14	43,6	0,01	1.499,0	0,00
21	RELATED	684	0,07	15	26,8		2.031,8	0,00
22	LATIN	647	0,07	15	15,1		2.532,3	0,00
23	EARLY	634	0,07	15	43,2	0,01	1.273,9	0,00
24	THUS	628	0,07	13	62,7	0,01	875,96	0,00

N	Key word	Freq.	%	Texts	RC Freq.	RC %	Keyness	P
1	CARTA	2.541	0,25	15	50,27		12.865,67	0,00
2	PORTUGUÊS	2.083	0,20	31	43,44		10.338,07	0,00
3	LÍNGUA	1.960	0,19	33	50,04		8.978,24	0,00
4	PALAVRAS	1.632	0,16	30	102,5	0,0	4.754,32	0,00
5	FORMA	1.456	0,14	33	492,6	0,0	550,50	0,00
6	SÉCULO	1.159	0,11	32	126,7	0,0	2.259,97	0,00
7	SUFIXO	1.154	0,11	11	516		12.964,85	0,00
8	VERBO	1.109	0,11	19	6,761		8.062,20	0,00
9	LATIM	962	0,09	24	3,370		7.950,14	0,00
10	EXEMPLO	938	0,09	32	264,2	0,0	539,54	0,00
11	FUTURO	888	0,09	16	84,66	0,0	1.936,49	0,00
12	PALAVRA	855	0,08	23	184,8	0,0	776,66	0,00
13	DOCUMENTO	841	0,08	19	50,13		2.529,30	0,00
14	FORMAS	837	0,08	29	102,6	0,0	1.474,01	0,00
15	AMIGO	809	0,08	12	35,70		2.877,64	0,00
16	NOME	803	0,08	22	145,3	0,0	930,13	0,00
17	BARÃO	778	0,08	3	13,46		4.135,32	0,00
18	PORTUGUESA	742	0,07	30	34,51		2.567,39	0,00
19	MARGEM	735	0,07	13	31,14		2.669,66	0,00
20	ESCRITA	710	0,07	25	32,16		2.491,15	0,00
21	FORMAÇÃO	692	0,07	26	167,6	0,0	524,97	0,00
22	LIVRO	680	0,07	20	128,0	0,0	748,63	0,00
23	SILVA	672	0,06	29	160,2	0,0	523,39	0,00
24	LÍNGUAS	662	0,06	26	11,85		3.475,73	0,00
25	REMETENTES	649	0,06	3	584		6.747,39	0,00
26	SUFIXOS	621	0,06	6	221		7.118,19	0,00

Figura 11. Lista de palavras-chave do *corpus* de LH (visão parcial).

Observando a Figura 11, na lista de palavras-chave em inglês (visão parcial), notamos a predominância de termos da sintaxe latina e de gramaticalização, enquanto que na lista em português vemos que além da gramaticalização, há termos ligados à etimologia e filologia. Analisando os contextos de concordância para extração de termos, vemos que nos textos em inglês há uma tendência ao retorno às línguas clássicas como o latim para explicar a gramaticalização nas línguas europeias. Já nos textos em português vemos uma LH preocupada em explicar as mudanças sofridas na língua portuguesa em seu processo de mudança do galego ao português atual.

Os candidatos a termos escolhidos para este trabalho foram substantivos. Os substantivos foram escolhidos pelo seu caráter denominativo, por agregarem em si a representação dos conceitos que trazem consigo sentido completo. Além disso, cumprem “a função de núcleo do sujeito, objeto direto e indireto e do agente da passiva. Do ponto de vista semântico, é o representante ideal do signo linguístico, pois representa objetos do mundo sensível e também conceitos abstratos” (SILVA, 2006, p. 33). Segundo a autora, ele está ligado ao ato de referência e nomeia coisas e seres. Estatisticamente, corresponde, em geral, de 50 a 60% do léxico de uma língua, podendo denotar “tudo o que o espírito humano pode propor como objeto do pensamento, permite ao verbo dizer algo sobre o substantivo sujeito e, se transitivo, do substantivo objeto” (p.33). Nesse trabalho, esses substantivos foram escolhidos a partir da equivalência conceitual dos termos extraídos da

lista de palavras-chave com 500 palavras, anterior à análise de equivalência conceitual. A equivalência conceitual foi pré-requisito determinante para que a palavra fosse incluída na lista de candidatos a termos. Essa análise foi feita pela leitura manual, a partir dos cotextos identificados em torno do termo por meio do Concordanciador.

Em relação à equivalência conceitual dos termos, como os *corpora* são distintos (equivalentes), não há como ter certeza de que uma palavra-chave ocorrerá nas duas listas e que, mesmo que existam palavras com equivalência conceitual em ambas as listas, essas palavras tragam em seus cotextos traços distintivos suficientes para construção de uma definição completa de uma entrada. A equivalência que buscamos nesse trabalho não é a equivalência no campo da sinonímia ou polissemia, mas no campo conceitual. Segundo Finatto, em terminologia, “conceitos são tidos como unidades de conhecimento e contêm apenas conhecimento factual ou técnico e não deveriam abrigar elementos emotivos ou conotados” (2001, p.214). Logo, faremos uma seleção de termos que tenham equivalentes nos dois *corpora* e, após análise dos contextos em que ocorrem, selecionaremos aqueles que atenderem à proposta de construção das definições desta obra terminográfica. Havendo equivalência terminológica e conceitual dos termos, eles constituiram o banco de entradas do vocabulário.

Para que a lista de palavras-chave trouxesse basicamente os substantivos, fizemos uma limpeza manual eliminando:

- (i) abreviações;
- (ii) algarismo romanos;
- (iii) prefixos e sufixos;
- (iv) vocabulário estrangeiro;
- (v) nomes próprios: *Jeremoabo, Bahia, Visconde, Pagotto, Vasconcellos etc*;
- (vi) termos pertencentes ao português arcaico: *hontem, meos, acção, anno, seos*, e às palavras gramaticais (conjunções, preposições, pronomes e advérbios) e lexicais (adjetivos e verbos), como por exemplo: *pode-se, trata-se, perifrástico, gândavo, românicas, latinos, verbais, borrado, tupi*;
- (vii) termos neolatinos, tais quais *adjectivos, mentum, corpus, abstractos*;
- (viii) desinências verbais *-aria*. Dentre eles, podemos mencionar *-mento, -ção*, dentre outros.

Após essa limpeza, tivemos uma lista que continha 189 palavras-chave em português e 163 palavras-chave em inglês (Apêndices 1 e 2). Além de substantivos, as listas também contêm alguns adjetivos que poderiam vir a compor as unidades sintagmáticas do vocabulário, como em **mudanças linguísticas** e **língua românica**.

Concluído o passo com as listas de palavras e palavras-chave, iniciamos a análise qualitativa dos dados com a ferramenta Concordanciador. Esse é o passo cuja explicação segue na próxima seção.

3.4 Concordanciador

A partir da identificação do candidato a termo na lista de palavras-chave e a certificação da existência desse nas duas línguas, iniciamos a análise dos contextos. Ao selecionarmos o termo, clicamos no botão *Compute* e clicamos em *Concordance*, ou então clicamos Shift+Control+C. Este comando abre uma nova janela, na qual podemos visualizar as linhas de concordâncias. Este processo se dá com o uso da terceira ferramenta, o Concordanciador, o que permite a localização de uma palavra em seu cotexto¹³.

Como exibido a seguir na Figura 12, observamos que na 19ª linha, encontramos o verbo **ser** à direita (R1) do termo **Linguística**. Para chegarmos até esta construção, clicamos na opção *collocates* que se encontra na aba inferior; selecionamos as formas *é* em português ou *is* em inglês. Em geral, o verbo *ser* é o verbo mais comumente usado para trazer os contextos definitórios para o pesquisador ou usuário do console.

¹³ Cotexto em Linguística de *Corpus* refere-se ao ambiente linguístico, isto é, os itens lexicais que estão à direita e à esquerda de uma dada palavra. (VIANA, 2010, p. 71).

N	Word	With	Relation Set	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	LINGÜÍSTICA	lingüística	0,00	18	621	19	19	4	6	7	2		583		2	7	6	4
2	DA	lingüística	0,00	16	305	246	59	9	24	20	70	123		9	9	9	18	14
3	A	lingüística	0,00	17	275	199	76	20	18	18	68	75		13	21	17	16	9
4	DE	lingüística	0,00	16	225	141	84	25	19	23	50	24		8	10	24	20	22
5	E	lingüística	0,00	16	170	88	82	25	13	15	19	16		34	16	10	12	10
6	QUE	lingüística	0,00	13	134	61	73	14	18	17	12			25	10	11	15	12
7	HISTÓRICA	lingüística	0,00	12	133	17	116	2	5	4	5	1		111				3
8	O	lingüística	0,00	14	98	43	55	18	13	11	1			5	10	28	6	6
9	NO	lingüística	0,00	11	89	28	61	5	10	13				28	12	10	5	6
10	EM	lingüística	0,00	13	72	42	30	9	7	7	4	15		4	7	7	7	5
11	DO	lingüística	0,00	14	65	29	36	6	11	9	3			18	2	3	6	7
12	COMO	lingüística	0,00	11	57	17	40	5	3	2	4	3		7	18	4	4	7
13	MUDANÇA	lingüística	0,00	10	53	49	4		2					1			1	2
14	UMA	lingüística	0,00	14	51	33	18	7	3	5	15	3			7	4	2	5
15	HISTORIOGRAFIA	lingüística	0,00	4	51	49	2	1			17	31						
16	NA	lingüística	0,00	9	49	31	18	6	3	4	7	11		2	2	6	3	5
17	PARA	lingüística	0,00	12	48	36	12	4	11	12	9			3	3	2	2	2
18	HISTÓRIA	lingüística	0,00	10	44	25	19	1	5	4	12	3			2	13	3	1
19	É	lingüística	0,00	13	43	11	32	2	2	3	2	2		10	9	5	4	4
20	SE	lingüística	0,00	12	40	9	31	4	3	2				4	7	4	7	9
21	BRASIL	lingüística	0,00	6	39	7	32	4		2	1			1	23	1	2	5
22	COM	lingüística	0,00	15	38	16	22	4	5	4	3			6	6	4	4	2
23	FILOLOGIA	lingüística	0,00	7	32	12	20	1	1	5	4	1		2	6	8	2	2
24	UM	lingüística	0,00	11	30	10	20	3	7					2	1	6	6	5
25	SOBRE	lingüística	0,00	11	30	23	7	8	1	9	4	1		2	3			2
26	À	lingüística	0,00	9	29	19	10	1	3	1	10	4		1	1	1	5	2
27	NÃO	lingüística	0,00	11	25	6	19	3	2		1			2	6	4	5	2

Figura 12. Vista da seleção do colocado é do termo Linguística.

Na Figura 12, vemos a seleção do verbo **É** para o termo Linguística. Há dez linhas de concordância que trazem os colocados para o termo **Linguística**. É o clique duplo que redireciona à frase e conteúdos diversos. Para acessar os contextos, clicamos duas vezes sobre o número abaixo da linha do verbo desejado, nesse caso *é/is*, e na coluna R1. Com um clique duplo, acessamos o contexto no qual o termo se insere, fazemos a análise do mesmo, principalmente quanto à presença de traços semânticos, e completamos a lista de conceitos, necessária para a construção final da definição. Vejamos a Figura 13 para o termo **Linguística** sobre os cotextos que o circundam.

N	Concordance	Set Tag	Word #	Ser	Ser	Par	Par	Hes	Hes	Ser
1	na escrita Admitindo a hipótese de que a mudança linguística é implementada primeiramente na fala e só		45	5	1	2	0	7	0	0
2	correntes contemporâneas. E, na sua percepção, 'A linguística é uma ciência muito nova [que] começou a		6.98	2	2	0	7	0	0	0
3	daquele que se lança à tarefa de escrever a história da linguística é estabelecer o que deve ser incluído no		407	1	3	0	4	0	0	0
4	, têm-se debruçado sobre a atividade historiográfica em linguística, é que um trabalho de tal natureza não se		8.53	3	4	0	9	0	0	0
5	produtos históricos. Se uma das tarefas da historiografia linguística é (re)estabelecer os pressupostos, nem		8.00	2	1	0	8	0	0	0
6	19, editadas pela autora. O objetivo central de sua análise linguística é averiguar se havia convergênc.W. entre a		6.84	2	6	0	4	0	0	0
7	Leite Vasconcellos, ao contrapor Linguística/Filologia: 'A Linguística é uma ciência de princípios gerais, aplicáveis		2.02	6	7	0	3	0	0	0
8	co-extensivas (Swiggers 1989). O objeto da historiografia linguística é construído, no sentido de que é uma		8.77	3	1	0	9	0	0	0
9	o recorte que define, para cada um, seu objeto formal. A Linguística é uma disciplina e um campo do		884	3	1	0	1	0	0	0
10	confirma a CRH e as autoras concluem que a mudança linguística é, de fato, gradual e não abrupta. Como se		5.73	1	6	0	1	0	0	0
11	por leis internas. 7. Curso de Linguística Geral, cap III: Linguística Estática e Linguística evolutiva; p.94 em		2.46	8	5	0	1	0	0	0
12	. É a vocação do colonizado, em busca da identidade linguística própria. Mas mexer com o problema da		176	7	9	0	3	0	0	0
13	, com maiores ou menores problemas de homogeneidade linguística, conforme a individualidade cultural dos grupos		11.9	6	8	0	5	0	0	0
14	relativas às bases teórico- metodológicas da Historiografia Linguística e, concomitantemente, mapear as		628	1	5	0	7	0	0	0
15	desenvolvimento da Linguística Histórica, a Historiografia Linguística se constitui como ciência em ascensão no		301	1	1	0	3	0	0	0
16	e na Universidade de São Paulo. Em síntese: faz-se hoje Linguística Histórica stricto sensu no Brasil na direção da		5.02	1	7	0	7	0	0	0
17	Revista argentina de historiografia linguística, I, 2, 115-136, 2009 Retrospectivas e		5	0	1	0	0	0	0	0
18	, deberían guiar la investigación sobre la historia de la linguística en Brasil. Palabras clave: Historia de la		323	1	9	0	3	0	0	0
19	clave: Historia de la linguística, historiografia de la linguística, Brasil, Mattoso Câmara. 1. Introdução1 O		335	1	7	0	4	0	0	0
20	una reivindicación del modelo disciplinar configurado por la linguística decimonónica. Finalmente, y a partir de lo		294	9	9	0	3	0	0	0
21	tiempo que inauguran la reflexión sobre la historia de la linguística brasileña, constituyen una reivindicación del		283	9	7	0	3	0	0	0
22	de la linguística en Brasil. Palabras clave: Historia de la linguística, historiografía de la linguística, Brasil, Mattoso		331	1	5	0	4	0	0	0
23	Interface Linguística Histórica e Filologia – Marilza de Oliveira		2	0	4	0	0	0	0	0
24	para alterar conceitos (arraigados) da inferioridade linguística do índio, a que corresponderia sua		6.52	2	7	0	7	0	0	0

Figura 13. Linha de concordância para o termo *linguística* seguida do colocado *é* (vista parcial).

Nas linhas de concordância, encontramos os fragmentos dos textos ou cotextos que são exibidos, a partir do termo **Linguística**, selecionado pelo usuário, como podemos visualizar na Figura 13 acima.

Na figura 14 podemos ver o termo **Linguística** inserido em contexto definatório com o verbo *é* a sua direita.

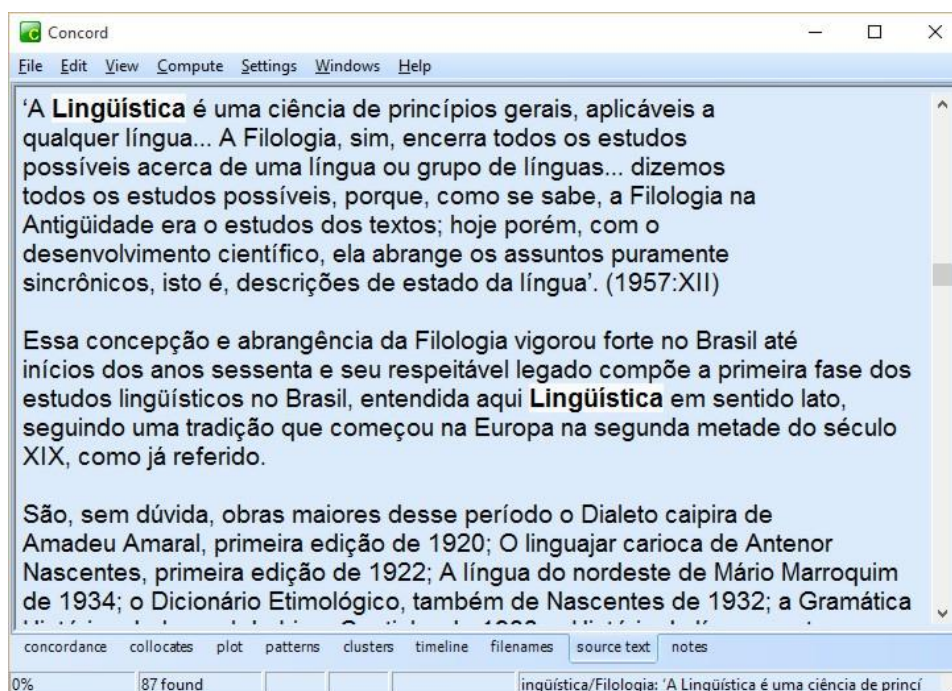


Figura 14 . Exemplo do termo **Linguística**, acompanhado do colocado *é* na primeira posição à direita (R1).

Na figura 14 observamos o termo **Linguística** inserido em um contexto definatório, diferenciando a Linguística da Filologia. O pesquisador valeu-se desses contextos para extrair os dados que foram organizados em fichas terminológicas e posteriormente fomentaram a construção das definições.

Na seção seguinte, apresentaremos a lista de candidatos a termos, explicaremos como vieram a fazer parte dela e ilustraremos a classificação disponibilizada, em forma de cores, na legenda que as explica.

3.5 Candidatos a termos

Após concluirmos as listas de palavras e a de palavras-chave, partimos para a construção da lista de candidatos a termos. A lista de candidatos a termos é uma lista na qual catalogamos os possíveis termos que podem vir a compor os verbetes do vocabulário de LH. Os candidatos a termos foram classificados levando em consideração três princípios: (i) a lista de palavras-chave em português como língua de partida e (ii) a lista de unidades sintagmáticas (*clusters*), e (iii) a relação conceitual e lexical que apresentaram, no contexto em que se inseriam no *corpus* analisado.

Abaixo, no Quadro 3, segue uma visão da lista de candidatos a termos e sua classificação quanto à presença ou ausência de relação conceitual. Logo abaixo há a legenda com a explicação das cores e seus conceitos.

Quadro 4. Candidatos a termos da LH em português e inglês.

CANDIDATOS A TERMOS DE LH					
Português	Ranking Freq.	Nº ocorr.	Inglês	Ranking Freq.	Nº ocorr.
1. Língua	40	1960	Language	29	2686
2. Etimologia	333	294	Etymology	110	785
3. Sufixo	69	1154	Suffix	280	350
4. Escrita	116	710	Writing	1639	65
5. Gramática	158	563	Grammar	259	382
6. Palavra	87	855	Word	45	1871
7. Verbo	70	1109	Verb	70	1265
8. Linguística	457	221	Linguistics	143	624
9. Mudança	166	549	Change	62	1392
10. Fala	209	440	Speech	297	338
11. Linguagem	210	439	Language2	29	2686
12. Texto	270	349	Text	669	164
13. Variação	232	402	Variation	544	200
14. Derivação	248	376	Derivation	529	204
15. Substantivo	265	354	Noun	219	444

16. Adjetivo	941	111	Adjective	513	209
17. Prefixo	1717	57	Prefix	735	148
18. Léxico	598	174	Lexicon	1007	109
19. Analogia	701	148	Analogy	1117	97
20. Clítico	1286	80	Clitic	707	155
21. Advérbio	1307	78	Adverb	1036	105
22. Preposição	1367	75	Preposition	1135	96
23. Particípio	1461	69	Participle	686	161
24. Alternância	1381	73	Alternation	918	118
25. Sufixação	1581	63	Suffixation	9826	7
26. Prefixação	1769	55	Prefixation	6898	11
27. Cognato	4020	21	Cognate	369	270
28. Acusativo	2048	46	Accusative	654	166
29. Afixo	3846	22	Affix	2033	50
30. Vogal	233	400	Vowel	158	577
31. Cultismo	2093	45	Formalism	22166	2
32. Filólogo	2326	40	Philologist	25053	2
33. Linguística Histórica	457	9	Historical Linguistics	143	147
34. Mudança linguística	549	47	Linguistic change	62	55
35. Morfologia derivacional	712	6	Derivational morphology	183	3
36. Pronome pessoal	1072	8	Personal Pronoun	482	16
37. Historiografia Linguística	1021	31	History of Linguistics	129	8
38. Língua românica	40	7	Romance Language	29	6

No quadro 4 acima, temos os candidatos a termos em português na primeira coluna, a ordem de frequência na segunda coluna e o número de ocorrências na terceira. A mesma ordem se segue para os termos em inglês da quarta à sexta coluna. Podemos observar que tanto a ordem quanto o número de ocorrências em ambas as línguas são diferentes. Como explicitado anteriormente, a opção metodológica para a escolha dos candidatos a termos deste trabalho baseou-se na lista de palavras do português.

No Quadro 5 a seguir, podemos encontrar a relação estabelecida entre as cores e os conceitos que serviram para classificar os candidatos a termos acima expostos. Esta classificação foi criada, considerando-se a equivalência conceitual dos termos em seus contextos.

Quadro 5. Legenda usada para classificação dos candidatos a termos.

ANÁLISE DE CANDIDATOS A TERMOS		
PORTUGUÊS	INGLÊS	TERMO
1. S	S	Termo
2. S/N	N/S	Ñ termo (falta de equivalência conceitual)
3. S	N	Equivalentes, mas sem contexto explicativo/definitórios suficiente em inglês
4. N	S	Equivalentes, mas sem contexto explicativo/definitórios suficiente em português
5. N	N	Sem contextos exp. ou def. nas duas línguas

No quadro 5, a cor cinza número 1 refere-se aos termos que são correspondentes nas duas línguas, isto é, há uma relação conceitual entre os termos. Essa relação é identificada a partir das linhas de concordâncias disponibilizadas pelo Concordanciador do WST. Após a identificação dos contextos, fizemos a análise para a extração de traços conceituais e posterior construção da definição. Para exemplificar esta relação, vejamos os contextos nos quais se inserem os termos *Linguística Histórica*/*Historical Linguistics* nos excertos abaixo.

(1) *Linguística Histórica*:

Na tradição da linguística histórica, os estudos linguísticos diacrônicos incidem predominantemente sobre a mudança fonológica e a analogia. No âmbito da analogia, salvo raras exceções, o linguista histórico estuda mudanças operadas nos sistemas flexionais, enquanto as mudanças que dizem respeito à formação de palavras são quase sempre tratadas dentro das mudanças lexicais e semânticas.

(2) *Historical Linguistics*:

[...] nevertheless, historical linguistics is a major, thriving area of linguistics, as well it should be, given the role it has played and continues to play in contributing towards the primary goals of linguistics in general.

A segunda linha, número 2, de cor azul, marcaria os termos sem equivalência conceitual. Contudo, na análise da lista de palavras-chave com 500 ocorrências, tal situação não ocorreu. Caso houvesse, identificaríamos a equivalência por meio de contextos associativos, mas haveria a ausência de contextos definitórios ou explicativos suficientes para a construção de uma definição. A lista de palavras-chave com 500 ocorrências é um padrão do WST, e foi o tamanho escolhido para o dimensionamento deste trabalho. Após a limpeza mencionada no item 3.3, a lista que continha 500 ocorrências foi reduzida para o número de 189 palavras-chave em português e 163 palavras-chave em inglês

(Apêndice1). Como nossa proposta objetivou o desenvolvimento de um vocabulário, não alteramos este dimensionamento a fim de aumentá-lo.

A terceira linha, de cor salmão, refere-se aos termos equivalentes nas duas línguas. Contudo, não foi possível identificar traços conceituais suficientes na língua inglesa para construção da definição.

Como exemplo, podemos mencionar o termo sufixo e *suffix*, encontrados nos seguintes excertos:

(3) Sufixo

Câmara Jr. (1975: 218) define a noção de produtividade (aqui tida como produtividade sufixal) como a possibilidade de podermos "destacar", i.e., a capacidade de identificarmos (segmentarmos) um determinado sufixo, em "palavras derivadas que vieram tais do latim ou, por empréstimo, de outra língua. Ou, em outros termos, dadas palavras derivadas passam a servir de modelo para a estruturação de novas palavras, fornecendo no seu elemento final um meio permanente na língua para novas derivações.

(4) *Suffix*

Whatever the origin of gaulick, -ick in it was mistaken for a suffix, because otherwise the form gaulish, also attested in North Country dialects, would not have arisen. However, this is a secondary development, and we need not be deceived by folk etymology.

A quarta linha, de cor laranja, marca os termos equivalentes nas duas línguas. Contudo, o termo em português não apresenta contextos definitórios ou explicativos que tragam traços conceituais suficientes para a construção de uma definição. Para exemplificar esta ocorrência, vejamos os termos cognatos/*cognates*.

(5) Cognatos

Consideramos estas como base para a formação em –MENTO, já que foram a base para a formação de um verbo em –AR que, em seguida, derivou um substantivo em –MENTO, ou seja, as bases desses cognatos são diferentes.

(6) Cognates

Cognates are recognized by the regular correspondences between their sounds that are the direct result of regular sound changes; see especially Hoenigswald 1960 for discussion of this fundamental point.

A quinta e última linha marca termos equivalentes nas duas línguas, inseridos em linhas de concordância que não apresentam contextos definitórios ou explicativos suficientes para a construção de definições. Temos, nesse caso, os termos *preposição/preposition*. Vejamos:

(7) *Preposição*

Um dos dados com essa variante apresentou a preposição *a* entre os dois verbos, como nas construções perifrásticas de futuro documentadas para o espanhol e para o italiano. Esse dado foi único em todo o material analisado nesta tese.

(8) *Preposition*

When, this derived nominal was used after a preposition to, the dative form of writan, i.e. writenne, was used. Their nominal status is also shown by the fact that there were no passive and perfective constructions available until Middle English (ME).

Estas foram as cinco formas diferentes de classificar as diferentes ocorrências dos candidatos a termos encontrados nos *corpora*. As formas equivalentes nas duas línguas, que traziam consigo contextos definitórios ou explicativos, com traços semânticos suficientes para a construção das definições foram as escolhidas para serem lançadas na plataforma do VoTec.

Considerando as ocorrências de termos nos *corpora*, observamos que quanto mais específico for o termo, maior será a probabilidade de ocorrência de contextos definitórios; quanto mais genéricos, antigos ou difundidos forem os termos, menor será a probabilidade de ocorrência de contextos explicativos e definitórios. Por genéricos, entendemos serem aqueles termos que são mais conhecidos ou difundidos na comunidade acadêmica ou geral, não carecendo de serem definidos.

Na seção 3.5.1 tratamos da validação dos termos, em geral, feita por especialistas, mas nesse trabalho, esse passo será baseado na confiabilidade de dados proporcionada pela metodologia e abordagem da LC.

3.5.1 Validação de termos

A validação dos termos por parte dos especialistas é um procedimento terminográfico sugerido pelos teóricos da área (BARROS, 2004; KRIEGER E FINATTO, 2004; SILVA, 2006). Contudo, apesar deste projeto ter sido submetido ao Comitê de Ética

e Pesquisa da Universidade Federal de Uberlândia (UFU), não houve tempo hábil para aprovação e liberação para autorização das entrevistas. Observamos que o ritmo de processamento dos projetos é moroso, e as várias idas e vindas do processo mostram que uma *checklist* para este tipo de departamento seria desejável.

Obviamente seria importante haver um retorno dos especialistas, quer fossem filólogos, etimólogos e linguistas históricos, principalmente porque o olhar do outro sobre o mesmo objeto traz uma contribuição que produz crescimento e confiabilidade ao trabalho.

Contudo, o fato de termos adotado Linguística de *Corpus* como abordagem nos respalda e nos permite ter um retorno com o mesmo nível de qualidade, ou até melhor, devido ao uso de estatística no WST. Tagnin (2012) afirma que trabalhos de Linguística de *Corpus* direcionados por *corpus* são constituídos pelo vocabulário mais frequente da área, incluindo as colocações. Isto se deve ao fato de que a LC é uma abordagem empírica para o estudo da língua, baseado nas observações.

A autora explica que a Terminologia direcionada por *corpus*, podendo ser chamada de fraseologia especializada (TAGNIN, 2012, p.169), extrai os termos do *corpus*. Devido a esta extração cuidadosa dos *corpora*, o que reflete a língua usada naquela área de especialidade, o produto é altamente confiável, o que dispensa a validação dos especialistas. Além da identificação dos termos nas listas de palavras e palavras-chave, a checagem do contexto no qual eles ocorrem trazem mais uma confirmação de que o termo escolhido pertence àquela área e ocorre em situações reais de uso.

Neste trabalho chegamos à situação descrita pela autora, na qual identificamos os termos e as colocações frequentes que ocorrem com os substantivos escolhidos pela lista de palavras-chave, tais quais: mudança linguística/*linguistics change*; morfologia derivacional/*derivational morphology*, historiografia linguística/*history of Linguistics*.

Na próxima seção, apresentamos os passos do fazer terminográfico tomados para o registro de dados na plataforma do VoTec. Explicaremos os passos tomados desde a inserção de dados, a construção da macro e microestrutura até a construção da definição final, disponibilizada aos consulentes.

3.6 VoTec

Em primeiro lugar, faz-se necessário justificar a escolha deste trabalho pela plataforma do VoTec – Vocabulário Técnico online - em detrimento de outras escolhas possíveis de ferramentas terminológicas. Escolhemos esta plataforma por ter sido criada por Fromm (2007), sendo ele o orientador desta dissertação. Em segundo lugar, por ser um banco de dados disponível *on-line*, configurando-se uma plataforma de gestão terminológica na *web*, na qual podemos inserir os dados que poderão ser posteriormente consultados na página de visualização.

Outra razão é o fato de ser uma plataforma que permite que as definições para os termos sejam feitas a partir dos dados provenientes dos *corpora*. A partir das colunas de traços semânticos, preenchidas com dados dos *corpora*, o pesquisador agrupa os dados recorrentes para a construção da definição. Esta característica da plataforma permite reduzir a subjetividade do pesquisador e torna a definição mais acurada e mais objetiva.

Finalmente, conforme propõem Wilkens *et alii* (2012), uma plataforma terminológica deve ser de fácil acesso, uso e de uma fonte confiável. Devido a sua disponibilização *on-line*, o usuário pode acessar o VoTec de qualquer lugar onde haja acesso à Internet. Por ser uma plataforma cujas opções de uso são mais objetivas (normal ou descritiva; total, tradutor ou modular) o usuário não se confunde com o *layout*. O aspecto confiável se faz presente pelos tipos de *corpora* usado para extração de contextos definitórios: artigos científicos, dissertações e teses. Os autores também propõem que haja a opção de imagens disponíveis na plataforma. No entanto, como nosso foco é o vocabulário da LH, trabalhamos com a metalinguagem desta subárea da Linguística. Consequentemente, por ser um vocabulário abstrato, a possibilidade de disponibilizarmos imagens não se aplica.

Para definirmos essa plataforma, dizemos que o “VoTec é uma ferramenta que se vale de *corpora* técnicos para a construção de seus verbetes e de um banco de dados (ambos exaustivamente descritos) para o seu funcionamento” (FROMM, 2007, p. 8). É uma ferramenta que oferece várias formas de visualização (normal e descritiva), e as consultas podem ser nos módulos total, tradutor e modular. A visualização normal é o formato que segue o padrão dos dicionários impressos, conforme é possível ver na Figura 15, abaixo.

Vocabulário Técnico Online Tela Cheia | English | Ajuda

Linguística ▼ Linguística Teórica ▼ Linguística Histórica ▼

Buscar

Tipos de Exibição

Normal

Descritiva

Tipos de Consulta

Total

Tradutor

Modular

Consultas Externas

Corpus NILC

Google

Answers.com

Wikipedia

CORTEC

▼ Português

[Voltar ao resultado da busca](#)

verbo. *Linguística Histórica.* **s.m.s/p.** classe de palavras que em orações matriz afirmativas do século 19 é precedido de um item referencial

NOTA: em orações coordenadas, é precedido de oração; há uma tendência à ênclise quando o verbo é o limite dos períodos de tempo; pode ser precedido de constituintes pelo clítico ou por uma sentença dependente;. Ex.: Por esse motivo o verbo *ßap..d.*, no futuro, torna-se *ßap.....*. Quando esse verbo é transformado em substantivo, o (sigma) permanece antes do sufixo *-µ..*, tornando-se em *ßap...µ..*, que é o ato da imersão.. **Sinônimos:** núcleo do predicado das sentenças. **Hipônimo de:** palavras; noções de ação, processo ou estado. **Corpus:** *Posição na Ordem de Frequência:* (48); *Nº de Ocorrências do termo:* (1109). **Informações Enciclopédicas:** Verbo é toda palavra que encerra ideia de ação ou estado. A palavra verbo vem do latim *verbum*, que significa palavra. Em: *Verbo* - [Wikipédia](#)

Figura 15. VoTec – visualização normal – termo verbo

A **visualização normal total**, como mostra a Figura 15, traz os dados em sequência. Dados como NOTA e exemplos e as relações de sinonímia, hiponímia, antonímia, bem como dados do termo no *corpus* vêm em ordem ininterrupta.

Diferentemente da visualização **normal**, a **descritiva** apresenta os dados de forma hierárquica e detalhada, como podemos ver na Figura 16.

Vocabulário Técnico Online Tela Cheia | English | Ajuda

Linguística ▼ Linguística Teórica ▼ Linguística Histórica ▼

Buscar

Tipos de Exibição

Normal

Descritiva

Tipos de Consulta

Total

Tradutor

Modular

Consultas Externas

Corpus NILC

Google

Answers.com

Wikipedia

CORTEC

▼ Português

[Voltar ao resultado da busca](#)

verbo

classe de palavras que em orações matriz afirmativas do século 19 é precedido de um item referencial

NOTA: em orações coordenadas, é precedido de oração; há uma tendência à ênclise quando o verbo é o limite dos períodos de tempo; pode ser precedido de constituintes pelo clítico ou por uma sentença dependente;

Abreviação/Acrônimo: *Nada encontrado*

Categoria Gramatical: substantivo

Gênero: masculino

Número: dual

Córpus

- **Posição na Ordem de Frequência:** 48
- **Nº de Ocorrências do termo:** 1109

Ontologia: [Linguística](#) > [Linguística Teórica](#) > [Linguística Histórica](#)

Figura 16. VoTec – visualização descritiva parcial – termo: **verbo** (português)

Na Figura 16, observamos como os dados são dispostos de forma diferente da **visualização normal total**. Na opção **visualização descritiva total**, os dados são dispostos de forma separada, facilitando ao leitor a localização de dados específicos como categoria gramatical, posição no *corpus*, entre outros, mais rapidamente.

Abaixo segue a visualização da apresentação **descritiva total** do termo *verb* em inglês, conforme mostra a Figura 17.

The screenshot displays the VoTec interface for the term **Verb** in English. On the left sidebar, there are options for 'Normal' and 'Descritiva' (selected), and a list of 'Tipos de Consulta' (Total, Tradutor, Modular) and 'Consultas Externas' (Corpus NILC, Google, Answers.com, Wikipedia, CORTEC). The main content area, titled 'English', includes a link 'Go back to search results' and the following information:

- Verb**: part of speech that has a base form, can have a prefix, varies in singular or plural and can be followed by an object
- Acronym**: Nothing found
- Part of Speech**: noun
- Gender**: neutral
- Number**: singular
- Corpus**:
 - Frequency order position: 76
 - Term number of occurrences: 1916
- Ontology**: [Linguistics](#) > [Theoretical Linguistics](#) > [Historical Linguistics](#)
- Morphosyntactic Variants**: Nothing found
- Synonyms**: Nothing found
- Antonyms**: Nothing found
- Hyponym of**: part of speech, words, phrase, construction
- Co-hyponyms**: Nothing found
- Hypernym of**: Nothing found
- Examples**:
 - The ACI constructions which are found in OE involve this type of verb. The verb is a three-place predicate, (while ECM verbs are two-place verbs, i.e. onotransitive) and the relevant constructions found in OE are analyzed as a double object construction (see Lightfoot 1991; Gelderen 1993:43):

Figura 17. VoTec – visualização descritiva parcial – termo : *verb*

A **visualização descritiva total** expõe todos os dados na posição vertical, separados uns dos outros. Na primeira parte há a definição; logo após, a classificação gramatical do termo; em seguida, os dados do termo no *corpus*, a ontologia e dados lexicais e os exemplos na parte final.

Dentre as opções de consulta, a de **módulo total** (Figuras 15, 16 e 17) disponibiliza todos os campos do banco de dados, sendo ela de caráter lexicográfico, na qual temos o termo, área a qual pertence, classificação gramatical, definição, exemplos, sinônimos, hipônimos, hiperônimos, frequência no *corpus*, ocorrência, nota e informações enciclopédicas. Já a consulta tradutor (Figura 18) é de cunho mais tradutológico. Nela há

o termo, área a qual pertence, classificação gramatical, definição, nota (se houver), exemplos e sinônimos. Nesse modo de consulta, as informações são mais objetivas, no sentido de que o que importa ao tradutor é a equivalência terminológica e semântica. Isto é, a visualização em módulo tradutor permite ao consulente acessar as informações mais frequentes usadas por tradutores, tais quais: definição, área de especialidade, tradução, sinônimos, entre outros. Observe a apresentação na imagem abaixo.

Vocabulário Técnico Online Tela Cheia | English | Ajuda

Linguística Histórica

Tipos de Exibição

Normal

Descritiva

Tipos de Consulta

Total

Tradutor

Modular

Consultas Externas

Corpus NILC

Google

Answers.com

Wikipedia

CORTEC

Português

[Voltar ao resultado da busca](#)

Verbo. *Linguística Histórica.* s.m/f.s. classe de palavras precedida de sujeitos ou oração, com categoria de tempo, modo, número, pessoa e aspecto, cujas formas nominais são de infinitivo, de gerúndio ou de participio. Ex.: Como se viu na seção 5.2.1., os verbos irregulares que mantêm o futuro simples na fala são ser, ir, haver e ver. Para efeito de comparação com os dados de fala, procedeu-se ao exame dos 22 verbos irregulares encontrados na escrita da década de 90.. *Veja Também:* [Palavra](#).

English

[Go back to search results](#)

Verb. *Historical Linguistics.* n.m/f.s. part of speech that has a meaning, a base form, can have a prefix, varies in singular or plural and can be followed by an object. Ex.: The ACI constructions which are found in OE involve this type of verb. The verb is a three-place predicate, (while ECM verbs are two-place verbs, i.e. onotransitive) and the relevant constructions found in OE are analyzed as a double object construction (see Lightfoot 1991; Gelderen 1993:43). *Synonyms:* three-place predicate. *See Also:* [Word](#).

25/06/2015 07:12 © 2007 Guilherme Fromm - ICMC Jr.
Termo elaborado por [Marcio Issamu Yamamoto](#) (pt)

Figura 18. VoTec – visualização normal tradutor – termo: **verb**

Finalmente, a consulta modular (Figura 19) exhibe a microestrutura de acordo com a busca do consulente, acelerando sua busca pela seleção mais específica de dados, ou seja, é uma opção de caráter mais híbrido (lexicográfica e terminológica). Nesse modo o consulente pode personalizar sua busca por meio de um clique na informação desejada (abreviatura, categoria gramatical, gênero, número, ontologia, variações morfossintáticas, definição, forma por extenso, *corpus*, exemplos, informações enciclopédicas, remissiva). Além dessas opções de visualização, o *site* disponibiliza ao usuário acesso para consultas externas. Esse modo de pesquisa apresenta características lexicográficas pela disponibilização de dados mais detalhada sobre o termo, como a microestrutura de um dicionário de língua geral; terminológicas, pois ao personalizar a busca, o consulente tem à disposição somente os dados que desejar, para que a tarefa da tradução seja concluída. Ou seja, se o consulente buscar somente o termo e a definição para confirmação de

conceitos correspondentes, teremos uma exibição mais recorrente aos dicionários de cunho terminológico.

<div>Tipos de Exibição</div> <div>Normal</div> <div>Descritiva</div> <hr/> <div>Tipos de Consulta</div> <div>Total</div> <div>Tradutor</div> <div>Modular</div> <hr/> <div>Pré-definição</div> <div>Abreviatura</div> <div>Categoria Gramatical</div> <div>Gênero</div> <div>Número</div> <div>Ontologia</div> <div>Variações Morfossintáticas</div> <hr/> <div>Definição</div> <div>Definição</div> <div>Forma por Extenso</div> <div>Cópus</div> <hr/> <div>Exemplificação</div> <div>Exemplos</div> <div>Informações Encilopédicas</div>	<div>▼ Português</div> <div>Voltar ao resultado da busca</div> <div>Verbo. classe de palavras precedida de sujeitos ou oração, com categoria de tempo, modo, número, pessoa e aspecto, cujas formas nominais são de infinitivo, de gerúndio ou de particípio. <i>Corpus: Posição na Ordem de Frequência: (71); N° de Ocorrências do termo: (1113).</i></div> <div>▼ English</div> <div>Go back to search results</div> <div>Verb. part of speech that has a meaning, a base form, can have a prefix, varies in singular or plural and can be followed by an object. <i>Synonyms: three-place predicate. Corpus: Frequency order position: (70); Term number of occurrences: (1265).</i></div>
---	---

Figura 19. VoTec – visualização normal modular – termo: *verb*

Na Figura 19, observamos que a quantidade de informação é restrita, sendo que as informações disponibilizadas aos consulentes são a definição do termo *verb* e os dados referentes a sua ocorrência no *corpus*.

A abordagem do banco de dados adotada para o VoTec é a terminográfica, na qual os campos usados para a criação do verbete do dicionário constam na ficha terminológica. O acesso do pesquisador ao banco de dados se faz por meio do cadastro de um usuário e de uma senha pessoal. Na próxima tela, ele pode clicar na opção **Novo Termo** e iniciar o processo de registro de dados na plataforma.

Vocabulário Técnico Online Tela Cheia | English

Novo Termo

Voltar ao painel

Passo 1

Termo

Língua
 Escolha uma língua: Inglês ▼

Ontologia
 Grande Área: Linguística Histórica ▼
 Sub-área 1: Etimologia ▼

Próximo Passo (Contextos)

Figura 20. Visualização do passo 1 para cadastro de termos no VoTec.

Nessa primeira página (Figura 20), o pesquisador deve inserir o termo a ser registrado no item **Termo**, devendo optar, na sequência, pela língua do referido termo (português ou inglês) e, finalmente, escolher sua ontologia (que é uma representação da árvore de domínio aqui anteriormente citada) desde a sua primeira subárea, nesse caso LH, até a última subárea: Etimologia, LH ou Filologia. Selecionados esses dados, o próximo passo foi o cadastro de contextos, como podemos ver na Figura 21.

Cadastro de Contextos para Etymology

Voltar ao Painel Próximo Passo

Novo Contexto

Passo 2

Dados do Contexto

Exemplo*:

Conceito*:

Fonte*: OPDF ▼ Cadastrar Nova

Fonte

Data de Coleta*: 16/02/2013 (dia/mês/ano ex.: 18/03/2007)

Salvar

Figura 21. Visualização da página Cadastro de contextos para *etymology* (visão parcial).

Nesta etapa do procedimento, como podemos verificar na Figura 21, no espaço chamado Exemplo*, registramos o contexto no qual a unidade terminológica se insere,

dados esses provenientes do Concordanciador do WST. A partir desse exemplo, faz-se um recorte de possíveis contextos definitórios ou explicativos¹⁴ e o transpomos para o campo Conceito*. Logo abaixo do conceito, há o campo de fonte de documento usada para inserção de dados na plataforma, neste caso a forma 0PDF. A escolha deste formato, 0PDF, foi devido ao fato de trabalharmos com textos de fontes diversas, logo a padronização do campo fonte se fez necessária por uma questão metodológica. Abaixo desse campo, há o campo Data de coleta* onde inserimos a data em que o texto do *corpus* foi coletado, registrado no documento em formato txt. Nesse tipo de documentos, inserimos no rodapé os dados de coleta daquele texto: fonte e data como podemos ver na figura 22.

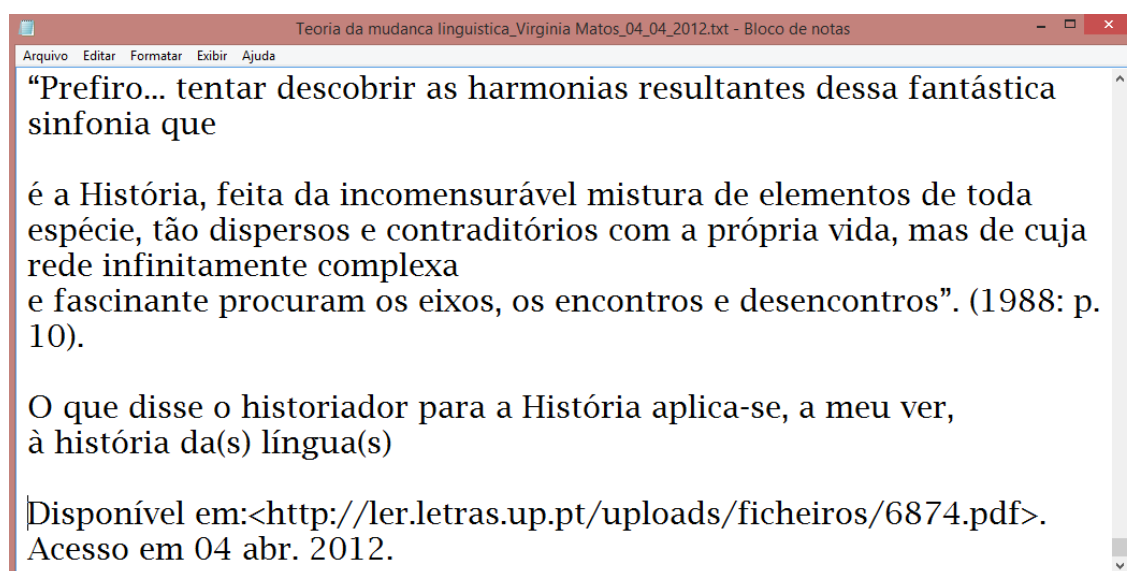


Figura 22. Visualização parcial do *corpus* salvo em txt com dados de coleta.

Na parte inferior dessa página, há o registro dos contextos cadastrados. À direita desse campo, há as opções editar e excluir que podem ser usadas para alteração ou exclusão de dados, de acordo com a avaliação do pesquisador, conforme mostra a Figura 23.

¹⁴ Segundo Aubert (1996, p. 66-67), “os contextos explicativos apresentam alguns traços conceituais pertinentes específicos do termo sob observação, frequentemente relativos à materialidade, finalidade, funcionamento e similares. [...] Talvez mais desejáveis, mas certamente menos encontradiços, os contextos definitórios proporcionam um conjunto completo dos traços conceituais distintivos do termo.” Além desses dois contextos, Aubert também menciona os contextos associativos: aqueles no qual o termo se relaciona ao tema da pesquisa, mas não indica seus traços conceituais.

Novo Contexto

Passo 2

Dados do Contexto

Exemplo*:

Conceito*:

Fonte*: [Cadastrar](#)

[Nova Fonte](#)

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

Contextos Cadastrados

Exemplo	Conceito	Fonte	Ações
Etymology is a vague concept, as shown by the polite warfare between those who look on it as the science of reconstruction and those who allow it to subsume "the history of words."	the science of reconstruction; the history of words	PDF 16/02/2013	editar - excluir
With one exception, Etymology is the least valuable element in the making of a dictionary, as it is of interest only to those who wish to study the history of language. It helps no man in his use of the word bishop to know that it comes from two Greek words, epi, meaning upon, and scopos, meaning a looker, still less to be told into what forms those words have passed in Spanish, Arabic, and Persian.	to study the history of language	PDF 16/02/2013	editar - excluir
When etymology tries to solve the riddle of the origin of language, it fails. In Indo-European, one can seldom go beyond the roots posited by Brugmann, Walde, and Pokorny (with or without laryngeals): the circumstances in which allegedly primitive sound strings like *bhlag-, *ster, and *wegh- (in so far as they are not obviously onomatopoeic) came by their meanings is hidden.	tries to solve the riddle of the origin of language	PDF 16/02/2013	editar - excluir

Contextos Cadastrados: 3

Figura 23. Visualização da página Cadastro de contextos para o termo *etymology* (visão parcial).

O acesso ao próximo procedimento é possível ao se clicar sobre o ícone Próximo Passo no canto superior direito da página. Ao acessar a próxima página, um grupo de fichas terminológicas serão preenchidas conforme podemos identificar na parte inferior da Figura 24.

Vocabulário Técnico Online Tela Cheia | English

Termo: Verb

[Passo Anterior](#)
[Salvar](#)
[Sair sem salvar](#)

Contextos Conceito Final / Definição

Exemplo	Conceito	Fonte
1 The ACI constructions which are found in OE involve this type of verb. The verb is a three-place predicate, (while ECM verbs are two-place verbs, i.e. onotransitive) and the relevant constructions found in OE are analyzed as a double object construction (see Lightfoot 1991; Gelderen 1993:43):	is a three-place predicate	PDF 13/02/2013
2 In metonymic extensions to the context in which the P/I occurs, the meaning of the verb	the meaning of the verb is extended to denote certain features of the broader event	PDF 15/02/2013

Dados Traços Distintivos Semântica Termo Equivalente Termos Remissivos Informações Enciclopédicas

Dados

Ontologia: Linguística > Linguística Teórica > Linguística Histórica

Figura 24. Visualização de página da elaboração da microestrutura de um verbete no VoTec.

Nesse momento do preenchimento da ficha terminográfica, o pesquisador preenche as abas (i) Dados, (ii) Traços distintivos, (iii) Semântica, (iv) Termo equivalente, (v) Termos remissivos, (vi) Informações enciclopédicas e (vii) Conceito Final/Definição, como exposto na Figura 24. Abaixo segue uma descrição de cada um deles e o que deve ser inserido.

i. Dados: na parte superior desta aba, temos os dados ontológicos da unidade terminológica. Nesse caso, Linguística > Linguística Teórica > LH. Logo abaixo, há o campo categoria gramatical, onde selecionamos a opção substantivo para o termo *língua*, no canto esquerdo. À direita, há o campo Número, onde selecionamos a opção singular, plural ou dual (como *lápis*, por exemplo). À esquerda, abaixo da categoria gramatical, há o campo Gênero, onde selecionamos as opções masculino, feminino ou neutro. À direita, há o campo Sigla/Acrônimo, caso o termo em análise também apresente, dentre seus exemplos, esta forma de escrita. Em seguida, abaixo, há a Entrada por extenso – o significado de uma abreviação ou acrônimo (caso o termo em análise seja uma sigla ou um acrônimo, e a sua respectiva explicitação se encontre nos exemplos), Variações morfosintáticas (como *center* e *centre* no inglês), e Acepção nº- campo usado para unidades terminológicas polissêmicas, como memória em informática (distinção feita para

software e hardware). Logo abaixo desses itens, há o registro de dados em relação ao *corpus*: dispostos lado a lado, eles são (a) Posição na ordem de frequência e (b) N° de ocorrências do termo, dados provenientes da lista de palavras do WST (veja Figura 25 abaixo).

The screenshot displays the 'Dados' (Data) tab of the VoTec - verb. interface. The interface features a top navigation bar with the following tabs: 'Dados', 'Traços Distintivos', 'Semântica', 'Termo Equivalente', 'Termos Remissivos', and 'Informações Enciclopédicas'. The 'Dados' tab is currently selected. The main content area contains a form with the following fields and values:

- Gênero:** Neutro (dropdown menu)
- Sigla/Acrônimo:** (empty text box)
- Entrada por Extenso:** (empty text box)
- Var. Morfossintáticas:** (empty text box)
- Acepção N°:** 39
- Cópus:** (empty text box)
- Posição na Ordem de Frequência:** 3
- N° de Ocorrências do termo:** 1916

The footer of the interface displays the date and time '18/08/2014 23:21' and the copyright notice '© 2007 FFLCH - ICMC Jr.'.

Figura 25. Visualização da aba Dados da página do VoTec - *verb.*

ii. Traços Distintivos: nesse campo da ficha terminográfica são inseridas as informações extraídas dos conceitos (elaborados a partir dos contextos/exemplos) que serviram à construção da definição final. Baseados na proposta de Análise Componential, de Ilari (2002), os dados foram organizados em colunas, nas quais os diferentes traços distintivos foram listados. Cada linha representa os dados provenientes de cada conceito. Todas as vezes que os traços pertencerem ao mesmo campo semântico ou forem sinônimos, eles foram listados nas mesmas colunas. Como podemos verificar na Figura 23, o pesquisador pode consultar os exemplos e conceitos elaborados, na parte superior da tela, enquanto distribui os conceitos em forma de semas na tabela (Figura 26) disponíveis na parte inferior da tela:

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
1	acquired	learners	generates	output				
2				descriptiv...				
3					formed	suffixes a...		
4					change ove...	grammar el...		
5						grammar tr...	conjugatin...	auxiliarie...
6						grammatica...		
7		individual...	to	internaliz...				

25/06/2015 11:11 © 2007 FFLCH - ICMC J

25/06/2015 11:11 © 2007 FFLCH - ICMC Jr.

Figura 26. Visualização da aba Traços Distintivos da página do VoTec.

iii. Semântica: indicamos se o termo é dicionarizado e se a definição dicionarizada é correspondente total ou parcialmente à definição em construção - o que geralmente será parcial e só serve como norte para o pesquisador, já que, por questões de direito autoral, definições já dicionarizadas não podem ser disponibilizadas na página de visualização do programa. Em seguida seleciona-se o dicionário a partir do qual provém a definição. No campo abaixo, insere-se a definição de acordo com a fonte mencionada. Isto feito, classificamos os exemplos dentro dos campos da hiperonímia, hiponímia, co-hiponímia, sinonímia e antonímia, se possível for¹⁵. Se necessário, o pesquisador pode deixar alguma **nota** que será verificada pelo administrador do sistema, conforme é possível ver no rodapé da Figura 27.

Dados	Traços Distintivos	Semântica	Termo Equivalente	Termos Remissivos	Informações Enciclopédicas
Semântica					
Termo Dicionarizado?		Definições Coincidentes?			
<input type="checkbox"/>		<input type="text"/>			
Fonte:		Escolha uma fonte			
Definição Dicionarizada:		<input type="text"/>			
Hiperônimo de:		<input type="text"/>			
Hipônimo de:		<input type="text"/>			
Co-hipônimo de:		<input type="text"/>			
Sinônimo(s):		<input type="text"/>		Antônimo(s):	
Notas:		<input type="text"/>			

Figura 27. Visualização da aba Semântica da página do VoTec.

¹⁵ De acordo com a proposta do VoTec (FROMM, 2007), todos campos de relações semânticas só podem ser preenchidos se essas relações forem traçadas dentro das linhas de concordância escolhidas a partir do *corpus*.

iv. Termo Equivalente: nessa aba, buscamos, dentro do sistema, a unidade terminológica equivalente na outra língua (inglês/português) para associá-lo à unidade terminológica em português/inglês. Esse processo só é possível se a unidade terminológica foi registrada dentro das mesmas características ontológicas do termo que está sendo registrado, conforme podemos visualizar na Figura 28 abaixo.

Dados Traços Distintivos Semântica **Termo Equivalente** Termos Remissivos Informações Enciclopédicas

Termo Equivalente

O termo equivalente deve ser um termo cadastrado em inglês, com a mesma ontologia e que ainda não esteja marcado como equivalente de outro termo.

Buscar Termo Equivalente:

- verbo

18/08/2014 23:21 © 2007 FFLCH - ICMC Jr.

Figura 28. Visualização da aba Termo Equivalente, para o termo *verb*, página do VoTec.

v. Termos Remissivos: nessa aba, podem ser resgatados os termos que se relacionam dentro do mesmo campo semântico tais quais: sinônimos, antônimos, hipônimos, co-hipônimos e hiperônimos (FROMM, 2007, p. 105). Essa relação e ligação só é possível se os termos já tiverem sido cadastrados e aprovados pelo administrador do sistema (veja Figura 29 abaixo).

3 The ACI constructions which are found in OE involve this type of verb. The verb is a three-place predicate, (while ECM verbs are two-place verbs, i.e. onotransitive) and the relevant constructions found in OE are

a three-place predicate

PDF 13/02/2013

Dados Traços Distintivos Semântica Termo Equivalente **Termos Remissivos** Informações Enciclopédicas

Termos Remissivos

Os termos remissivos devem ser termos cadastrado em português e com os respectivos cadastro já aprovados pelo administrador.

Buscar Termo Remissivo:

Termo	Apagar
Word	<input type="button" value="Apagar"/>

Figura 29. Visualização da aba Termos Remissivos da página do VoTec do termo Verbo.

vi. Informações Enciclopédicas: adiciona-se as definições provenientes de uma fonte enciclopédica, o artigo, a fonte, o *link* da informação e o tipo de fonte, não disponível na versão atual. Normalmente as informações desta aba são fornecidas, novamente por uma questão de direitos autorais, pela Wikipédia – as mesmas podem ser visualizadas na página de consulta do programa, como podemos visualizar na Figura 30 abaixo.

Definição	Artigo	Fonte	Link	Apagar
is a word (part of speech) that in syntax conveys an action (bring, read, walk, run, learn), an occurrence (happen, become), or a state of being (be, exist, stand)	verb	Wikipedia		

Figura 30. Visualização da aba Informações Enciclopédicas da página do VoTec para o termo *verb*.

vii. Conceito Final/Definição: nesse campo, o pesquisador formula as definições para a unidade terminológica selecionada em duas janelas denominadas Conceito Final e Definição. Na primeira, o pesquisador insere os dados para uma definição prévia da unidade terminológica, construída a partir dos traços distintivos. A segunda janela, a Definição, já com um texto mais encaixado dentro de parâmetros do projeto de pesquisa em questão, é a definição final que é visualizada pelos consulentes ao acessar a página do dicionário na Internet (Figura 31).

Figura 31. Visualização da aba Conceito Final, termo *verb*, na página do VoTec.

Preenchidos todos os campos, o pesquisador deve salvar as informações antes de sair do sistema. Essas informações foram enviadas ao administrador do sistema que, após avaliação, as disponibiliza para o acesso público. Do contrário elas podem ser re-editadas, a pedido do administrador, se necessário for.

Concluída a análise e aprovação do administrador, o termo é visualizado em português e inglês na página do Votec (como, por exemplo, na Figura 25), fica acessível ao público em geral, e disponibiliza visualizações alternativas de acordo com a necessidade do consulente.

The screenshot shows the VoTec interface. At the top, there is a search bar with the text 'Linguística Histórica' and a 'Buscar' button. Below the search bar, there is a sidebar with three sections: 'Tipos de Exibição' (Normal, Descritiva), 'Tipos de Consulta' (Total, Tradutor, Modular), and 'Consultas Externas' (Corpus NILC, Google, Answers.com, Wikipedia, CORTEC). The main content area displays the definition of the term 'Gramática' in Portuguese and English. The Portuguese definition is under the 'Português' tab, and the English definition is under the 'English' tab. Both definitions include a 'NOTA' (Note) and a 'Corpus' (Corpus) section. The Portuguese definition also includes a 'Hipônimo de' (Hyponym of) and a 'Hiperônimo de' (Hypernym of) section. The English definition includes a 'Hyponym of' and a 'Hypernym of' section. The interface is designed to provide a comprehensive view of the term, including its historical context and its use in various corpora.

26/06/2015 06:06 © 2007 Guilherme Fromm - ICMC Jr.
Termo elaborado por [Marcio Issamu Yamamoto](#) (pt)

Figura 32. VoTec - Visualização do termo *gramática* na área de LH.

Na figura 32, ao visualizar a tela do VoTec que é disponibilizada ao consulente, é possível observar as definições em contraste, construídas a partir dos *corpora* em português e inglês. Na definição do termo, em português, observamos que há um campo onde encontramos a palavra *NOTA*. Este campo foi criado na plataforma, como meio de suprir a necessidade de escrevermos mais de uma acepção para o termo pesquisado. É, de certa forma, uma limitação do sistema (que objetivamos sanar na construção e lançamento de uma nova plataforma) para o termo *Gramática/Grammar*.

Na próxima seção, discutiremos os resultados, as dificuldades e percursos do fazer terminográfico, específicos da plataforma do VoTec.

4. RESULTADOS

Nesta seção, objetivamos tratar das dificuldades e nuances encontradas no desenvolvimento desta dissertação, durante o fazer terminológico. Para construir a redação, seguimos a ordem dos objetivos para que pudéssemos contemplar cada um dos itens.

Para a condução do primeiro objetivo, a reelaboração da árvore de domínio da LH (BARROS, 2004), fizemos uma pesquisa bibliográfica com obras de autores referência das áreas, para que pudéssemos delimitá-las com rigor metodológico e científico adequados a esta proposta de dicionário bilíngue. Este rigor se fez necessário por dois motivos: (i) porque a Etimologia, a Filologia e Linguística Histórica são subáreas da Linguística; e (ii) devido ao fato de serem disciplinas que emitem um olhar diacrônico sobre a língua, o limiar que as separa às vezes parece ser tênue.

Na área da Filologia, pudemos analisar, a partir daquela proposta por Fromm e Yamamoto (2013), as áreas relacionadas à LH, para que traços conceituais de cada uma delas fossem delineados; pesquisamos autores como Iordan, 1982; Vidos, 1996; Basseto, 2001 e Faraco, 2005. A partir das pesquisas, entendemos que a Filologia é uma das ciências mais antigas e trabalha com a pesquisa de textos, no sentido de buscar o contexto autoral, geográfico, social e linguístico da produção de obras. Segundo Basseto (2001), o filólogo era um dos sábios que tinha um conhecimento amplo em várias áreas, era, diríamos, um profissional de caráter enciclopédico. A Filologia se interessa pelo estudo do texto, usa o método da crítica textual e o histórico comparativo para um texto, o contexto de produção e língua, no qual se insere. Além da Filologia como área, ela se subdivide em áreas menores, na tentativa de atender línguas diversas, como Filologia Clássica, Germânica, Românica, Portuguesa, dentre outras.

A Etimologia, como ciência, apresenta característica semelhante à Terminologia; pode denominar uma área da ciência ou o significado das palavras. Por isso, foi comum encontrarmos linhas de concordância que traziam ocorrências como “a etimologia da palavra x é y”. Os autores Viaro (2011) e Durkin (2009) apresentam a área e mostram que o objeto central desta disciplina é o étimo. A partir dele, os gregos acreditavam que podiam resgatar o conceito que levou os primeiros profissionais da linguagem a nomear os objetos, numa tentativa de conciliar o conceito com a referente.

A LH é uma ciência que vem a se consolidar bem depois dessas duas, mencionadas anteriormente. Ela surge com a busca dos filólogos europeus para encontrar a origem das

línguas, usando o método comparativo e tecendo as leis fonéticas que serviriam de base para explicar as mudanças linguísticas. Friedrich Diez será um dos filólogos renomados, responsável pela consolidação da disciplina na academia. Esses estudos se amadurecem, os linguistas propõem o indo-europeu e o sânscrito como língua que antecede as línguas atuais como o grego, latim, português e inglês, dentre outras. No Brasil, uma grande cientista dessa área foi Mattos e Silva, professora que se dedicou aos estudos profundos do português, juntamente com Mary Kato, e contribuiu para a consolidação da disciplina no Brasil. O objeto de estudo da LH são as mudanças fonéticas, mórfica, sintáticas e semântico-lexicais que ocorrem na língua. A partir do estudo dessas mudanças, também chamadas metaplasmos, consegue-se traçar a origem da língua e explicar suas mudanças no tempo, espaço numa sociedade.

O estudo dessas áreas nos levou a concluir que, apesar de serem áreas que dialogam entre si e apresentam muitas semelhanças, elas possuem certa distinção que as impedem de serem aglutinadas sob uma única subárea da Linguística. Isso se deu devido ao fato de elas terem estatutos diferentes, tais quais: são nomeadas de forma diferente, têm objeto, método e objetivo que são distintos entre si. Contudo a LH é uma área que absorve os dados provenientes dessas disciplinas para compor a subárea da LH, independentemente das outras.

O segundo objetivo deste trabalho foi a compilação de *corpora* bilíngue, para que servissem como banco de dados, de onde os traços conceituais foram extraídos, a partir dos contextos definitórios e explicativos. Este trabalho foi breve, devido à rapidez proporcionada pelos computadores e pela internet. Algumas dificuldades que enfrentamos foram: (1) nem todos os *corpora* disponíveis puderam ser salvos em formato txt, devido ao fato de estarem em formato PDF criptografado; logo, às vezes, fez-se necessário o uso de ferramentas cibernéticas que nos ajudam a editar os textos em PDF, convertê-los para versões legíveis pelo WST, isto é, em formato txt; (2) em inglês, a possibilidade de conseguirmos artigos na área de Etimologia foi menor, já que a política de popularização do conhecimento nos países anglófonos é mais dificilmente adotada. Isso nos obrigou a mudar a busca na internet, buscando obras diferentes daquelas propostas no projeto original. Optamos por aumentar o *corpus* de inglês com um dicionário de Etimologia, que traz definições e comparações de diferentes propostas de etimologias dos verbetes, com um parecer final do autor. Este caráter comparativo da obra, permitiu que houvesse um

material não só de cunho lexicográfico, mas também enciclopédico e também acadêmico-científico, o que contribui para a diminuição de interferências na busca por contextos definitórios. Tomamos o cuidado de manter os *corpora* balanceados em números de ocorrências (aproximadamente 500 mil), apesar de não ter sido possível o mesmo balanceamento em número de textos. O que já foi explicado na seção de Metodologia (3.2).

A extração de candidatos a termos, provenientes das listas de palavras-chave, após processamento dos *corpora* pelo WordSmith Tools 6.0 (SCOTT, 2012), foi um passo dado a partir da lista em português, pelo fato de o nosso público alvo ser o brasileiro, em primeiro lugar. Isto é, partimos dos termos selecionados pela lista de palavras-chave do português, para depois buscá-los na mesma lista em inglês. Nos casos em que não foi encontrado um termo correspondente no inglês, fizemos a busca na lista de palavras dessa língua. Uma ocorrência que nos chamou a atenção foi o caso do termo *ablaut* em inglês, traduzido como apofonia em português. Este foi um termo bastante recorrente no inglês, aparecendo 157 vezes, em três textos, correspondendo a uma porcentagem de 0,02% do *corpus*. Em português, o termo apofonia aparece 6 vezes na lista de palavras, não tendo a mesma correspondência de porcentagem no texto¹⁶.

A disponibilização de dados oriundos de fontes confiáveis aos usuários concretizou-se por meio da plataforma, já que as definições foram construídas a partir de traços semânticos provenientes de *corpora* acadêmicos. Esta foi uma das grandes vantagens proporcionadas pela Linguística de *Corpus*, adotada como metodologia e como abordagem neste trabalho. Caso não a tivéssemos adotado, com certeza, o passo de confirmação do termos, que teria sido feito com especialistas, comprometeria a conclusão desta pesquisa, já que não haveria como certificar-nos de que o termo pertenceria ou não à LH. Outra limitação seria saber se o candidato a termo seria ou não uma palavra-chave; também não teríamos como localizar contextos definitórios e explicativos para a construção de definições adequadas para os consulentes. Os *corpora* coletados são todos provenientes da comunidade científica, conforme mencionado e especificado na seção de metodologia.

A organização de dados, na qual incluímos o termo, a área a qual pertence, a classe gramatical, a definição, os exemplos do *corpus*, entre outros, foi feita em fichas de registro,

¹⁶ Devido à baixa ocorrência do termo, cuja frequência é menor que 0,01%, esta porcentagem não é disponibilizada pelo WST 6.0.

e os dados dispostos nas macro e microestrutura do vocabulário. Primeiramente fizemos a análise componencial (ILARI, 2002), para que pudéssemos organizar os dados nas fichas terminológicas do VoTec. Além dos dados referentes ao termo em si, agregamos a definição de dicionários e da Wikipédia, de modo que o consulente pudesse acessar as informações de formas diversas, ou seja, customizar a plataforma com suas opções de visualização, para que o uso fosse prático e de rápido acesso. Este passo objetivou disponibilizar aos leitores uma obra de leitura acessível, organizada e padronizada. Isto se fez necessário, principalmente ao considerarmos alunos ingressantes de graduação, que ainda não dominam a terminologia da Linguística e podem não acompanhar os conceitos que jazem por detrás das palavras. Esta é uma das vantagens do VoTec, já que as definições são redigidas de forma curta, objetiva e restritas a uma oração.

A elaboração de definições para os termos selecionados foi baseada no padrão GPDE, gênero próximo e diferença específica, adotado nas obras terminológicas em nível nacional e internacional. Esta definição partiu das noções mais amplas, da relação de hiperonímia, para aquelas mais específicas, as relações de hiponímia (ILARI, 2002, p.39). Ao analisar os padrões adotados pela academia canadense, foi possível observar a importância do padrão GPDE, haja vista a importância do Canadá como país bilíngue, bem como referência mundial em terminologia bilíngue e em teóricos da Lexicologia, Lexicografia, Terminologia e Terminografia.

A disponibilização dos resultados no VoTec, foi feita sem dificuldades, já que esta plataforma se encontra em funcionamento há mais de 7 anos. Nosso objetivo era disponibilizar os dados em uma nova plataforma (ainda em construção), onde cada projeto seria mais personalizado; contudo, a plataforma ainda apresenta problemas de funcionamento, o que nos fez postergar os planos para seu uso. Nessa nova plataforma, o campo NOTAS fará parte de sua estrutura, sem que precisemos criar esse campo juntamente com a definição principal. Quanto ao processo de ligar o termo principal aos termos remissivos, tivemos a seguinte limitação: os termos polissêmicos e recorrentes em várias subáreas da Linguística eram disponibilizados pela plataforma para serem escolhidos como remissivos. Por exemplo, ao definir o termo **gramática**, um dos termos remissivos era **língua**. Ao solicitarmos a busca deste termo no sistema, todos os termos **língua**, pertencentes a várias subáreas da Linguística eram disponibilizados pelo sistema. Não havia como identificar qual deles era o termo pertencente à LH. Logo para sanar esta

limitação, tivemos que descadastrar todos os termos **língua**, pertencentes às outras subáreas da Linguística, e deixar somente aquele pertencente à subárea da LH. Dessa forma, conseguimos cadastrar o termo como remissivo da LH, ligado ao termo gramática.

Quanto à escrita e constituição da definição, um desafio a ser enfrentado foi conseguir evitar uma definição que fosse muito técnica¹⁷ e incompleta. Na construção da definição do termo *palavra*, deparamo-nos com a seguinte possibilidade: **unidade que expressa um conceito abstrato, composta por uma base e um sufixo**. Contudo, ao ser lida por um consulente, com certeza a definição seria muito abstrata, e o sentido do termo *palavra* ficaria confuso e obscuro. Como NOTA teríamos o seguinte: **nó formado a partir do mesmo afixo/base que identifica as estruturas morfológicas formadas pela adjunção dos sufixos de flexão e de número para os adjetivos e nomes**. Para o público-alvo do VoTec, que seria alunos de Tradução, alunos de Letras e tradutores (FROMM, 2007), o objetivo final seria somente o de checar a correspondência dos termos na duas línguas: inglês e português. O recurso da inserção da nota foi motivado pelas leituras do padrão terminográfico proposto por Bevilacqua e Finatto (2006) para que a definição não ficasse muito longa. Além de longa, uma definição redigida por orações separadas por ponto e vírgula seria confusa e não desejável para a compreensão do aspecto conceitual do termo.

O termo *etimologia* foi difícil de ser definido, apesar da recorrência de quase 50 vezes do termo, no *corpus* de estudo. O termo, assim como acontece com a combinação terminologia vs. Terminologia, além de ser usado como subárea da Linguística, também é usado como a etimologia do termo X ou Y. Neste trabalho, a segunda opção foi mais recorrente no *corpus* de inglês. No *corpus* de português ela foi mencionada como o nome da ciência mais vezes que no *corpus* de inglês. Observamos então que, como ciência, ela se coloca mais vezes, no caso do português, que no caso do inglês. Isso mostra certa consolidação da disciplina em inglês, diferentemente do português. Chegamos a esta conclusão uma vez que o termo recorria no *corpus* de inglês sem que estivesse inserido em contextos definitórios ou explicativos. As ocorrências, em sua maioria, referem-se a etimologia de dado termo e não à explicação ou definição do termo etimologia. Essa

¹⁷ A expressão “muito técnico” se explica pelo fato de o nosso público alvo ser de alunos de graduação e tradutores que não dominam o sistema linguístico inglês, nem a área científica (LH) muito bem.

característica induz à compreensão de que, para o leitor de fala inglesa, o termo já está difundido sem que sua explicação ou definição se faça necessária.

Nesse caso, observamos que a bibliografia disponível em português brasileiro, para o estudo da Etimologia como ramo da ciência linguística, é mais limitada. A ocorrência foi em quinze textos de português e cinco de inglês. Em inglês, o termo recorreu 783 vezes, enquanto que no português ele foi recorrente 294 vezes. Em contrapartida, o que observamos no *corpus* em inglês é que os estudos de Etimologia estão mais consolidados, provavelmente pela tradição filológica, etimológica e linguística mais antiga e tradicional na Europa que no Brasil.

Uma das abas das fichas terminológicas disponibiliza um espaço onde cadastramos as relações de sinonímia, hiponímia, hiperonímia, co-hiponímia, antonímia, traçadas dentro do texto, das linhas de concordância dos *corpora*. Seguem abaixo alguns exemplos encontrados nos *corpora*.

1. Sinonímia:

Etimologia = origem dos vocábulos

Seu autor foi Santo Isidoro de Sevilha (c.560-636AD) que, por ter sido o compilador da primeira base de dados do Ocidente, é hoje considerado o padroeiro da Internet. “A etimologia é a origem dos vocábulos, cuja força (de significação), quer em nomes, quer em palavra é desvendada pela interpretação,” afirma o bispo da igreja católica.

2. Hiponímia:

Gramática > Língua¹⁸

A definição de Gramática Histórica de Francisco J. Martins Sequeira (1938b) é mais completa no segundo trabalho publicado, salientando-se neste que no estudo da evolução de uma língua devem ser estudados todos os fenómenos e que se deve ter em consideração tanto o eixo temporal como o eixo espacial.

3. Hiperonímia

Gramática < sons e letras¹⁹

Na análise que faz desta corrente gramatical filosófica, Duarte (1995: 15) considera que a "ideia de que as categorias gramaticais reflectem elementos e processos mentais universais, mas de que é necessário analisar cada língua particular porque tais processos se podem exprimir diferentemente de língua para língua, o afastamento do objectivo normativo (a gramática é a arte de falar e não a arte de bem falar), a distinção clara entre sons e letras e a prioridade atribuída à análise dos primeiros são características importantes e

¹⁸ Usamos o símbolo > para indicar que o termo gramática é hipônimo de língua.

¹⁹ Usamos o símbolo < para indicar que o termo gramática é hiperônimo de sons e letras.

singularmente contemporâneas desta escola de pensamento que Chomsky, na década de 60, invoca centralmente ao procurar os fundamentos da gramática generativa na tradição racionalista de teorização e análise das línguas naturais".

4. Co-hiponímia

Verbo = substantivo; advérbios; conjunção de coordenação; adjetivo.²⁰

Outros tipos de predicação envolvem, diga-se de passagem verbo e o que vem a seguir, substantivo ou adjetivo, que, juntos com o verbo remetem ao sujeito, por exemplo: eu sinto dor, em que sinto dor refere-se ao sujeito, não apenas sinto.

5. Antonímia

Dentro da lista de candidatos a termos, extraída de uma lista de palavras-chave de 500 palavras, não encontramos contextos que trouxessem esta ocorrência.

Um aspecto interessante encontrado durante o processamento de dados foi o termo **Futuridade**, termo com contexto definitório e explicativo, presente em um único texto no *corpus*, somente em português, sem correspondência no inglês.

Quanto à disponibilização de dados para o consulente, um detalhe observado foi quanto à sinonímia. Na presença de mais de um sinônimo, a plataforma só permite a visualização de alguns termos. Para que haja a visualização de todos os termos, é necessário solicitar o modo descritivo de visualização.

As relações semânticas entre termos, às vezes, são difíceis de serem traçadas, mesmo analisando-se o contexto, como foi o caso da tentativa de extração dos traços conceituais para elaborar a definição da expressão *linguistic composition*, como podemos ver no excerto abaixo:

Once one child converges on some new grammar, then the linguistic composition of the population will change, because the child with the new grammar produces different structures; this may have domino effects.

Todas essas experiências e tentativas contribuíram para uma melhor compreensão do fazer terminológico, da compreensão do processo de produção de sentidos para a identificação de traços semânticos e construção de definições.

²⁰ Usamos o símbolo = para indicar a co-hiponímia entre termos.

Abaixo podemos ver os termos que compuseram o Vocabulário de LH, produto final desta pesquisa. Podemos visualizar as duas colunas, a da esquerda com os termos em português, e a da direita, com os termos em inglês.

Quadro 6. VoTec - Termos do Vocabulário bilíngue de LH.

VoTec – Vocabulário de LH – lista final.	
Português	Inglês
1. Língua	Language
2. Etimologia	Etymology
3. Gramática	Grammar
4. Palavra	Word
5. Verbo	Verb
6. Linguística	Linguistics
7. Linguística Histórica	Historical Linguistics
8. Mudança	Change
9. Fala	Speech
10. Linguagem	Language
11. Texto	Text
12. Variação	Variation
13. Derivação	Derivation
14. Substantivo	Noun
15. Adjetivo	Adjective
16. Prefix	Prefix
17. Léxico	Lexicon
18. Analogia	Analogy
19. Clítico	Clitic
20. Advérbio	Adverb
21. Particípio	Participle
22. Sufixação	Suffixation
23. Prefixação	Prefixation
24. Afixo	Affix
25. Mudança linguística	Linguistic change
26. Língua românica	Romance Language

Os termos apresentados no Quadro 6 foram selecionados por terem equivalência conceitual e por serem termos recorrentes na área de LH. Dentre os termos, salientamos aqueles que trazem conceitos mais usados nessa área, bem como nas áreas que a alimentam, a Etimologia e a Filologia; termos tais quais: língua, etimologia, linguística histórica, mudança, mudança linguística, variação, língua românica, analogia, sufixação e prefixação.

5. CONSIDERAÇÕES FINAIS

A partir da pesquisa desenvolvida e da prática terminológica proporcionada pela plataforma do VoTec, podemos observar alguns resultados quanto ao que vem a ser a produção de uma obra terminográfica. Primeiramente foi necessário conhecer o que é o campo ou quais são os limites da Terminologia e Terminografia. Apesar da prática já estar instituída, a área em si ainda é estranha ou nova para linguistas de outras subáreas da Linguística, o que pode levar a uma confusão entre a Terminografia e a Lexicografia como áreas que trabalham com o registro do léxico. Daí a importância da diferenciação conceitual dos objetos dessas disciplinas, quais sejam a unidade lexical versus a unidade terminológica.

Estar inscrito em uma escola que norteia os conceitos e princípios a serem seguidos e defendidos também é crucial no fazer científico da academia. A TCT permite que entendamos a unidade terminográfica que pode se identificar como unidade lexical ou terminológica, dependendo do seu contexto comunicativo de uso. A LC, enquanto metodologia, que parte da visão probabilística da língua, possibilita ao seu usuário o manejo de dados concretos, e significativos quanto à representatividade e quanto ao uso em *corpora* autênticos.

Os resultados finais obtidos para a elaboração da árvore de domínio foram que a área da LH é difícil de ser definida como uma área estanque. Isso se deve ao fato de que ela se alimenta de subáreas da Linguística, principalmente como *corpus*, e de outras para fazer sua análise de dados. Isto é, falar em LH significa envolver a Etimologia, a Filologia e a Linguística Diacrônica. Quanto às unidades terminológicas, observamos que nos *corpora* de português e inglês há termos correspondentes que servem de candidatos a termos. Há também aqueles que trazem consigo traços semânticos em uma língua e não em outras, o que nos impossibilitou de manter os candidatos a termos nas listas de termos a serem inseridos na plataforma. Os *corpora* nos permitiram ver que a LH vive momentos diferentes no Brasil e Europa, o que já era esperado, já que grande parte da tradição filológica e linguística ocidental iniciou-se na Europa, mais especificamente na Alemanha, berço da língua inglesa, e na Itália e França origem dos estudos de Linguística e Filologia Românica.

A aplicabilidade que podemos dar a esta pesquisa é a condução de projetos em áreas de Letras e das Ciências Agrárias e Biológicas, bastante fortes na instituição onde atua o pesquisador. Dominados e aprimorados os procedimentos do fazer terminológico e terminográfico, desde a delimitação das áreas por meio da construção da árvore de domínio até a disponibilização dos dados na plataforma do VoTec, a construção de novos vocabulários técnicos se tornará uma realidade.

A plataforma do VoTec vem comprovar a evolução tecnológica no fazer terminográfico. Ela traz vantagens e desvantagens quanto ao fazer terminográfico, mas o ganho é maior que as perdas e, uma vez que uma nova plataforma está sendo concebida, as limitações desta versão serão certamente aprimoradas. É uma plataforma que possibilita a organização eficaz de dados, por meio das fichas virtuais, de forma que a rapidez com que se conclui o fazer definitório seja incomparavelmente maior em relação ao modo como era feito antigamente. Outro aspecto relevante é a disponibilidade dos dados, bem como seu alcance em nível mundial, gratuitamente, via rede mundial de computadores. O Vocabulário Técnico de Linguística Histórica possibilita a compreensão de termos técnicos por parte dos tradutores, pesquisadores e docente que estão envolvidos com a LH e com a pesquisa científica da história das mudanças linguísticas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, G. M. B., PINO, D. H. P., SOUZA, D. S. L. **A definição nos dicionários especializados**: proposta metodológica. RITerm – Debate Terminológico, n. 3, janeiro 2007. Disponível em: <http://www.riterm.net/revista/n_3/Art_Barcellos_Almeida.pdf>. Acesso em: 15 abril 2007.

ANTHONY, L. **AntConc** (Version 3.2.2) [Computer software]. Tokyo, Japan: Waseda University, 2011. Disponível em: <<http://www.antlab.sci.waseda.ac.jp/>>.

AUBERT, F. H. **Introdução à metodologia da pesquisa terminológica bilíngue**. São Paulo: Humanitas, 1996.

BARBOSA, M. A. Lexicologia, Lexicografia, Terminologia, Terminografia, Identidade científica, Objeto, Métodos, Campos de atuação. In: **Simpósio latino-americano de terminologia e I Encontro brasileiro de terminologia técnico-científica**. 2., 1990, Brasília, Anais. p. 152-158.

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

BASSETO, B. F. **Elementos de filologia românica**: história externa das línguas. São Paulo: EDUSP, 2001, p.17-42.

BÉJOINT, H. **The Lexicography of English**. From Origins to Present. Oxford: Oxford University Press, 2010.

_____. **Modern Lexicography** – an introduction. Oxford: Oxford University Press, 2004.

BEVILACQUA, C. R.; FINATTO, M. J. B. **Lexicografia e Terminografia**: alguns contrapontos fundamentais. São Paulo: Alfa, 2006. Disponível em: <<http://www.ufrgs.br/textecc/textquim/arquivos/03-Bevilacqua-Finatto.pdf>>. Acesso em 20 jul. 2015.

BLIKSTEIN, I. (Dir. e Coord. Geral de Trad.). **Dicionário de Linguística**, SP: Ed. CULTRIX LTDA, 2004.

CABRÉ, M. T. **La Terminología**: representación e comunicación. Barcelona: IULA/UPF, 1999.

_____. El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en terminología (I). In: **Organon**. Universidade Federal do Rio Grande do Sul, Instituto de Letras, vol.1, n.1 (1956). Porto Alegre: UFRGS, semestral, 2011.

CALLADO, A. A. O texto em veículos impressos. In: CALDAS, A. (org.). **Deu no jornal**: o jornalismo impresso na era da internet. São Paulo: Loyola, 2002.

CÂMARA JR, M. **Dicionário de Linguística e gramática**: referente à língua portuguesa. Petrópolis: Vozes, 1986.

CAVALCANTI, C. B. O. Moderna Perspectiva Das Classes De Palavras. In: **Cadernos do CNLF**, Série VIII, nº03, 2004. Disponível em < <http://www.filologia.org.br/viiicnlf/anais/caderno03-15.html> >. Acesso em 20 maio 2015.

CREMONESE, L. E. **Bases epistemológicas para a elaboração de um dicionário de Linguística da Enunciação**. Dissertação (Mestrado). Universidade Federal do Rio Grande do Sul. Instituto de Letras, Programa de Pós-Graduação em Letras. Orientação: Profª Drª Maria José Bocorni Finatto. Porto Alegre, 2007. Disponível em < <https://www.lume.ufrgs.br/bitstream/handle/10183/11399/000611386.pdf?sequence=1> >. Acesso em ago. 2015.

CRYSTAL, D. **Dicionário de Linguística e Fonética**. Trad. e adap. DIAS, Maria Carmelita Pádua. RJ: Jorge Zahar Ed., 2000.

DAVIES, M. COCA – **The Corpus of Contemporary American English [corpus]**. Provo, UT, USA: Brigham Young University, 2008. Disponível em: <<http://corpus.byu.edu/coca/>>.

DAVIES, M.; FERREIRA, M. J. (2006-) **Corpus do Português**: 45 million words, **1300s-1900s**. Disponível em < <http://www.corpusdoportugues.org/> >. Acesso em 25 jun. 2015.

DORNELLES, M. S. **Bases teórico-metodológicas para elaboração de um glossário bilíngue (português-inglês) de treinamento de força**: subsídios para o tradutor. Disponível em < <http://hdl.handle.net/10183/117567> >. Acesso em 15 jun. 2015.

DUBOIS, J. *et al.* **Dicionário de linguística**. 16ª ed. São Paulo: Cultrix, 1999. Editorial, 2002.

DURKIN, P. **The Oxford Guide to Etymology**. New York: Oxford University Press, 2009.

FARACO, C. A. **Linguística Histórica**: uma introdução ao estudo da história das línguas. São Paulo: Parábola editorial, 2005.

FINATTO, M. J. B. A definição terminológica do dicionário TERMISUL: expressões linguísticas de relações conceptuais complexas. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (Orgs.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. Campo Grande, MS: Ed. UFMS, 2001. p. 211-223.

_____. 2001. **Definição terminológica**: fundamentos teórico-metodológicos para sua descrição e explicação. Porto Alegre: UFRGS. Tese de doutorado.

_____. **Elementos Lexicográficos e Enciclopédicos na Definição Terminológica**: Questões de Partida. Porto Alegre, RS: Organon, 1998. Disponível em < <http://seer.ufrgs.br/index.php/organon/article/view/29563> >. Acesso em 01 ago. 2015.

FLORES, V. do N. [et. al.] **Dicionário de Linguística da enunciação**. São Paulo: Contexto, 2009.

FROMM, G. **VoTec**: a construção de vocabulários eletrônicos para aprendizes de tradução. São Paulo, 2007. Tese (Doutorado em Estudos Linguísticos e Literários em Língua Inglesa). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

_____. A questão da taxonomia num *corpus* colaborativo para construção de um vocabulário na área de linguística. In: **Anais do SILEL**. Volume 3, Número 1. Uberlândia: EDUFU, 2013.

_____. **Vocabulário de Linguística**: treinamento em Terminografia Bilíngue, uso de *corpora* e ambiente de gestão terminológica. Disponível em <<http://150.164.100.248/gtlexNovo/>>. Acesso em 10 jun. 2015.

FROMM, G.; YAMAMOTO, M. I. Terminologia, Terminografia, Tradução e Linguística de *Corpus*: a criação de um vocabulário bilíngue sobre Linguística. In: TAGNIN, S.; BEVILACQUA, C. **Corpora na Terminologia**. São Paulo: Hub Editorial, 2013.

HARTMANN, R. R. K.; JAMES, G. **Dictionary of Lexicography**. New York: Routledge, 2002.

ILARI, R. **Introdução ao estudo do léxico** - brincando com as palavras. São Paulo: Contexto, 2002.

IORDAN, I. **Introdução à Linguística Românica**. Trad. De Júlia D. Ferreira. Lisboa: Fundação Calouste Gulbenkian, 1982.

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

KYTÖ, M. *Corpora* and historical linguistics. In: **Rev. bras. linguist. apl.**, Belo Horizonte, v. 11, n. 2, 2011. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-63982011000200007&lng=en&nrm=iso>. Acesso em 18 ago. 2014. <http://dx.doi.org/10.1590/S1984-63982011000200007>.

LANDAU, S. **Dictionaries**: The art and Craft of Lexicography, 2nd. Edn. Cambridge: CUP, 2001.

MATTOS E SILVA, R. V. **Caminhos da linguística histórica** – “ouvir o inaudível”. São Paulo: Parábola Editorial, 2008.

MOISÉS, M. **Dicionário de termos literários**. SP:Ed. CULTRIX LTDA, 1995.

MUSSALIM, F.; BENTES, A. (Orgs.). **Introdução à linguística: domínios e fronteiras**. Vol. 1. São Paulo: Cortez, 2001.

REY-DEBOVE, J. **Etude Linguistique et sémiotique des dictionnaires français contemporains**. The Hague: Mouton, 1971.

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

SCOTT, M. **WordSmith Tools**. Versão 6, 2012. Disponível em: <<http://lexically.net/wordsmith/version5/>>. Acesso em 23 mar. 2013.

SILVA, M. C. P. Lexicografia bilíngue: uma verificação dos substantivos mais frequentes em dicionários bilíngues francês-português e português-francês. In: LONGO, B. N. de O.; SILVA, B. C. (Org.). **A construção de dicionários e de bases de conhecimento lexical**. Araraquara: Ed. Cultura Acadêmica, 2006, p. 13-44.

TAGNIN, S. E. O. *Corpus-driven terminology in Brazil*. In: Armelle Le Bars Poupet; Claudia Xatara. (Org.). **Cahiers de lexicologie - Dynamique de la recherche en lexicologie, lexicographie et terminologie au Brésil**. 101ed.Paris: Classiques Garnier, 2012, v. 2012-2, p. 169-182.

TARALLO, F. L. **Tempos linguísticos: itinerário histórico da língua portuguesa**. São Paulo: Editora Ática, 1990.

TEIXEIRA, E. D. **A Linguística de Corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual**. São Paulo, 2008. Tese (Doutorado – Programa de Pós-Graduação em Estudos Linguísticos e Literários em Inglês – Departamento de Letras Modernas). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

VIANA, V.; TAGNIN S. E. O. **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010.

_____. **Corpora na tradução**. São Paulo: Hub Editorial, 2015.

VIARO, M. E. **Etimologia**. São Paulo: Contexto, 2011.

_____. **Morfologia Histórica**. São Paulo: Cortez, 2014.

VIDOS, B. E. **Manual de linguística românica**. Trad. de José Pereira da Silva. Revisão técnica de Evanildo Bechara e Marlit Cavalcante Bechara. Rio de Janeiro: Eduerj, 1996.

WILKENS, Rodrigo; PEREIRA BOCORNY, Ana Elisa; KRAUSE KILIAN, Cristiane; VILLAVICENCIO, Aline. Ambientes web de gestão terminológica para a criação de produtos terminológicos on-line. In: **Debate Terminológico**. No. 08, Jun. 2012; pp. 16-22.

YAMAMOTO, M. I. . Da etimologia à linguística histórica: considerações diacrônicas sobre o estudo da lingua(gem). In: **IV Congresso Internacional de História: Cultura**,

sociedade e poder. Anais Eletrônicos. JATAÍ: Universidade Federal de Goiás/Campus Jataí, 2014. p. 1-14. Disponível em <<http://www.congressohistoriajatai.org/2014/anais2014.html>>. Acesso em 14 ab. 2015.

_____. Linguística Histórica e Linguística de *Corpus*: uma proposta para dicionário bilíngue português-inglês. In: **Revista Philologus**. Rio de Janeiro, Ano 19, n.55 Suplementos, p.691-705. Disponível em <<http://www.filologia.org.br/revista/55supl/55supl.html>>. Acesso em 15 set. 2013.

WEEDWOOD, B. **História concisa da linguística**. São Paulo: Parábola Editorial, 2002.

APÊNDICES

Apêndice 1. Lista de palavras-chave em português.

N	Key word	Freq.
1	CARTA	2541
2	PORTUGUÊS	2083
3	LÍNGUA	1960
4	PALAVRAS	1632
5	FORMA	1456
6	SÉCULO	1159
7	SUFIXO	1154
8	VERBO	1109
9	LATIM	962
10	EXEMPLO	938
11	FUTURO	888
12	PALAVRA	855
13	DOCUMENTO	841
14	FORMAS	837
15	AMIGO	809
16	NOME	803
17	BARÃO	778
18	PORTUGUESA	742
19	MARGEM	735
20	ESCRITA	710
21	FORMAÇÃO	692
22	LIVRO	680
23	LÍNGUAS	662
24	REMETENTE	649
25	SUFIXOS	621
26	NOMES	616
27	AUTOR	590
28	LINGÜÍSTICA	583
29	GRAMÁTICA	563
30	ORIGEM	552
31	MUDANÇA	549
32	CARTAS	542
33	LETRA	524
34	NOTAS	464
35	EXEMPLOS	464
36	LETRAS	462
37	VERBOS	451
38	FALA	443
39	LINGUAGEM	439
40	SIMPLES	438
41	TENHO	436
42	DERIVADOS	428
43	TEXTOS	418

44	NASCIMENTO	412
45	VARIAÇÃO	402
46	VOGAL	400
47	DERIVAÇÃO	376
48	SEÇÃO	356
49	FONTES	355
50	SUBSTANTIVO	354
51	PAUTAS	344
52	SUJEITO	326
53	CAPÍTULO	307
54	COMPADRE	298
55	GRAMÁTICAS	294
56	ETIMOLOGIA	294
57	VOGAIS	288
58	OCORRÊNCIAS	264
59	PROFISSÃO	264
60	OBSERVAÇÕES	264
61	COMPLETO	264
62	VOCÁBULOS	263
63	INFINITIVO	252
64	VARIANTES	252
65	PESSOAS	252
66	GRAMÁTICOS	251
67	PASTA	251
68	TUCANO	244
69	NATURALIDADE	242
70	ORAÇÕES	241
71	ORTOGRAFIA	233
72	FILIAÇÃO	233
73	NACIONALIDADE	231
74	PRÓCLISE	230
75	FILOLOGIA	228
76	FALECIMENTO	225
77	LEÃO	222
78	LINGÜÍSTICA	221
79	VARIANTE	221
80	SUBSTANTIVOS	219
81	PATERNOS	218
82	PERÍFRASE	217
83	LATINOS	214
84	ÊNCLISE	211
85	ACEPÇÃO	210
86	SÉCULOS	201
87	HISTÓRICOS	193
88	CONSELHEIRO	191
89	CONSTRUÇÕES	186
90	COUTINHO	184
91	BACHAREL	174
92	LÉXICO	174
93	TUPI	173

94	CONTEXTOS	173
95	ACENTO	168
96	REMETENTES	165
97	PERIFRÁSTICO	158
98	HISTÓRICAS	158
99	FINITO	154
100	SEMÂNTICA	152
101	ANALOGIA	148
102	PRONOMES	146
103	MORFOLOGIA	146
104	GRAFIA	143
105	RUBRICA	142
106	PLURAL	138
107	TÓPICO	138
108	PRONÚNCIA	137
109	CULTO	137
110	PRONUNCIACÃO	136
111	ADESIVA	134
112	GRAMATICALIZAÇÃO	133
113	FONÉTICA	131
114	SÍLABA	130
115	NOTÍCIAS	129
116	GRAFISMO	125
117	FALANTES	123
118	GRAMATICAL	122
119	DATAÇÃO	121
120	CONSOANTES	119
121	FAMÍLIA	119
122	GRAMÁTICO	118
123	VOCÁBULO	118
124	ANOTAÇÕES	118
125	MICROESTRUTURA	113
126	ADJETIVO	111
127	FALANTE	109
128	DICÇÕES	108
129	SÍLABAS	105
130	LINGÜÍSTICOS	104
131	MANUSCRITO	104
132	TIMBRE	101
133	LINGUA	98
134	NOMINAIS	98
135	PRONOME	97
136	AUTÓGRAFO	96
137	PARÁFRASE	93
138	PREFIXOS	90
139	CLÍTICOS	89
140	ENCLISE	88
141	ACEPÇÕES	88
142	ANOTAÇÃO	88
143	SINTAXE	87

144	LEXICAIS	85
145	DATADA	84
146	CLÍTICO	80
147	ADVÉRBIOS	80
148	GALEGO	79
149	ADVÉRBIO	78
150	DIACRÔNICA	77
151	DICIONÁRIOS	76
152	PREPOSIÇÃO	75
153	DITONGO	73
154	PARTICÍPIO	69
155	ALTERNÂNCIAS	69
156	CASTELHANO	68
157	CLÍTICOS	64
158	SUFIXAÇÃO	63
159	LINGUÍSTICO	62
160	LINGUISTAS	60
161	FUTURIDADE	59
162	DECLARATIVAS	59
163	ORTOGRÁFICA	59
164	HISTORIÓGRAFO	58
165	CLÍTICO	57
166	FONÉTICAS	57
167	NEGOCIO	57
168	DITONGOS	56
169	SUFIXAIS	55
170	PREFIXAÇÃO	55
171	COGNATOS	53
172	DIACRONIA	53
173	ETIMOLÓGICA	50
174	DEVERBAIS	49
175	SEMIVOGAL	48
176	ACUSATIVO	46
177	CULTISMO	45
178	AFIXOS	44
179	MUDANCA	42
180	DERIVACIONAL	37
181	PARÁFRASES	37
182	ROMÂNICA	32
183	CULTISMOS	32
184	FILÓLOGOS	32
185	ORTÓGRAFOS	31
186	LINGUISTICA	31
187	FILOLÓGICA	31
188	GLOTOLOGIA	29
189	METAPLASMOS	29

Fonte: autor.

Apêndice 2. Lista de palavras-chave em inglês.

N	Key word	Freq.
1	LANGUAGE	2518
2	LANGUAGES	2170
3	WORD	1799
4	ENGLISH	1688
5	CHANGE	1330
6	FORM	1266
7	MEANING	1213
8	VERB	1150
9	FORMS	1140
10	CASE	1105
11	GERMANIC	855
12	ORIGIN	794
13	ETYMOLOGY	783
14	LINGUISTIC	762
15	DEVELOPMENT	757
16	ROOT	728
17	CENTURY	706
18	LATIN	647
19	LINGUISTICS	620
20	SUBJECT	570
21	SOUND	560
22	PRESENT	543
23	ORDER	542
24	VOWEL	536
25	NUMBER	532
26	HISTORICAL	528
27	STRUCTURE	526
28	ORIGINAL	512
29	MIDDLE	486
30	DIALECTS	485
31	NOTE	478
32	CONSTRUCTION	478
33	SEMANTIC	444
34	POINT	430
35	NOUN	421
36	SPEAKERS	414
37	CLAUSES	412
38	POSITION	410
39	CONTACT	409
40	LEXICAL	403
41	SENSE	399
42	GRAMMAR	381
43	CASES	377
44	COGNATES	349
45	PRONOUNS	347

46	SUFFIX	345
47	OBJECT	321
48	STUDIES	315
49	FORMATION	310
50	GRAMMATICALIZATION	300
51	INFLUENCE	285
52	SLANG	282
53	TEXTS	278
54	TERM	277
55	NOUNS	273
56	DICTIONARIES	271
57	DICTIONARY	270
58	COGNATE	269
59	PHONETIC	266
60	BORROWING	265
61	MEANINGS	263
62	CONSTRUCTIONS	261
63	ETYMOLOGICAL	260
64	VOWELS	256
65	DIALECT	256
66	FUNCTION	256
67	SANSKRIT	249
68	VARIETIES	244
69	RECONSTRUCTION	243
70	ROMANCE	243
71	AUXILIARY	228
72	SINGULAR	225
73	RESEARCH	224
74	PRONOUN	220
75	CLAUSE	216
76	ETYMON	215
77	TRANSITIVE	214
78	VARIANTS	213
79	PARADIGM	212
80	FACTORS	210
81	VARIANT	210
82	PATTERN	208
83	DERIVATION	204
84	ADJECTIVE	204
85	SYLLABLE	201
86	LINGUISTS	199
87	SYNTAX	194
88	CREOLES	191
89	GENITIVE	187
90	REFLEXIVE	187
91	CONSONANT	184
92	MORPHOLOGY	182
93	ROOTS	174
94	ETYMOLOGIES	172
95	INFLECTION	172

96	DATIVE	171
97	EPISTEMIC	168
98	REANALYSIS	167
99	PARTICIPLE	160
100	NOMINATIVE	159
101	ABLAUT	157
102	PHONOLOGY	155
103	ADJECTIVES	154
104	INTRANSITIVE	153
105	CONSONANTS	146
106	PREFIX	145
107	REGULAR	141
108	PREASPIRATION	139
109	SUBJECTS	138
110	STAGES	138
111	SPACE	132
112	ACCUSATIVE	131
113	CLITIC	130
114	TENSE	129
115	PRONUNCIATION	127
116	PREFIXES	123
117	SUFFIXES	120
118	CONJUNCT	119
119	GRAMMARS	118
120	ALTERNATION	118
121	INFINITIVE	116
122	COMPOUNDS	113
123	FACTOR	112
124	POSSESSIVE	112
125	BASE	111
126	SPELLING	109
127	CREOLIZATION	108
128	DECLENSION	106
129	ABSTRACT	105
130	AUXILIARIES	102
131	VOICELESS	102
132	ADVERB	98
133	PROGRESSIVE	98
134	SEMITIC	97
135	PREPOSITION	96
136	SYNCHRONIC	94
137	ADVERBIAL	92
138	UNSTRESSED	91
139	CORRESPONDENCES	88
140	DELETION	88
141	ASPIRATE	85
142	GRADIENCE	82
143	MORPHEMES	82
144	PREDICATE	81
145	ACCORDANCE	79

146	INFLECTED	76
147	I-LANGUAGES	75
148	MORPHEME	74
149	PRETERITE	74
150	GLOSSES	74
151	IMPERFECTIVE	73
152	GLOSSED	73
153	PERFECTIVE	73
154	PREPOSITIONAL	73
155	DITRANSITIVE	69
156	PERIPHRASTIC	69
157	COREFERENTIAL	68
158	COPULA	68
159	DEGRAMMATICALIZATION	63
160	GRAMMATICALIZED	62
161	GRAMMATICALISATION	62
162	DEMONSTRATIVES	58
163	PASSIVES	57

Fonte: autor.