

VoTec: vocabulário bilíngue on-line de Linguística Histórica

VoTec: on-line bilingual Historical Linguistics vocabulary

Márcio Issamu YAMAMOTO*

RESUMO: O objetivo deste trabalho é apresentar um recorte da nossa pesquisa de Mestrado, conforme Yamamoto (2015), em que criamos um vocabulário terminológico bilíngue, português-ínglês, de Linguística Histórica (LH), sob a metodologia e abordagem da Linguística de *Corpus* (LC). Teoricamente, este trabalho está embasado nos conceitos de vocabulário de Barbosa (1992), de Terminologia de Krieger e Finatto (2004), de contextos definitórios de Aubert (1996) e na Teoria Comunicativa da Terminologia de Cabré (1999). Do ponto de vista metodológico, selecionamos termos e elaboramos as definições deles a partir da análise de *corpora* constituídos de textos da área de LH. Os *corpora* caracterizaram-se por ser comparáveis e acadêmicos, compostos por teses, dissertações, artigos científicos e anais de eventos. A dimensão dos *corpora* totalizou aproximadamente um milhão de itens, dos quais 500 mil são de língua portuguesa e 500 mil são de língua inglesa. O *software* que usamos para o processamento dos dados linguísticos foi o *WordSmith Tools* versão 6.0 (SCOTT, 2012), e as ferramentas que utilizamos desse programa foram a lista de palavras, a lista de palavras-chave e o

ABSTRACT: This article deals with Historical Linguistics terminology conceived under the methodological perspective of Corpus Linguistics (Berber Sardinha, 2004). The theories that based this research are that of Barbosa (1992), on the concept of vocabulary; on Terminology by Krieger and Finatto (2004); on the Communicative Theory of Terminology by Cabré (1999); and on definition contexts by Aubert (1996). The purpose of this research was to build a terminological bilingual vocabulary, Portuguese-English, in which the definitions were extracted from a terminological bank in Historical Linguistics. The research corpus was comparable, scientific, composed of theses, dissertations and articles on Historical Linguistics - in Portuguese and English -, conference proceedings were also included. The corpora size was approximately one million items, of which 500 thousand items are in Portuguese and 500 thousand items are in English. The program used for processing linguistic data was WordSmith Tools 6.0 (Scott, 2012), and the used tools were: the wordlist, the keyword list, and concord. The target audience consists of translators, students and professionals in

* Doutorando em Linguística pela UFU. Professor assistente do curso de Letras Inglês da Universidade Federal de Jataí (UFJ). ORCID: <https://orcid.org/0000-0001-9792-8187>. marcioy@ufg.br.

concordanciador. Construimos as definições seguindo o padrão gênero próximo, diferença específica (GPDE), com vistas a tradutores, estudantes e profissionais da área de Letras, Linguística e Tradução. Como resultado, o nosso vocabulário bilíngue de Linguística Histórica possui 26 termos, está disponibilizado na plataforma *on-line* denominada Vocabulário Técnico (VoTec), de Fromm (2007), sendo de acesso público e gratuito.

PALAVRAS-CHAVE: Linguística Histórica. Terminologia. Linguística de *Corpus*. Vocabulário bilíngue português-inglês.

the area of Languages, Linguistics and Translation. Definitions were constructed according to proximate genus and specific difference patterns, and the result containing 26 terms was made available at the online platform of Technical Vocabulary – VoTec - being of public and free access.

KEYWORDS: Historical Linguistics. Terminology. Corpus Linguistics. Portuguese-English bilingual vocabulary.

1 Introdução

A Linguística Histórica (LH) é uma subárea da Linguística que se ocupa das mudanças das línguas, por meio de estudos comparados e interpretações de dados intra e extralinguísticos. Ela se beneficia de dados provenientes da Etimologia, da Filologia, da Linguística Diacrônica e da Filologia Românica, que se constituíram a partir das contribuições de Schlegel (1772-1829), Bopp (1791-1867) e Grimm (1785-1863), na medida em que eles realizaram análises diacrônicas do sânscrito, do latim, do grego, do persa moderno, do eslavo e do germânico pelo método comparativo na Europa (FARACO, 2005).

Desenvolvemos esta pesquisa com o objetivo de elaborar um vocabulário bilíngue de LH, português-inglês, disponibilizado *on-line* na plataforma denominada Vocabulário Técnico (VoTec)¹, criada por Fromm (2007), considerando o valor da LH (YAMAMOTO, 2015). Adotamos o conceito de vocabulário de Barbosa (1992) no qual

¹ A plataforma VoTec está locada no servidor do Instituto de Letras e Linguística (ILEEL) da UFU sob a coordenação do prof. Dr. Guilherme Fromm. Disponível em: <http://treino.votec.ileel.ufu.br/>. Acesso em: 6 abr. 2020.

o léxico de uma língua é subdividido em vocabulários, que possuem como unidade padrão o vocábulo, manifestado em atos de fala. Nessa perspectiva, os vocabulários representam a terminologia de uma área de conhecimento e são vistos como conjuntos. Neles, estão contidos os conceitos extraídos dos contextos das palavras-ocorrências, concebidos como um recorte do léxico da língua geral. Diante disso, chegamos à conclusão de que o produto de nossa pesquisa seria um vocabulário.

Dessa forma, este artigo é um recorte do trabalho que realizamos no Mestrado, no qual tivemos o intuito de possibilitar o acesso a uma rede conceitual da LH em contextos de língua portuguesa e inglesa.

Construímos o nosso vocabulário bilíngue da LH a partir de *corpora* acadêmicos bilíngues comparáveis², coletados e equilibrados em um número de aproximadamente 500 mil itens para cada língua. Esses dados linguísticos foram analisados quantitativamente a partir da abordagem e metodologia da Linguística de *Corpus* (LC). Em seguida, procedemos à análise qualitativa para extrairmos os traços distintivos que seriam utilizados na construção das definições e organizados em fichas terminológicas virtuais na plataforma VoTec.

No que diz respeito ao resultado da elaboração do vocabulário bilíngue, a macroestrutura dele contém a definição de 26 termos da LH, em português e em inglês, e a microestrutura apresenta as seguintes informações em relação aos termos: definição terminológica, classificação morfológica, ordem de frequência (posição no *corpus*), número de ocorrências no *corpus*, exemplos extraídos do *corpus* e definição enciclopédica³ quando possível.

² Segundo McEnery e Hardie (2012, p. 240), **corpus comparável** é aquele no qual há amostras de línguas diferentes de forma que possam ser comparáveis. Nesta pesquisa, trabalhamos com *corpora* na área da Linguística, subárea Linguística Histórica, gênero acadêmico em português e inglês, considerados multilíngues.

³ A definição enciclopédica foi disponibilizada em forma de NOTA; isso ocorreu quando o *corpus* analisado provia traços distintivos suficientes para a construção da definição terminológica e da enciclopédica.

Nas seções seguintes, detalhamos os princípios teóricos e metodológicos que fundamentaram a elaboração do vocabulário de LH no VoTec e, por fim, os resultados obtidos.

2 Pressupostos teóricos

Nesta seção, apresentamos brevemente alguns conceitos que subsidiaram o nosso estudo, quais sejam a Terminologia, a Teoria Comunicativa da Terminologia, os contextos definitórios, o paradigma definicional e a Linguística de *Corpus*. É importante ressaltar que, nesta pesquisa, a Linguística de *Corpus* foi adotada como abordagem e como metodologia.

2.1 Terminologia e a Teoria Comunicativa da Terminologia (TCT)

A nossa pesquisa está inserida na área da Terminologia e da Terminografia, já que o nosso objetivo no Mestrado foi o de identificar, conceituar e definir as unidades terminológicas da área da LH. A Terminologia tem como objeto de estudo o termo ou a unidade terminológica (UT) de uma área do conhecimento, também concebida como a terminologia de uma área de especialidade.

De acordo com Krieger e Finatto (2004), a palavra terminologia pode ser escrita com letra maiúscula e minúscula, havendo uma acepção para cada forma. A Terminologia, grafada com letra maiúscula, representa a subárea da Linguística que tem como objeto de estudo o termo e que se dedica à pesquisa, ao desenvolvimento e ao tratamento de termos como saber técnico científico. Nesse contexto, o termo é um signo bipartido cujas faces constitutivas são o conceito e a denominação. O conceito é o conjunto de traços distintivos ou semânticos que identifica e distingue um termo de outros termos em um contexto de uso específico da língua, e a denominação é o signo linguístico ou uma unidade lexical que materializa o conceito em um contexto de uso técnico científico da língua.

Tendo isso em vista, termo diferencia-se de signo linguístico, conforme concebido por Saussure, pois um termo apresenta um único conceito, não admitindo polissemia quando aplicado a uma área de especialidade. Em oposição ao caráter monossêmico presente em obras terminológicas, o significado de um signo linguístico pode ser denotativo (literal) e conotativo (figurado), dependendo do contexto de uso da língua geral na perspectiva lexicológica. Esse fenômeno permite que um signo linguístico ou uma unidade lexical tenha várias acepções em uma obra lexicográfica, ou seja, que seja polissêmica (FINATTO, 2001).

A terminologia grafada com letra minúscula, segundo Krieger e Finatto (2004), representa o conjunto de termos pertencente a uma rede conceitual de uma área de especialidade, para comunicação do saber técnico-científico. Nesse caso, o termo está presente em diferentes áreas do saber como, por exemplo, na terminologia da LH, da Biomedicina e da Agronomia.

Para embasar teoricamente a elaboração do nosso vocabulário bilíngue, adotamos a Teoria Comunicativa de Terminologia (TCT) de Cabré (1999), que concebe a existência do termo a partir do contexto em que este se insere, sendo também constituinte da linguagem natural e da gramática das línguas. Isto é, o uso de uma unidade lexical em um contexto de especialidade é o que determina seu *status* como termo, segundo critérios semânticos, pragmáticos e discursivos. A este fenômeno Cabré (1999) denomina de “caráter poliédrico do termo”, no qual uma unidade lexical deixa de pertencer ao universo da língua geral e passa a ser uma UT do conhecimento especializado ou vice-versa.

Em detalhes, a referida autora propõe que a unidade lexical pode ser uma unidade linguística, cognitiva ou sociocomunicativa. Unidade linguística, uma vez que parte de um contexto real de uso da língua; unidade cognitiva, pois serve ao propósito de diferenciar o conhecimento especializado do geral; e sociocomunicativa,

em virtude de os aspectos pragmáticos definirem o uso das unidades lexicais em situações comunicativas determinadas.

Para exemplificar os princípios acima expostos, observemos o termo língua. Este pode ocorrer em vários contextos: na linguagem cotidiana, na Anatomia, na Zoologia e na Linguística. Contudo, é seu uso que determinará o grau de especificidade dele, ou seja, se pertence a uma área técnico-científica ou não.

2.2 Contextos definitórios e paradigma definicional

Os padrões para a elaboração das definições de um termo advêm das diretrizes da Terminologia. Quanto aos traços semânticos, sua extração é feita dos contextos nos quais os termos ocorrem dentro de textos da área da LH.

De acordo com Aubert (1996), os contextos definitórios, provenientes dos contextos de uso, apresentam os traços semânticos utilizados na representação conceitual dos termos. Segundo Pavel e Nolet (2002, p. 48), “os contextos definitórios apresentam características essenciais do conceito em estudo, enquanto que os explicativos fornecem informação sobre algumas das características”. São essas características essenciais dos conceitos que extraímos do *corpus*, organizamos nas fichas terminológicas do Votec e, posteriormente, utilizamos na elaboração da definição dos termos.

Vale ressaltar que esses traços semânticos precisam ser organizados sob um padrão que esteja em consonância com os princípios da Terminologia. Finatto (1998) propõe três diferentes tipos de paradigmas definicionais: o enciclopédico, o lexicográfico e o terminográfico. O paradigma definicional enciclopédico visa à produção de uma definição abrangente e “diversificada no teor de informações”, que contenha “traços relevantes e irrelevantes” ao caracterizar uma palavra na descrição de “coisas” (FINATTO, 1998, p. 1). O paradigma definicional lexicográfico caracteriza-se “pela predominância de informações linguísticas tratando mais de palavras”

(*Ibidem*, p. 2). Finalmente, o paradigma definicional terminográfico “traz predominantemente conhecimentos formais sobre ‘coisas’ ou fenômenos” (*Ibidem*, p. 2).

Em nosso estudo, adotamos o paradigma definicional de cunho terminológico, o padrão de definição gênero próximo e diferença específica (GPDE) e, excepcionalmente, o padrão do LIDE⁴, do jornalismo (CALLADO, 2012). Vale ressaltar que essas escolhas se justificam pelo fato de termos definido termos mais abstratos e também pelo fato de o *corpus* ter nos proporcionado informações suficientes para procedermos dessa maneira.

2.3 A Linguística de Corpus

Escolhemos a LC como metodologia e abordagem para o desenvolvimento da nossa pesquisa por razões como: (i) possibilitar a descrição da língua em uso; (ii) permitir a análise qualitativa e quantitativa dos dados linguísticos; (iii) fornecer resultados probabilísticos e estatísticos confiáveis utilizados por especialistas; (iv) permitir que os dados analisados sejam todos provenientes dos *corpora* (as definições foram elaboradas a partir da análise qualitativa das linhas de concordância, nas quais havia os traços semânticos ou semas); (v) propiciar a observação da regularidade dos termos disponibilizados nas linhas de concordância (elas mostram a repetição dos itens que coocorrem); e (vi) possibilitar, a partir da lista de palavras-chave, a identificação dos contextos definitórios em que os termos ocorrem.

Na LC, *corpus* é todo material de estudo ou pesquisa que serve à extração de dados linguísticos, qualquer que seja o tamanho ou gênero; é um conjunto de textos, legíveis por computador, que representa uma língua e serve à pesquisa linguística,

⁴ O termo LIDE, originalmente LEAD (guiar, conduzir) em inglês, é um padrão adotado para a redação do primeiro parágrafo de uma notícia na área de Comunicação e Jornalismo. Quando seguido fielmente, o parágrafo responde as *WH questions* que são: o quê? Quem? Quando? Onde? e Por que? (how).

cumprindo requisitos como autenticidade, representatividade e probabilidade (BERBER SARDINHA, 2004). Essa visão é confirmada por Tagnin (2015, p. 20) ao definir *corpus* como “bancos de textos de linguagem autêntica, criteriosamente construídos, destinados à pesquisa e legíveis por computador”. Além desses aspectos, o fator equilíbrio e normalização também é levado em consideração, já que os *corpora* tendem a ser padronizados em termos de gêneros textuais e tamanho. Em nossa pesquisa, esses pré-requisitos foram atendidos, com *corpora* constituídos por textos de gênero científico (teses, dissertações, artigos científicos e anais de eventos), dimensionados e equilibrados em 500 mil itens cada.

3 Metodologia

Os passos metodológicos que tomamos para a condução da pesquisa são oriundos da Terminologia e da LC com o uso do console do *WordSmith Tools 6.0* (WST) (SCOTT, 2012). Adiante apresentamos os passos metodológicos que fizeram parte da pesquisa.

- a) Compilação dos *corpora* em português e inglês;
- b) Extração dos termos com o WST 6.0 (SCOTT, 2012).
- c) Escolha do VoTec como plataforma para administração dos dados.;
- d) Registro dos dados, elaboração das definições e disponibilização no Votec. (ILARI, 2002, p. 39);

3.1 Compilação dos *corpora* em português e em inglês

Como mencionamos no item 2.4 deste artigo, a nossa pesquisa é baseada em *corpus* a partir da perspectiva da LC. Logo, é fundamental expormos como coletamos os *corpora* e as características deles.

Coletamos os *corpora* de LH da internet, por meio de buscas que realizamos no *Google*. Cabe pontuar que a coleta se restringiu a arquivos de acesso público e gratuito,

para que os direitos autorais não fossem infringidos, visto que iríamos usar os textos para produzir definições que divulgaríamos na plataforma VoTec. Os arquivos possuem produções científicas, por exemplo, teses, dissertações e artigos da área de LH, e são provenientes de *sites* de caráter científico, como revistas da área e de anais de eventos (no caso do *corpus* de língua inglesa).

Originalmente, compilamos os arquivos em formato pdf e, em seguida, os salvamos em txt. Descartamos os arquivos bloqueados, pois entendemos que não estavam disponíveis para serem alterados em sua formatação. Para compilá-los, usamos atalhos da Computação com os seguintes dados em português e em inglês: Linguística*Linguística Histórica*teses/dissertações/artigos científicos:pdf.

O *corpus* de LH em português foi composto por 33 textos, totalizando 536.330 itens. Desses 33 textos, 28 eram artigos, dois (2) eram dissertações e três (3) eram teses. Em relação à busca, observamos que as revistas de acesso livre têm sido a fonte de propagação das pesquisas e artigos na área de LH. As dissertações e teses são facilmente encontradas nas bases de teses e dissertações das universidades e também na Biblioteca Digital Brasileira de Teses e Dissertações⁵.

O *corpus* de língua inglesa da LH é constituído de oito textos, dimensionado em 521.794 palavras. O balanceamento do *corpus* foi possível em número de palavras, o que não ocorreu quanto ao número de textos. Esta divergência justificou-se pelo gênero textual, pois os anais de eventos em inglês são mais extensos que os encontrados no Brasil. Os anais corresponderam a sete (7) textos e o subgênero acadêmico tese correspondeu a um (1). É importante notar que o subgênero anais não foi encontrado em português e, em inglês, há uma média de dez (10) artigos por anais.

⁵ Biblioteca virtual. Disponível em: <http://bdtd.ibict.br/vufind/>. Acesso em: 20 nov. 2019.

3.2 Extração dos termos com o WST 6.0

O WST é um console arquitetado para análises linguísticas de *corpora* escritos, geralmente de grandes dimensões. As ferramentas do WST que utilizamos em nossa pesquisa foram o gerador da lista de palavras, o extrator de palavras-chave e o concordanciador (VIANA; TAGNIN, 2010, p. 43).

Primeiramente, geramos a lista de palavras dos arquivos dos *corpora* de LH que estavam em formato txt, pois os arquivos só podem ser lidos pelo WST se estiverem em formato txt (bloco de notas), que é o tipo de texto sem formatação. Assim, o WST disponibilizou os dados quantitativos de frequência e de recorrência das palavras dentro do *corpus*. Esta lista não foi publicada, pois o foco deste artigo é a Terminologia. Conseqüentemente, apresentamos adiante os detalhes da lista de palavras-chave.

Em segundo lugar, usamos o gerador de palavras-chave a fim de identificar as palavras-chave da LH em ambas as línguas. Para que o WST extraia as palavras-chave, é necessário que o *corpus* de estudo seja comparado com um *corpus* de referência três vezes maior. Em nossa pesquisa, usamos os seguintes *corpora* de referência: o *Corpus* de Português, para contrastá-lo com os textos de nosso *corpus* de LH em língua portuguesa, e o *Corpus of Contemporary American English* (COCA), para compará-lo com o nosso *corpus* de LH em língua inglesa. Ambos os *corpora* de referência apresentavam 100 mil types⁶ cada.

Quando extraímos as palavras-chave de um *corpus*, lidamos com o princípio da chavicidade. Segundo Viana e Tagnin (2010, p. 64), “a chavicidade reporta o resultado de um procedimento estatístico pelo qual a ferramenta levanta o quão importante cada palavra-chave positiva⁷ é para o *corpus* de pesquisa em relação ao de referência”.

⁶ Em LC, *types* refere-se ao número de palavras contidas em um *corpus* que não se repetem (TAGNIN, 2015).

⁷ “As palavras positivas são aquelas que encontramos mais frequentemente no *corpus* de estudo, enquanto as negativas são mais frequentes no *corpus* de referência” (YAMAMOTO, 2015, p. 66).

As palavras-chave úteis foram aquelas que recorriam no *corpus* de português e no de inglês. O fruto dessa seleção indicou o grupo de palavras que seriam candidatas a termos, previamente à análise qualitativa que definiria quais seriam mantidas ou não. Na sequência, na Figura 1, podemos visualizar algumas palavras-chave dos *corpora* de LH em português e em inglês; nela destacamos a posição em que os candidatos a termo *língua* e *language* aparecem.

Figura 1 – Lista de palavras-chave dos *corpora* de LH em português e em inglês (visão parcial).

N	Key word	Freq.	%	Tests	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	102,6	4.754,32	0,00	
5	FORMA	1.456	0,14	33	492,6	550,50	0,00	
6	SÉCULO	1.159	0,11	32	126,7	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	539,54	0,00	
11	FUTURO	888	0,09	16	84,66	1.936,49	0,00	
12	PALAVRA	855	0,08	23	184,8	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	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	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	524,97	0,00	
22	LIVRO	680	0,07	20	128,0	748,63	0,00	
23	SILVA	672	0,06	29	160,2	523,39	0,00	
24	LÍNGUAS	662	0,06	26	11,86	3.475,73	0,00	
25	REMETENTE	649	0,06	3	584	6.747,39	0,00	
26	SUFIXOS	621	0,06	6	221	7.118,19	0,00	

N	Key word	Freq.	%	Tests	RC. Freq.	RC. %	Keyness	P
1	LANGUAGE	2.518	0,26	15	60,1	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	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	3.396,5	0,00	
6	FORM	1.266	0,13	15	59,2	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	527	11.380,0	0,00	
11	GERMANIC	855	0,09	12	372	8.888,3	0,00	
12	ORIGIN	794	0,08	13	9,07	4.168,0	0,00	
13	ETYMOLOGY	783	0,08	5	267	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	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	1.142,1	0,00	
20	MODERN	697	0,07	14	43,6	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	1.273,9	0,00	
24	THUS	628	0,07	13	62,7	875,96	0,00	

Fonte: análise no WST (SCOTT, 2012).

Os candidatos a termo foram selecionados a partir do parâmetro de maior frequência no singular, por isso os termos no plural e no singular não foram lematizados. No *corpus* em português e em inglês, *língua* e *language* encontram-se próximas na posição de frequência. À primeira vista parecem similares, mas uma análise semântica nos leva a considerar que o termo *language* em inglês, quando traduzido para o português, pode significar *linguagem* ou *língua*. Na seção Resultados, explicamos o percurso seguido para lidar com esta nuance semântica e terminológica. Outro dado interessante é a similaridade da posição de *palavra*, em português, na 12ª

posição (corresponde à porcentagem de 0,08% de frequência no *corpus* de estudo), e de *word*, em inglês, na 3ª posição (corresponde à porcentagem de 0,19% de frequência no *corpus* de estudo) de frequência, o dobro do português. Independentemente da porcentagem de frequência dos candidatos a termo, observamos que, em geral, os termos foram recorrentes em ambos os *corpora*.

A classe gramatical escolhida para a elaboração do vocabulário de LH foi a dos substantivos, já que ela tem um caráter denominativo maior que as outras. Além desse fator, ela cumpre a função de sujeito, objeto direto e indireto e agente da voz passiva nas construções frasais. Ademais, os substantivos podem corresponder a uma porcentagem de 50% a 60% do léxico de uma língua (SILVA, 2006), o que faz dos substantivos uma classe fundamental. A lista de palavras-chave trouxe um total de 500 candidatos a termos, dentre os quais selecionamos os 26 termos que compuseram o vocabulário: língua, etimologia, gramática, palavra, verbo, linguística, linguística histórica, mudança, fala, linguagem, texto, variação, derivação, substantivo, adjetivo, prefixo, léxico, analogia, clítico, advérbio, particípio, sufixação, prefixação, afixo, mudança linguística, língua românica. Fizemos a escolha dos termos a partir da análise dos contextos linguísticos e de sua equivalência conceitual.

Finalmente, a análise dos contextos de ocorrência dos candidatos a termos foi possibilitada pelo concordanciador. O concordanciador é uma ferramenta que busca o termo, seu cotexto⁸ e seu contexto no *corpus* de ocorrência. Dessa forma, foi possível fazer a leitura e analisar os aspectos do conceito subjacente ao termo, a equivalência conceitual nas línguas, a análise dos contextos definitórios e a extração dos exemplos para serem usados no VoTec.

⁸ 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; TAGNIN, 2010, p. 71).

3.3 Escolha do VoTec como plataforma para administração dos dados.

O VoTec é uma plataforma bilíngue para administração de dados terminológicos, criada pelo prof. Dr. Guilherme Fromm em 2007. Ela tem sido utilizada por alunos de graduação e de pós-graduação, servindo aos objetivos terminográficos de estudos de Iniciação Científica, de práticas de componentes curriculares e de pesquisas de Mestrado e Doutorado. Para os níveis de graduação, ela se encontra disponível em <http://ic.votec.ileel.ufu.br> e, para os estudos de pós-graduação, em <http://treino.votec.ileel.ufu.br>. Vale salientar que o acesso à plataforma para edição de dados é restrito aos pesquisadores sob a supervisão do administrador.

Para o desenvolvimento da nossa pesquisa de Mestrado, lançamos mão do VoTec para a inserção, administração e publicação dos resultados do vocabulário bilíngue da LH. Os dados que compõem o resultado da obra terminográfica da área de LH estão disponíveis *on-line*⁹, sendo de acesso público e gratuito.

Selecionamos o VoTec como forma de publicação do nosso trabalho, pois ele permite que as definições sejam construídas a partir de dados provenientes do *corpus* de estudo, por meio de colunas de traços distintivos. Esta forma de preenchimento com dados do *corpus* reduz a subjetividade do autor e torna a definição mais objetiva. Mais do que isso, a plataforma oferece as formas de visualização normal (resumida, ou seja, com dados agrupados – padrão de dicionários impressos) e descritiva (com os dados separados, facilitando a identificação deles) e consultas no modo total (disponibiliza todas as informações na posição vertical), no tradutor (seleciona as informações mais relevantes para tradutores, a saber: definição, área de especialidade, tradução, sinônimos) e no modular (modo personalizado, no qual o consulente define as seções a serem exibidas).

⁹ Disponível em: <http://treino.votec.ileel.ufu.br>. Acesso em: 21 out. 2019.

Na próxima seção, descreveremos o passo metodológico que nos auxiliou a definir quais aspectos usar na construção da definição: o preenchimento das fichas terminológicas no VoTec.

3.4 Registro dos dados, elaboração das definições e disponibilização no Votec.

Na sequência, apresentamos algumas figuras para que a compreensão desse processo fique mais concreta para o leitor. Na tela inicial do VoTec, há a opção de inserir o nome do usuário e sua senha, gerando o acesso personalizado. Após a liberação de acesso, o pesquisador é direcionado à página de registro dos termos, conforme mostra a Figura 2.

Figura 2 – Visualização da tela de boas-vindas do VoTec.

The screenshot shows the VoTec user interface. At the top, it says 'Vocabulário Técnico Online' and 'Tela Cheia | English'. Below that, a grey banner says 'Bem Vindo, Marcio Issamu Yamamoto'. Underneath, it displays 'Seu último acesso foi em: 23/03/2018 19:56' and four buttons: 'Trocar Senha', 'Novo Termo', 'Cadastro de Fontes', and 'Sair'. There are two sections for term management:

Fichas ainda não aprovadas

Termo	Língua	Coletado em	Cadastrada por	Ações
Nenhum termo para revisar				

Total de Termos: 0

Fichas aguardando revisão

Termo	Língua	Coletado em	Cadastrada por	Ações
Nenhum termo para revisar				

Total de Termos: 0

Fonte: VOTEC, 2015.

A tela de registro dos termos permite a inserção do termo em português e em inglês e sua ontologia. Na ontologia, classificamos a grande área e a subárea às quais o termo pertence e, depois, gravamos os dados, conforme mostra a Figura 3.

Figura 3 – Passo 1: classificação da área e subárea de um termo para cadastro dele no VoTec.

Fonte: VOTEC, 2015.

Adiantando a descrição dos passos, na Figura 4, há a página de registro dos traços distintivos dos termos para posterior utilização na redação da definição terminológica¹⁰. Nas fichas terminológicas existentes na página, inserimos os traços semânticos em linhas e colunas organizadas pelo número de exemplos.

Figura 4 – Visualização da aba Traços distintivos no VoTec.

	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...				

Fonte: VOTEC, 2015.

A Figura 4 mostra como os traços distintivos podem ser organizados. Cada linha corresponde a um exemplo extraído de um contexto no qual o termo ocorre no

¹⁰ Aos interessados, o passo a passo do processo de cadastro e definição de termos da LH no VoTec encontra-se em Yamamoto (2015). Disponível em: https://lexically.net/wordsmith/corpus_linguistics_links/theses_using_wordsmith.htm. Acesso em: 25 out. 2019.

corpus, enquanto as colunas servem para a organização dos traços distintivos identificados. Cada vez que o traço distintivo repete, ele é inserido na mesma coluna que o anterior, o que é bastante visível nas colunas D, E e F. A organização desses traços distintivos, seguindo o padrão GPDE – gênero próximo, diferenças específicas -, pode ser visualizada na Figura 7, na qual apresentamos a definição final conforme disponível on-line.

As Figuras 5 e 6, a seguir, apresentam as linhas de concordância dos candidatos a termo *language* e *língua*. Em seguida, apresentamos a aplicação da análise qualitativa dos dados na definição terminológica do VoTec.

Figura 5 – Linhas de concordância do termo *language* em inglês seguido do colocado *is* (visão parcial).

N	Concordance	Set	Tag	Word #	Sent	Para	Hear	Sect	Date	%
2	; he had also demonstrated that "a language is not just a collection of	3.660	12	49'	0	70'	0	70'	HISTORICAL 2014/jun/27 C	69%
3	on the necessity of the form of the language is already [63] mixed with the	23.42	42	15'	0	23'	0	23'	Winfred P. Le 2014/jun/27 C	23%
4	one another. The true preeminence of a language is simply to develop from a	24.71	42	17'	0	24'	0	24'	Winfred P. Le 2014/jun/27 C	24%
5	. II. To assume one original universal language is impossible; there are	34.67	42	35'	0	34'	0	34'	Winfred P. Le 2014/jun/27 C	34%
6	length contrast in the language is a necessary condition for	52.70	2	158'	0	44'	0	44'	Laurel J. Brin 2014/jun/27 C	41%
7	is linguistically inferior. The resulting language is composite, truly mixed in	2.451	80	19'	0	9%	0	9%	degraff2005f: 2014/jan/30 C	9%
8	is nothing to be surprised by: any E-language is 'an epiphenomenon at	7.193	22	71'	0	21'	0	21'	degraff2009k: 2014/jan/30 C	21%
9	, and especially the rich metaphoric language is at least as important as	42.01	2	370'	0	76'	0	76'	Jan Keessen 2014/jan/30 C	76%
10	absence of external possessors from a language is in any way the normal	127.1	33	66'	0	79'	0	79'	Laurel J. Brin 2014/jun/27 C	99%
11	by the extent to which work on the language is integrated in a meaningful	1.576	51	63'	0	6%	0	6%	degraff2005f: 2014/jan/30 C	6%
12	stressed two other points: that a language is not an object with a reality	798	41	42'	0	13'	0	13'	HOW LONG ' 2014/jan/30 C	13%
13	of phonemes in the lexicon of a language is what is meant by lexical	4.840	16	41'	0	33'	0	33'	New perspec 2014/jan/30 C	34%
14	became integrated into the Egyptian language is the etymological word-play	9.683	46	87'	0	14'	0	14'	David Potter, 2014/jan/30 C	14%
15	braiding yarn into rope, what in sailor language is called "a line," soon they	41.77	2	341'	0	76'	0	76'	Jan Keessen 2014/jan/30 C	76%
16	This formal organization of the Icelandic language is much simpler than the	12.07	33	35'	0	12'	0	12'	Winfred P. Le 2014/jun/27 C	12%
17	the lives of individuals and of societies, language is a factor of greater	1.764	56	59'	0	7%	0	7%	degraff2005f: 2014/jan/30 C	7%
18	and almost pragmatic view. Language is for him a human	1.579	55	8%	0	30'	0	30'	HISTORICAL 2014/jun/27 C	30%
19	yields the further result that the Indic language is the older, the others	7.520	20	48'	0	7%	0	7%	Winfred P. Le 2014/jun/27 C	7%
20	of any I-language, whether or not such I-language is called 'Creole'. This is the	4.530	14	94'	0	13'	0	13'	degraff2009k: 2014/jan/30 C	13%
21	regional; the meaning in the standard language is 'play the violin'. He	99.20	1	81%	0	60'	0	60'	Anatoly Liber 2014/jan/30 C	45%
22	certain amount of work with the spoken language is indispensable for any	99.59	1	375'	0	97'	0	97'	Winfred P. Le 2014/jun/27 C	97%
23	of the individual sounds from which a language is constructed be an end in	98.40	1	279'	0	96'	0	96'	Winfred P. Le 2014/jun/27 C	96%
24	shows in what community he grew up. Language is no infallible sign of race,	97.29	1	28%	0	95'	0	95'	Winfred P. Le 2014/jun/27 C	95%
25	forms show, the difference in this language is still clearly distinct. In Old	78.42	64	80'	0	77'	0	77'	Winfred P. Le 2014/jun/27 C	76%
26	and primitive output, namely a "Creole" language, is to be evaluated with the	2.708	90	74'	0	10'	0	10'	degraff2005f: 2014/jan/30 C	10%
27	by combining words. Correspondingly, language is a social institution which	2.211	79	13'	0	42'	0	42'	HISTORICAL 2014/jun/27 C	42%
28	lie behind the historical tradition of a language is like a sea voyage without	81.77	74	96'	0	80'	0	80'	Winfred P. Le 2014/jun/27 C	80%

Fonte: análise no WST (SCOTT, 2012).

A Figura 5 será analisada na seção de resultados, na qual mostraremos a adequação da tradução de *language* para *língua* ou *linguagem* em português, bem como as definições dos termos no VoTec. A Figura 6 retrata alguns contextos do termo *língua* no *corpus* de LH.

Figura 6 – Linhas de concordância do termo *língua* em português seguido do colocado *é* (visão parcial).

N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	Hea	Hea	Secl	Secl	File	Date	%
1	permanente. Mas, uma vez que a língua é um sistema dinâmico, em que			48.02	1.435'	0	32'			0	32'		L_LH_PT_A_2018/dez/01		33%
2	, visto que possibilita visualizar como a língua é usada por uma determinada			314	15 74'	0	5%			0	5%		L_LH_PT_A_2018/dez/01		5%
3	de Ferdinand de Saussure); 8) a língua é uma estrutura autorregulada e			5.276	15 75'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
4	, além de se analisar a estrutura da língua, é aconselhável que se explique			33.22	1.240'	0	90'			0	90'		L_LH_PT_D_2018/dez/01		90%
5	relacionadas. A ideia de que a língua é um "organismo vivo" com			16.67	51 16'	0	11'			0	11'		L_LH_PT_A_2018/dez/01		11%
6	as análises do pensamento; 5) a língua é um conjunto de formas que			5.215	15 47'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
7	como ciência, a questão da língua é afetada pela relação do			13.54	42 33'	0	21'			0	21'		L_LH_PT_T_2018/dez/01		22%
8	linguística dos séculos XVI e XVII); 4) a língua é um conjunto de formas que			5.201	15 40'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
9	típica dos gramáticos); 3) a língua é um objeto histórico, derivado			5.175	15 28'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
10	que o falante tem da realidade; 7) a língua é uma convenção social, que			5.249	15 62'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
11	língua em especial: Glotologia de uma língua é a História d'essa língua" (Leite			12.67	36 66'	0	8%			0	8%		L_LH_PT_A_2018/dez/01		9%
12	(tradição platônica e aristotélica); 2) a língua é um conjunto de formas			5.163	15 22'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
13	as suas respectivas "filiações" --: 1) a língua é um meio de expressão do			5.148	15 15'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
14	de uma mudança no sistema da língua, é aí que haverá o espaço para			25.16	76 59'	0	26'			0	26'		L_LH_PT_T_2018/dez/01		26%
15	também do pressuposto de que a língua é composta por seus falantes e			33.34	1.239'	0	90'			0	90'		L_LH_PT_D_2018/dez/01		90%
16	questão, WL&H argumentam que toda língua é um sistema em transição,			29.59	1.017'	0	80'			0	80'		L_LH_PT_D_2018/dez/01		80%
17	nativa de um grupo de falantes. Essa língua é mantida, contudo, é			9.071	35 15'	0	24'			0	24'		L_LH_PT_T_2018/dez/01		22%
18	iniciais da IELPB, o autor afirma que a língua é uma expressão da sociedade,			42.04	1.230'	0	44'			0	44'		L_LH_PT_T_2018/dez/01		44%
19	nas diferentes formas de como o termo língua é definido em diferentes			12.69	49 59'	0	76'			0	45'		L_LH_PT_D_2018/dez/01		46%
20	, uma concepção psicológica da língua é iniciada, preocupada em			6.174	24 48'	0	37'			0	22'		L_LH_PT_D_2018/dez/01		23%
21	, partindo do princípio de que a língua é expressão de cultura, e a que			285	19 66'	0	10'			0	10'		L_LH_PT_A_2018/dez/01		12%
22	um jogo (visão de Wittgenstein); 10) a língua é um objeto de operações			5.301	15 87'	0	5%			0	5%		L_LH_PT_T_2018/dez/01		6%
23	com o pensamento de Durkheim. A língua é vista como um fato social,			27.43	1.010'	0	74'			0	74'		L_LH_PT_D_2018/dez/01		74%
24	organizada na alma. Defendia que a língua é energia, ou seja, é uma			19.82	71 8%	0	54'			0	54'		L_LH_PT_D_2018/dez/01		54%
25	(p.75). Para Dubois et al. (2001), a língua é considerada um sistema no			1.881	57 20'	0	3%			0	3%		L_LH_PT_A_2018/dez/01		3%
26	, parte-se do pressuposto de que a língua é uma entidade social, formada			13.70	49 47'	0	37'			0	37'		L_LH_PT_D_2018/dez/01		37%
27	, uma vez que a sua visão de língua é fundamentada pela relação			50.94	1.464'	0	53'			0	53'		L_LH_PT_T_2018/dez/01		53%

Fonte: análise no WST (SCOTT, 2012).

A Figura 6 apresenta vários contextos definitórios do termo *língua* em português. Na seção 4, discutiremos alguns resultados provenientes da análise desses contextos definitórios e faremos a análise do termo *language* (Figura 5).

Na Figura 7, a título de exemplificação, apresentamos a definição do termo *língua*, em português e em inglês, já concluída na plataforma VoTec.

Figura 7 – Visualização do termo *língua* e *language* no vocabulário bilíngue da LH presente no Votec.

▼ Português

[Voltar ao resultado da busca](#)

Língua. *Linguística Histórica. s.f.s.* produto social de uma comunidade linguística, que sofre mudanças devido a fatores sócio-histórico-culturais, usado para comunicação. Ex.: Para Mattos e Silva, um estudo que trabalhe com a sincronia contemporânea não exige tanto quanto o trabalho que objetive mergulhar, por exemplo, no português arcaico. Por essa razão, os aspectos sociais são tão importantes, visto que possibilita visualizar como a língua é usada por uma determinada comunidade.. *Sinônimos:* organismo vivo. *Hipônimo de:* comunicação social; comunidade linguística; falantes. *Hiperônimo de:* mudança; evoluções; latim vulgar; Português. *Veja Também:* [Analogia](#), [Fala](#), [Língua românica](#), [Mudança linguística](#), [Texto](#), [Variação](#). **Cópus:** *Posição na Ordem de Frequência:* (40); *Nº de Ocorrências do termo:* (1960). **Informações Enciclopédicas:** Língua natural (língua humana, língua idiomática, ou somente língua ou idioma) é qualquer linguagem desenvolvida naturalmente pelo ser humano, de forma não premeditada, como resultado da facilidade inata para a linguagem possuída pelo intelecto humano. Em: *Língua* - [Wikipédia](#)

▼ English

[Go back to search results](#)

Language. *Historical Linguistics. n.m/f.s.* social system structured to express human thoughts made by elements such as grammar, lexicon, syntax and pragmatics. Ex.: They are generally characterized in terms of their ideas about regularity and phonetic conditioning of sound change, but the manifesto stressed two other points: that a language is not an object with a reality of its own independent of its speakers (contra the Hegelian, group-psychology notions of Schleicher and others), and that the psychological and physical aspects of language must always have been the same.. *Synonyms:* system; social institution. *Hyponym of:* thoughts; community; people; speakers; change.. *Hypernym of:* sound; speech sounds; words; morphology; syntax; pragmatics; words; grammar; lexicon.. *See Also:* [Analogy](#), [Clitic](#), [Derivation](#), [Grammar](#), [Language2](#), [Lexicon](#), [Participle](#), [Prefix](#), [Romance language](#), [Speech](#), [Suffixation](#), [Text](#), [Variation](#), [Word](#). **Corpus:** *Frequency order position:* (29); *Term number of occurrences:* (2686). **Encyclopedic Information:** Language is the ability to acquire and use complex systems of communication, particularly the human ability to do so, and a language is any specific example of such a system. em: *Language* - [Wikipedia](#)

Fonte: VOTEC, 2015.

A Figura 7 retrata a definição final do termo *língua* na subárea da LH. É possível observar os aspectos mais recorrentes no *corpus*, usados para definir o termo *língua*, como o aspecto social e as mudanças que a língua sofre devido aos fatores externos (sociais, históricos e culturais). É importante considerar que os aspectos mudança linguística ou da língua são relevantes quando se trata da área de LH. A definição em língua inglesa também ressalta o aspecto social da língua, porém com foco mais estrutural, explicando as áreas da gramática, do léxico, da sintaxe e da pragmática. Essas questões são observáveis nos exemplos incluídos na plataforma VoTec, no modo de visualização Descritiva.

Na Figura 8, trazemos o termo *linguagem/language* definido para retratar a conclusão deste percurso da escolha do termo até sua definição.

Figura 8 – Visualização do termo *linguagem* e *language* no vocabulário bilíngue da LH presente no Votec.

▼ Português

[Voltar ao resultado da busca](#)

Linguagem. *Linguística Histórica.* s.f.s. sistema de sinais, escrito ou falado, usado pela humanidade para comunicação de ideias, marcado pela variabilidade, considerado uma ciência moral e histórica

NOTA: forma de expressão linguística do pensamento, de nível consciente e abstrato. Ex.: Afinal, a linguagem é ou não uma atividade do espírito?. *Sinônimos:* ciência histórica e moral. *Hipônimo de:* humanidade; ideias; pensamentos; espírito; criação social; Glotologia;. *Hiperônimo de:* sinais; sincronia; diacronia; sistemas linguísticos; vocábulos; linguagem escrita; linguagem falada;. *Veja Também:* [Analogia](#), [Léxico](#). *Corpus: Posição na Ordem de Frequência:* (210); *Nº de Ocorrências do termo:* (439). *Informações Enciclopédicas:* Linguagem pode se referir tanto à capacidade especificamente humana para aquisição e utilização de sistemas complexos de comunicação, quanto à uma instância específica de um sistema de comunicação complexo Em: *Linguagem* - [Wikipédia](#)

▼ English

[Go back to search results](#)

Language2. *Historical Linguistics.* n.m/f.s. system used to produce meaning, reference, naming and used by different groups of people for communication.

NOTE: types of language comprise philosophical, sacral, baby, hunters', legal, children's, thieves' and woovers'. Ex.: The ease with which OHG boununga, bauhning(a) 'significatio' (from bouhnen 'significare, innuere'), and OE gebeacnung 'categoria,' both of them derivatives of *baukn-, passed into religious and philosophical language points in the same direction.. *Hypernym of:* words, language, term.. *See Also:* [Language](#), [Word](#). *Corpus: Frequency order position:* (29); *Term number of occurrences:* (2686). *Encyclopedic Information:* any means of expressing or communicating, as gestures, signs, or animal sounds: body language em: *Language* - [Wikipedia](#)

Fonte: VOTEC, 2015.

O *corpus* de LH disponibiliza, em português e em inglês, a definição do termo *linguagem/language* como sistema usado pelos humanos para a comunicação, cujas definições são mais semelhantes entre si do que a definição dos termos *língua/language*, se comparadas. Em todas as entradas do VoTec, as definições provenientes da Wikipédia são disponibilizadas, por meio de *links*, para que o leitor possa ter acesso mais rápido e prático. Entendemos que a primeira linha de definição da Wikipédia é de padrão terminológico, mas a sua microestrutura em si é de caráter enciclopédico, o que a difere de nossa proposta.

4. Análise parcial dos resultados

No Quadro 1, apresentamos a lista de termos da subárea da LH que foram definidos na plataforma VoTec. Em seguida, faremos uma explicação do passo

metodológico de análise das linhas de concordância para os termos *língua/language*, bem como os resultados de sua análise.

Quadro 1 – Termos do vocabulário bilíngue da LH.

Votec – Vocabulário de LH – Lista final	
Português	Inglês
1. <i>Língua</i>	<i>Language</i>
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. Prefixo	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

Fonte: YAMAMOTO, 2015, p. 74.

Ao analisar o Quadro 1, é possível observar que os termos encontrados em um *corpus* de LH remetem a classes gramaticais (verbo, adjetivo e advérbio), ao campo de estudos do léxico (palavra, derivação, prefixo, léxico, clítico, particípio, sufixação, prefixação e afixo) e aos fenômenos estudados na Etimologia e na LH (língua,

Etimologia, gramática, Linguística, Linguística Histórica, mudança, fala, linguagem, texto, variação etc.).

Dentre os termos definidos, chamamos a atenção para o termo *language* em inglês, que, dependendo do contexto em que ocorre, pode ser traduzido por *língua* ou *linguagem*. Nesse caso, a análise dos contextos e sua equivalência conceitual, presentes nas linhas de concordância, é que definiram com qual termo do português houve correspondência do termo *language*. Logo, foi necessário examinarmos os contextos para constatar se *language* correspondia a *língua* ou a *linguagem*.

Na Figura 5, linha 18, encontramos a seguinte frase: “*Language is for him a human institution, an instrument made by man to meet human needs, and at no time beyond human control*¹¹”. Nesse contexto, identificamos que o autor se refere à *língua*, e não à *linguagem*. Já no exemplo “*One of the chief problems of the science of language is the inquiry into, and description of the classes of languages or speech stems, that is, of the languages which are derived from one and the same original tongue...*¹²”, observamos que o autor trata das ciências da linguagem, e não do termo *língua*.

Na Figura 6 observamos que alguns contextos do termo *língua* são semelhantes aos que ressaltam o papel social da língua e o *status* dela de organismo vivo. Outros contextos são diferentes e a definem de uma perspectiva mais estática, como um objeto histórico.

Quanto aos aspectos epistemológicos da LH, consideramos o imbricar existente entre as áreas que têm uma abordagem diacrônica dos estudos linguísticos. Essas áreas são: Filologia (Clássica e Românica), Etimologia e Linguística Diacrônica e/ou LH.

¹¹ Língua é, para ele, uma instituição humana, um instrumento criado pelo homem para atender às necessidades humanas, e nunca além do controle humano (todas as traduções contidas neste artigo são do autor).

¹² Um dos principais problemas das ciências da linguagem é a investigação e descrição dos tipos de línguas ou origens da fala, ou seja, línguas que são derivadas de um mesmo idioma... (texto adaptado – tradução do autor).

Nessa perspectiva, nos questionamos: em que medida essas áreas estão inter-relacionadas e em que pontos elas divergem?

Numa tentativa de lançar um pouco de luz sobre essa reflexão, baseando-nos na pesquisa bibliográfica que realizamos, chegamos a algumas considerações. Primeiramente, há a Etimologia, que tem como objeto de estudo o étimo e que usa os métodos gramático-histórico e etimológico (*corpus*, datação) para estudar a origem das palavras (DURKIN, 2009; VIARO, 2011). Em segundo lugar, temos a Filologia, cujo objeto de estudo é o texto. Os métodos filológicos são a crítica textual e o histórico comparativo. Essa área tem o objetivo de estudar o texto, o contexto de produção e a língua (VIDOS, 1996; BASSETO, 2001). Finalmente, a LH (às vezes, concebida como Linguística Diacrônica) tem como objeto de estudo a língua no tempo, por meio do método histórico-comparativo e da analogia. O objetivo de estudo dessa área é o estudo das mudanças linguística de uma língua (FARACO, 2005; OLIVEIRA E SILVA, 2008).

Obviamente delinear o estatuto dessas disciplinas não é tão simples como parece, nem é nosso objetivo fazê-lo. Tivemos que tentar definir o espaço de cada uma delas para que o *corpus* coletado realmente fosse da LH, e não dessas outras disciplinas. Concluindo esse desafio em partes, entendemos que elas realmente estão separadas entre si por linhas tênues. Diante do exposto, aceitamos que elas estão interligadas entre si e que não há como separá-las, já que uma contribui para outra, como em um movimento de simbiose.

Quanto ao estado da arte da produção acadêmica brasileira na área de LH, se contraposta ao da produção em língua inglesa, tendo como parâmetro os *corpora* da LH que coletamos, notamos que a necessidade de definir os termos técnico-científicos da LH foi maior no *corpus* de língua inglesa. Isso se tornou evidente com a menor quantidade de contextos definitórios para termos básicos da língua, como *verb*/verbo. Apesar de o termo ser recorrente nas duas listas de palavras-chave (português e

inglês), os contextos definitórios ou explicativos em inglês foram mais reduzidos, o que dificultou a construção de definições.

5. Considerações finais

Neste artigo, descrevemos brevemente a fundamentação teórico-metodológica que norteou a nossa pesquisa de Mestrado, os procedimentos de coleta dos *corpora* que utilizamos, os passos tomados para a construção do nosso vocabulário terminológico bilíngue (português-inglês) de LH, a análise quantitativa e qualitativa dos dados providenciados pelo WST e, ainda, a ordenação dos dados terminológicos na plataforma VoTec. Como resultado final, disponibilizamos 26 verbetes *on-line* para acesso gratuito, objetivando atender a tradutores, linguistas históricos e estudiosos de Linguística em geral.

Atualmente, como continuidade desta pesquisa, o *corpus* de LH foi reciclado e atualizado. As subáreas da Filologia, Etimologia e Linguística Diacrônica foram adicionadas a uma pesquisa maior que busca produzir um vocabulário bilíngue de Linguística e suas subáreas, em português e em inglês (YAMAMOTO, 2018). A plataforma do VoTec sofreu alterações e novas funcionalidades já estão disponíveis, como a inserção de áudio para pronúncia dos termos (CARDOSO, 2017).

Por fim, em relação ao vocabulário bilíngue de LH, podemos dizer que a elaboração dele aflorou realidades dos estudos diacrônicos e abriu caminhos para novas aplicações da LC em pesquisas de LH. Além disso, pudemos tratar de aspectos epistemológicos da LH, refletir sobre o estado da arte da produção acadêmica brasileira e também comentar a prática terminológica/terminográfica da academia brasileira.

Referências Bibliográficas

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*: II SIMPÓSIO LATINO-AMERICANO DE TERMINOLOGIA E I ENCONTRO BRASILEIRO DE TERMINOLOGIA TÉCNICO-CIENTÍFICA. 2., 1990, Brasília, **Anais** [...] Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT); Paris: União Latina, 1992. p. 152-158. Disponível em: <http://livroaberto.ibict.br/handle/1/992>. Acesso em: 20 nov. 2019.

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

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

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

CABRÉ, M. T. **La Terminología: representación y comunicación**. Barcelona: IULA/UPF, 1999. DOI <https://doi.org/10.1075/tlrp.1>

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, 2012. p. 46.

CARDOSO, S. A. F. **TermosTeo: a elaboração de vocabulários monolíngues de termos da Teologia em um estudo conduzido por corpus**. 2017. 340 f. Tese (Doutorado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2017.

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. Elementos Lexicográficos e Enciclopédicos na Definição Terminológica: questões de Partida. **Organon**, Porto Alegre, v. 12, n. 26, 1998. DOI: <https://doi.org/10.22456/2238-8915.29563>. Acesso em: 20 nov. 2019.

FINATTO, M. J. B. **Definição terminológica: fundamentos teórico-metodológicos para sua descrição e explicação**. Orientadora: Dra. Maria da Graça Krieger. 2001. 395 f. Tese (Doutorado em Estudos da Linguagem) – Instituto de Letras, Universidade

Federal do Rio Grande do Sul, Porto Alegre, 2001. Disponível em: <https://www.lume.ufrgs.br/handle/10183/1516>. Acesso em: 20 nov. 2019.

FROMM, G. **VoTec**: a construção de vocabulários eletrônicos para aprendizes de tradução. Orientadora: Dra. Stella Esther Ortweiller Tagnin. 2007. 214 f. Tese (Doutorado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008-150855/pt-br.php>. Acesso em: 22 nov. 2019.

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. p. 129-152.

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

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

KYTÖ, M. Corpora and historical linguistics. **Revista Brasileira de Linguística Aplicada**, Belo Horizonte, v. 11, n. 2, 2011. DOI: <http://dx.doi.org/10.1590/S1984-63982011000200007>. Acesso em: 22 nov. 2019.

MCENERY, T.; HARDIE, A. **Corpus Linguistics**: Method, Theory and Practice. Cambridge: Cambridge University Press, 2012. DOI <https://doi.org/10.1017/CBO9780511981395>

NASCENTES, A. **Dicionário Etimológico da Língua Portuguesa**. RJ: Livraria Acadêmica, 1955.

OLIVEIRA E SILVA, R. V. B. de M. **Caminhos da linguística histórica**: “ouvir o inaudível”. São Paulo: Parábola Editorial, 2008.

PAVEL, S.; NOLET, D. **Manual de terminologia**. Tradução de Enilde Faulstich. Canadá: Departamento de Tradução, 2002.

SCOTT, M. **WORDSMITH TOOLS**. Versão 6. Oxford: Oxford University. 2012. Disponível em: <http://lexically.net/wordsmith/version6/>. Acesso em 20 nov. 2019.

SILVA, M. C. P. Lexicografia bilíngue: uma verificação dos substantivos mais frequentes em dicionários bilíngues francês-português e português-francês. *In: LONGO, B.; 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. *Corpora na tradução*. São Paulo: Hub Editorial, 2015.

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

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

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

YAMAMOTO, M. I. *Linguística histórica e linguística de corpus*: Caminhos que se cruzam para desvelar a história da linguagem: um vocabulário bilíngue português – inglês. Orientador: Dr. Guilherme Fromm. 2015. 116 f. Dissertação (Mestrado em Linguística) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2015. Disponível em: https://lexically.net/wordsmith/corpus_linguistics_links/theses_using_wordsmith.htm. Acesso em: 25 out. 2019.

YAMAMOTO, M. I. Vocabulário Bilíngue Português/Inglês de Linguística Geral. *Revista Philologus*, V. Ano 24, p. 272-297, 2018. Disponível em <http://www.filologia.org.br/rph/ANO24/70supl/023.pdf>. Acesso em 06 Ab. 2020.

Artigo recebido em: 22.11.2019

Artigo aprovado em: 26.05.2020