

**LINGUÍSTICA HISTÓRICA E LINGUÍSTICA DE CORPUS:
UMA PROPOSTA PARA DICIONÁRIO BILÍNGUE
PORTUGUÊS-INGLÊS**

Márcio Issamu Yamamoto (UFU)
issamu2009@gmail.com

1. Introdução

O objetivo deste trabalho é apresentar, explicitar e detalhar os passos que serão tomados para confecção de um dicionário bilíngue português-inglês, na área de linguística histórica e, metodologicamente embasado na linguística de *corpus*, denominado *VoTec* – Vocabulário Técnico – disponível na rede mundial de computadores⁹⁷ e de acesso gratuito. Esse dicionário tem o intuito de servir a profissionais da Linguística, tradutores e quaisquer outros profissionais interessados na área de tradução e obras bilíngues. Para a confecção deste dicionário serão levantados *corpora* em português e inglês de linguística histórica, etimologia e filologia. Posteriormente estes *corpora* serão tratados com a ferramenta *Wordsmith Tools* (SCOTT, 2009)⁹⁸ para construção de uma lista de palavras (*Wordlist*) e de palavras-chave (*Keywords*). A partir dessas listas serão selecionados os termos que serão dicionarizados nessas duas línguas, cujas definições serão provenientes dos contextos nos quais eles estiverem inseridos. Esses contextos serão identificados pelo concordanciador (*Concord*), de onde extrairemos os conceitos para a formação das definições. As definições serão inseridas na plataforma do *VoTec* e posteriormente disponibilizadas na Internet.

2. Linguística histórica e linguística de corpus

A concepção desta obra terminográfica será beneficiada pela linguística histórica e pela metodologia da linguística de *corpus*. Aquela responsável por estudar a história das línguas e descrevê-las em seus contextos histórico-sociais (COSERIU, 1979, p. 236; SILVA, 2008, p. 30), e esta proveniente da evolução tecnológica e usada para o tratamento de *corpora* como modo de investigação empírica da linguagem (VIANA,

⁹⁷ Dicionário disponível em <www.pos.voteconline.com.br>. Acesso em: 23-03-2013.

⁹⁸ Programa disponível em <http://lexically.net/wordsmith/version5>. Acesso em: 23-03-2013.

2010, p. 27, 34).

Como definição de *corpus*, apoiamo-nos em Viana (2010, p. 27) que o define como uma compilação de textos de ocorrência natural que representa uma certa língua ou seus aspectos mais específicos para que possibilite uma análise linguística pré-estabelecida. Os *corpora* que serão utilizados para esta obra terminográfica serão compilados na área da linguística teórica, subárea da filologia, etimologia e linguística histórica. A linguística histórica 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 metaplasmos. Além dessa vantagem, há aspectos das áreas da fonologia e ortografia que permeiam o ensino da língua portuguesa e que podem ser elucidados com o auxílio da linguística histórica. Esses *corpora* são de caráter acadêmico, tais quais artigos científicos, dissertações e teses. Essa variedade de textos contribui para o enriquecimento da “gama de dados e aproxima um pouco mais o linguista da ideal onisciência dos dados da *huge chart bloomfieldiana*.” (VIARO, 2011, p. 104).

3. *Árvore de domínio*

Para iniciar, tomaremos a árvore de domínio da linguística com o objetivo de delimitar o campo de pesquisa terminográfica na área da linguística histórica. Para construção da obra terminográfica teremos uma macro e uma microestrutura que servem à elaboração das fichas terminológicas, nas quais os termos serão selecionados e inseridos por ordem de maior frequência nos *corpora*.

Árvore do Campo da Linguística

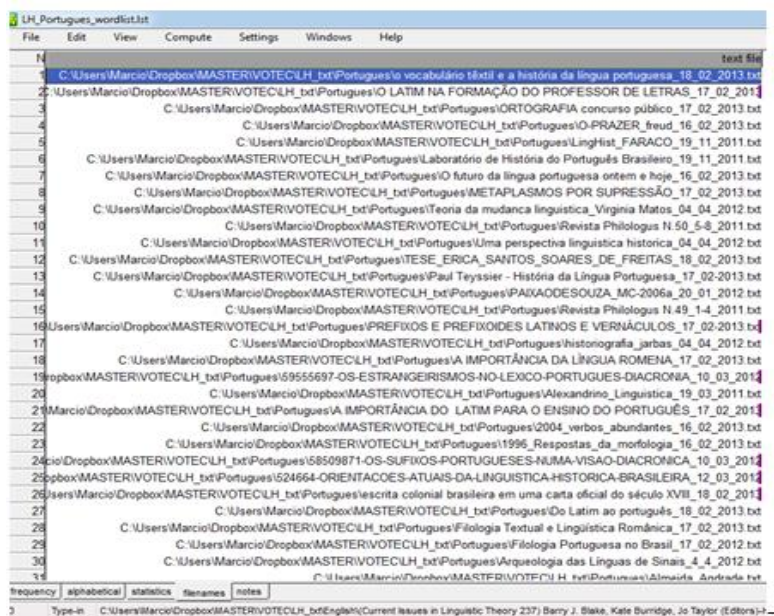


Figura 1. Árvore de domínio da linguística em construção segundo Fromm (2012).

4. Coleta e análise de corpora

Os *corpora* que estão sendo usados para extração de termos e contextos para criação do banco de dados são arquivos da área acadêmica, sendo eles artigos científicos, dissertações, teses ou livros nas subáreas de etimologia, filologia e linguística histórica. Os arquivos são de acesso público, baixados em formato pdf e salvos em formato txt a partir de *sites* específicos da internet. Dentre eles, citamos os *sites* da *Revista Philologus*, disponível em <<http://filologia.org.br/revista>>, e *Filologia e Linguística Portuguesa*, disponível em <<http://www.fflch.usp.br/dlev/lport/flp>> para o *corpus* de língua portuguesa e a base de dados Jstor, disponível em: <<http://www.jstor.org>> para o *corpus* de língua inglesa. Ressalvamos que este último disponibiliza

um maior número de artigos ao ser acessado de um servidor público federal devido a acordos interinstitucionais. Os arquivos são salvos em formato txt para possibilitar a leitura pelo programa *WordSmith Tool* - doravante *WST* -, ferramenta usada para análises lexicais. O tamanho dos *corpora* usados é de 536.330 palavras para o português, com 33 textos, e de 521.794 para o inglês, com 8 textos. De acordo com a classificação proposta por Viana (2011, p. 30), esses *corpora* classificam-se como: especializado, escrito, sincrônico, contemporâneo, estático, bilíngue não paralelo e de primeira língua. O *corpus* especializado, que objetiva a linguagem de especialidade contrasta com o de língua geral. O escrito contrasta com o oral; o sincrônico é aquele que traz o registro de uma dada língua num momento específico da história, enquanto o diacrônico cobre períodos diferentes dessa língua. É contemporâneo pois reflete a produção do tempo presente; estático significa que o *corpus* não permite ser alterado, ou seja, a inclusão ou exclusão de textos não pode ser realizada. Bilíngue, nesse caso, por abarcar a língua portuguesa e inglesa, não paralelo pois o *corpus* do português não é uma tradução do texto em inglês ou vice-versa. Um *corpus* de primeira língua significa que seus autores têm essa língua como língua nativa.



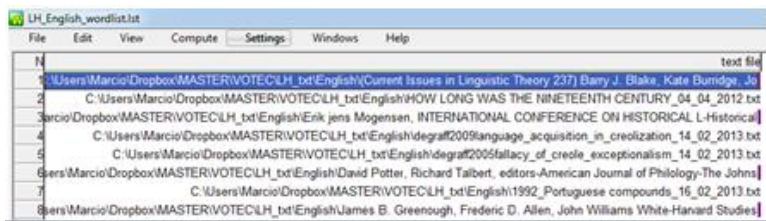


Figura 2. Visão parcial dos arquivos que fazem parte dos *corpora* de Linguística Histórica.

O primeiro passo tomado para o tratamento dos *corpora* é a utilização de três ferramentas do WST: o gerador de listas de palavras, o extrator de palavras-chave e o concordanciador (VIANA, 2011, p. 43). O gerador de lista de palavras levanta as formas diferentes de palavras no *corpus* com suas frequências a partir de uma seleção de textos estabelecida pelo pesquisador.

N	Word	Freq	%	Texts	%emma	Set
1	DE	22.242	4.15	33	100.00	
2	A	14.583	2.72	33	100.00	
3	E	11.448	2.13	33	100.00	
4	O	11.186	2.09	33	100.00	
5	QUE	10.692	1.99	33	100.00	
6	DO	8.056	1.50	33	100.00	
7	EM	7.017	1.31	33	100.00	
8	DA	6.486	1.21	33	100.00	
9	SE	6.151	1.15	33	100.00	
10	NO	4.038	0.75	33	100.00	
11	OU	3.852	0.72	33	100.00	
12	COM	3.494	0.65	33	100.00	
13	UM	3.446	0.64	33	100.00	
14	OS	3.426	0.64	33	100.00	
15	UMA	3.355	0.63	33	100.00	
16	E	3.337	0.62	33	100.00	
17	COMO	3.208	0.60	33	100.00	
18	POR	3.148	0.59	33	100.00	
19	PARA	3.101	0.58	33	100.00	
20	NA	2.853	0.53	33	100.00	

N	Word	Freq	%	Texts	%emma	Set
1	THE	29.241	5.60	8	100.00	
2	OF	17.846	3.42	8	100.00	
3	IN	13.829	2.65	8	100.00	
4	AND	11.046	2.12	8	100.00	
5	A	9.159	1.76	8	100.00	
6	TO	8.410	1.61	8	100.00	
7	IS	5.950	1.14	8	100.00	
8	THAT	5.001	0.96	8	100.00	
9	AS	4.456	0.85	8	100.00	
10	I	3.207	0.61	8	100.00	
11	FOR	3.042	0.58	8	100.00	
12	WITH	2.910	0.56	8	100.00	
13	BE	2.856	0.55	8	100.00	
14	IT	2.825	0.54	8	100.00	
15	THIS	2.809	0.54	8	100.00	
16	BY	2.551	0.49	8	100.00	
17	ON	2.324	0.45	8	100.00	
18	ARE	2.301	0.44	8	100.00	
19	FROM	2.220	0.43	8	100.00	
20	NOT	2.128	0.41	8	100.00	

Figura 3. Lista de palavras em ordem de frequência do *corpus* de Linguística Histórica (visão parcial).

Além do gerador de lista de palavras, usamos também o extrator de palavras-chave para levantar os termos que existem no *corpus* de estudo e compará-los com palavras de um *corpus* de referência. Nesse caso, usamos o *corpus* de referência BNC e ANC, de 122 milhões de palavras, para a língua inglesa, e o Banco do Português, 689 milhões de palavras, para o português. Após essa comparação, o WST exhibe as palavras-chaves positivas, aquelas que servem para identificar o *corpus* a ser descrito. Isto significa que as palavras que aparecem na lista de palavras-chave são aquelas que são mais recorrentes no *corpus* de estudo do que no *corpus* de referência, o que é chamado de chavicidade. “A chavicidade reporta o resultado de um procedimento estatístico pelo qual a ferra-

menta levanta o quão importante cada palavra-chave positiva é para o corpus de pesquisa em relação ao de referência” (VIANA, 2011, p. 64).

N	Key word	Freq	% Freq	FC	mes	Phonetic
1	LANGUAGE	3,178	0.253,910	6.488	00000	3141
2	VERB1	1,192	0.141,194	5.443	00000	7321
3	LINGUISTICS	489	0.09	895	3,568	00000
4	AMMATICALIZATION	291	0.06	5	3,128	00000
5	TRANSITIVE	242	0.05	85	2,268	00000
6	CREOLIZATION	194	0.04	2	2,098	00000
7	GENITIVE	182	0.03	30	1,818	00000
8	CONSTRUCTIONS	233	0.04	376	1,738	00000
9	VOWEL	235	0.05	465	1,678	00000
10	CHANGE	887	0.177,700	0.031	5,628	00000
11	LATIN	368	0.073,904	1,548	00000	
12	CLITIC	130	0.02	4	1,388	00000
13	SPEAKERS	301	0.062,674	1,358	00000	
14	FORMS	612	0.102,974	0.011	3,488	00000
15	PRONOUNS	175	0.03	278	1,308	00000
16	DATIVE	127	0.02	14	1,298	00000
17	CONSTRUCTION	404	0.087,494	1,288	00000	
18	ACCUSSATIVE	124	0.02	12	1,278	00000
19	NOMINATIVE	118	0.02	14	1,198	00000
20	REANALYSIS	114	0.02	27	1,108	00000
21	MEANING	399	0.088,804	1,078	00000	
22	SYNTAX	176	0.03	688	1,068	00000
23	DIALECTS	149	0.03	349	1,028	00000

Figura 4. Lista de palavras-chave do corpus de linguística histórica (visão parcial).

A terceira ferramenta utilizada é o concordanciador, ferramenta que permite a análise de uma palavra em seu cotexto⁹⁹. Os resultados dessa ferramenta são dispostos em linhas de concordância, fragmentos dos textos que são exibidos, a partir de uma palavra selecionada pelo usuário.

N	Concordance	Seleç	Vord	Cent	Int	Por	Para	Indec	Sect	File	%
1	é outra coisa. Para ela, a língua é uma expressão da	3,573	285	33%	0	3%	0	3%	as_18_02_2013	3%	
2	Neto (1986:18) declare. A língua é um produto social. é	3,283	104	78%	0	37%	0	37%	as_04_04_2013	37%	
3	. Nossa língua, como toda língua, é um rastro de velhos	1,165	51	18%	0	43%	0	43%	UA FORMACÃO	43%	
4	estática de língua – a língua é em si mesma um	3,152	99	34%	0	36%	0	36%	as_04_04_2013	35%	
5	em que a aquisição da nova língua é feita por indivíduos	1,566	54	65%	0	20%	0	20%	IA_10_03_2012	20%	
6	(2005: 24) que diz que “a língua é uma realidade	4,911	155	73%	0	77%	0	77%	colonia brasileira	76%	
7	preceito saussuriano de que a língua é forma não substância	19,763	985	35%	0	33%	0	33%	avista Philologus	33%	
8	da gênese e evolução da língua é um “vôo cego”, se tais	368	11	32%	0	9%	0	9%	Arqueologia das	9%	
9	O léxico de uma língua é constituído por	204	7	38%	0	7%	0	7%	ia_18_02_2013	6%	
10	perspectiva da crítica textual, a língua é apenas um dos	1,268	65	14%	0	34%	0	34%	ciologia Textual	32%	
11	língua e que o estudo da língua é objeto da Linguística,	247	11	62%	0	7%	0	7%	ciologia Textual	7%	
12	tanto, sabendo-se que cada língua é fruto da conjunção	51	1	10%	0	3%	0	3%	IMPORTÂNCIA	3%	
13	linguística. E, nesse sentido, a língua é mais do que uma	3,119	98	30%	0	35%	0	35%	as_04_04_2013	35%	
14	, acha que a sintaxe de uma língua é atin- gida primeiro que	43,034	2,395	32%	0	73%	0	73%	avista Philologus	73%	
15	de regras sintáticas de uma língua é também relativamente	4,013	277	95%	0	3%	0	3%	as_18_02_2013	3%	
16	a coesividade as repete. A língua é eminentemente	8,857	409	11%	0	73%	0	73%	as_18_02_2013	38%	
17	particularidades locais de sua língua é a consequência dos	8,947	414	60%	0	74%	0	74%	as_18_02_2013	39%	
18	dos fonemas – os sons da língua – é impressionante	39,873	1,695	66%	0	73%	0	73%	avista Philologus	73%	
19	com o historiógrafo da língua é um esquecimento de	8,352	283	27%	0	94%	0	94%	as_04_04_2013	93%	
20	períodos antigos de qualquer língua é algo difícil e que	401	24	23%	0	5%	0	5%	Jma perspectiva	6%	
21	possibilita visualizar como a língua é usada por uma	471	28	74%	0	8%	0	8%	Jma perspectiva	7%	
22	autoritária de uma língua é um instrumento que	12,965	826	60%	0	78%	0	78%	as_18_02_2013	55%	
23	partindo do princípio de que a língua é expressão de cultura,	311	20	65%	0	10%	0	10%	as_18_02_2013	10%	

Figura 5. Linha de concordância para o termo língua seguida do colocado é (vista parcial).

⁹⁹ 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, 2011, pg. 71).

5. VoTec

Feitos os levantamentos de *corpora*, as listas de palavras e de palavras-chave, os dados serão inseridos na plataforma do VoTec para construção da macro e microestrutura do dicionário. Para definirmos essa plataforma, “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, pg. 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. Diferentemente dessa, a descritiva apresenta os dados de forma hierárquica e detalhadamente. Nas opções de consulta, a de módulo total disponibiliza todos os campos do banco de dados, sendo de caráter lexicográfico; em seguida, a de módulo tradutor permite ao consulente acessar as informações mais frequentes usadas por tradutores, tais quais: área de especialidade, tradução, sinônimos, definição, entre outros. Finalmente, a consulta modular exhibe a microestrutura de acordo com a busca do consulente, acelerando sua busca pela seleção mais específica de dados. Além dessas opções de visualização, o *site* disponibiliza ao usuário acesso para consultas externas.

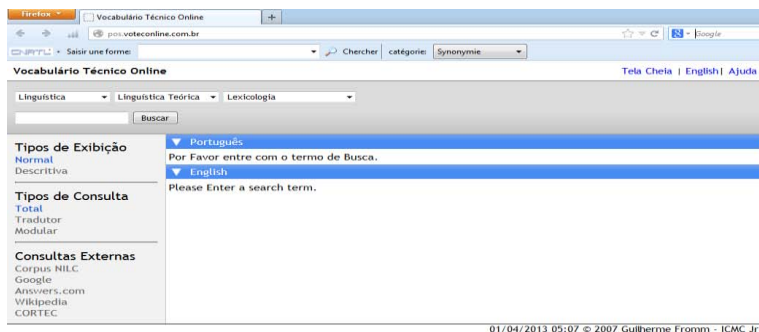


Figura 6. Visualização da página inicial do VoTec.

A abordagem do banco de dados adotada para o VoTec é a lexicográfica, 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.

The screenshot shows the 'Vocabulário Técnico Online' interface. The page title is 'Novo Termo'. At the top right, there are links for 'Tela Cheia' and 'English'. Below the title, there is a 'Voltar ao painel' button. The main section is labeled 'Passo 1' and contains three input fields: 'Termo' (a large text box), 'Língua' (a dropdown menu with 'Português' selected), and 'Ontologia' (a dropdown menu with 'Escolha uma Área' selected). Below the 'Ontologia' field is a 'Próximo Passo (Contextos)' button. At the bottom right of the page, there is a timestamp: '01/04/2013 06:10 © 2007 FFLCH - ICMC Jr.'.

Figura 7. Visualização do passo 1 para cadastro de termos no VoTec.

Nessa primeira página, o pesquisador deve inserir o termo a ser registrado no item Termo, após deve escolher a língua do referido termo (português ou inglês) e, finalmente, escolher sua ontologia desde a grande área, nesse caso linguística, até a subárea: linguística histórica. Selecionados esses dados, o próximo passo será o cadastro de contextos.

The screenshot shows the 'Vocabulário Técnico Online' interface for 'Cadastro de Contextos para língua'. At the top right, there are links for 'Tela Cheia' and 'English'. Below the title, there are two buttons: 'Voltar ao Painel' and 'Próximo Passo'. The main section is labeled 'Novo Contexto' and 'Passo 2'. It contains a 'Dados do Contexto' section with four input fields: 'Exemplo*', 'Conceito*', 'Fonte' (with 'OPDF' selected), and 'Data de Coleta*' (with a date format example '(dia/mês/ano ex.: 18/03/2007)'). There is a 'Salvar' button at the bottom and a 'Cadastrar Nova' link on the right.

Figura 8. Visualização da página Cadastro de contextos para língua (visão parcial).

Nesta etapa do procedimento, no espaço chamado Exemplo*, registramos o contexto no qual o termo se insere, dados esses provenientes do concordanciador do WST. A partir desse exemplo, faz-se um recorte de possíveis contextos definitórios ou explicativos¹⁰⁰ e o transpomos para

¹⁰⁰ Segundo Aubert (1996, p. 66-67), "os contextos explicativos apresentam alguns traços conceptuais pertinentes específicos do termo sob observação, frequentemente relativos à materialidade, finalidade, funcionamento e similares. [...] Talvez mais desejáveis, mas certamente menos encontrados, os contextos definitórios proporcionam um conjunto completo dos traços conceptuais distintivos do termo.

o campo Conceito*. 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. 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.

Contextos Cadastrados			
Exemplo	Conceito	Fonte	Ações
Para Mattos e Silva, um estudo que trabalhe com a sincronia contemporânea não exige tanto quanto o trabalho que objective 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.	aspectos sociais possibilitam visualizar como a língua é usada por uma determinada comunidade.	PDF 04/04/2012	editar - excluir
Embora encontremos em Ferdinand de Saussure esclarecimentos para aqueles que postulam uma concepção estática de língua - a língua é em si mesma um sistema imutável -, podemos observar que, por essa perspectiva, a língua que não muda não é a mesma exposta aos fatores externos, ou seja, aquela constituída pela historicidade e concretizada pela interação humana.	constituída pela historicidade e concretizada pela interação humana.	PDF 04/04/2012	editar - excluir
Nossa língua, como toda língua, 'é um rastro de velhos mistérios', no dizer de Guimarães Rosa, é um produto histórico, cujas marcas hão de ser perqueridas através dos tempos. O túnel do tempo, expressão consagrada por Tarallo, levar-nos-á aos mistérios tão indezessáveis, às raízes - qual a diferença entre radical e raiz? - às respostas a tantos porquês grotescamente dosados como regras e/ ou exceções gramaticais.	é um produto histórico	PDF 17/02/2013	editar - excluir
À língua é eminentemente mutável no tempo e o seu movimento de mudança tem o caráter de uma evolução,	mutável no tempo e o seu movimento de mudança tem o caráter de uma	PDF 18/02/2013	editar - excluir

Figura 9. Visualização da página Cadastro de contextos para o termo língua (visão parcial).

O acesso ao próximo procedimento é possível ao clicar-se sobre o ícone Próximo passo no canto superior direito da página.

Tela Chela | English

Vocabulário Técnico Online

Termo: língua

Passo Anterior Salvar Sair sem salvar

Contextos **Conceito Final / Definição**

Exemplo	Conceito	Fonte
1 A língua é um produto social, é uma atividade do espírito humano. Não é, assim, independente da vontade do homem, porque o homem não é uma folha seca ao sabor dos ventos veementes de uma fatalidade desconhecida e cega. Não está obrigado a prosseguir na sua trajetória, de acordo com leis determinadas, porque as línguas seguem o destino dos que a falam, são o que delas fazem as sociedades que as empregam.	um produto social, é uma atividade do espírito humano	PDF 04/04/0

Dados Traços Distintivos Semântica Termo Equivalente Termos Remissivos Informações Enciclopédicas

Dados

Ontologia: Linguística > Linguística Teórica > Linguística Histórica

Figura 10. Visualização da página de construção de microestrutura da página do VoTec.

Nesse momento do preenchimento da ficha terminográfica, o pesquisador preencherá as abas (i) Dados, (ii) Traços distintivos, (iii) Semântica, (iv) Termo equivalente, (v) Termos remissivos, (vi) Informações enciclopédicas e (vii) Conceito Final/Definição. 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 do termo. Nesse caso: Linguística > Linguística Teórica > Linguística Histórica. 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, muito usados para nomes de instituições ou termos da informática, tais quais CPU, CD-Rom, etc. Em seguida abaixo, há a Entrada por extenso – o significado de uma abreviação ou acrônimo, Variações morfossintáticas como *center* e *centre* no inglês, e Acepção n^o- campo usado para termos polissêmicos, 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^o de ocorrências do termo, dados provenientes da lista de palavras do WST.

Dados		Traços Distintivos	Semântica	Termo Equivalente	Termos Remissivos	Informações Enciclopédicas
Dados						
Ontologia:	Linguística > Linguística Teórica > Linguística Histórica					⋮
Categoria Gramatical:	Substantivo	Número:	Singular			
Gênero:	Masculino	Sigla/Acrônimo:	<input type="text"/>			
Entrada por Extenso:	Língua					<input type="text"/>
Var. Morfossintáticas:	<input type="text"/>					<input type="text"/>
Acepção N°:	<input type="text"/>					<input type="text"/>
Cópus						
Posição na Ordem de Frequência:	26	N° de Ocorrências do termo:	1964			

Figura 11. Visualização da aba Dados da página do VoTec.

(ii) **Traços distintivos:** nesse campo da ficha terminográfica são inseridos as informações extraídas dos contextos/exemplos que servirão à construção da definição final. Os dados serão organizados em colunas, nas quais os diferentes traços distintivos serão listados. Cada linha repre-

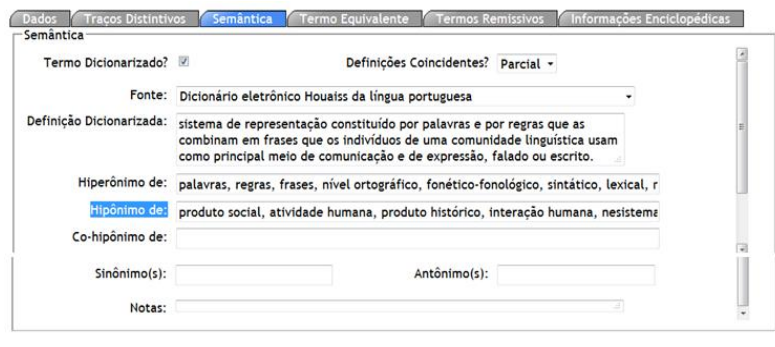
senta os dados provenientes de cada contexto/exemplo. Todas as vezes que os traços pertencerem ao mesmo campo semântico, eles serão listados nas mesmas colunas.



	A	B	C	D	E	F	G
1	produto so...	atividade ...					
2			produto hi...				
3	concretiza...		constituid...				
4		necessidad...		realidade ...	mudança in...	níveis ort...	
5			processo d...		mutável no...		
6	aspectos s...						
7	comunidade...				variações ...		organismo ...

Figura 12. 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 por questões de direito autoral e só serve como norte para o pesquisador, não sendo 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 for. Se necessário, o pesquisador pode deixar alguma nota que será verificada pelo administrador do sistema.



Semântica

Termo Dicionarizado? Definições Coincidentes? Parcial ▾

Fonte: Dicionário eletrônico Houaiss da língua portuguesa ▾

Definição Dicionarizada: sistema de representação constituído por palavras e por regras que as combinam em frases que os indivíduos de uma comunidade linguística usam como principal meio de comunicação e de expressão, falado ou escrito.

Hiperônimo de: palavras, regras, frases, nível ortográfico, fonético-fonológico, sintático, lexical, r

Hipônimo de: produto social, atividade humana, produto histórico, interação humana, nesistema

Co-hipônimo de: _____

Sinônimo(s): _____ Antonônimo(s): _____

Notas: _____

Figura 13. Visualização da aba Semântica da página do VoTec.

(iv) Termo equivalente: nessa aba, o sistema busca o termo equivalente na outra língua (inglês) para associá-lo ao termo em português. Esse processo só é possível se o termo foi registrado dentro das mesmas características ontológicas do termo que está sendo registrado.

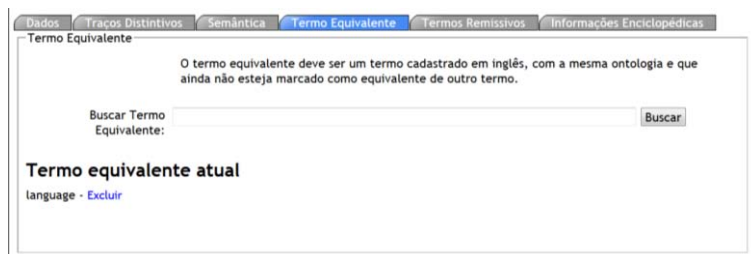


Figura 14. Visualização da aba Termo Equivalente da página do *VoTec*.

(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.

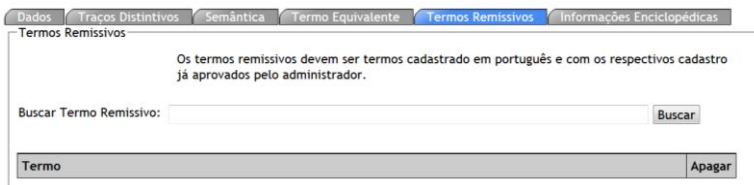


Figura 15. Visualização da aba Termos remissivos da página do *VoTec*.

(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.

Definição	Artigo	Fonte	Link	Apagar
Língua natural (língua humana ou somente língua) é qualquer linguagem desenvolvida naturalmente pelo ser humano, de forma não premeditada, como resultado da facilidade inata para a linguagem possuída pelo intelecto humano. Vários exemplos podem ser dados	língua	Wikipédia		

Figura 16. Visualização da aba Informações Enciclopédicas da página do VoTec.

(vii) Conceito Final/Definição: nesse campo, o pesquisador formula as definições para o termo selecionado em duas janelas denominadas Conceito Final e Definição. Na primeira, o pesquisador insere os dados para uma definição prévia do termo, 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 será visualizada pelos consulentes ao acessar a página do dicionário na Internet.

Conceito Final: produto socio-histórico mutável no tempo; atividade do espírito humano, que responde às necessidades socio-interativas organismo vivo caracterizado por um processo dinâmico, gradual e coerente de mutações nos níveis ortográfico, fonético-fonológico, sintático, lexical, morfológico, discursivo

Definição: atividade do espírito humano que responde às necessidades socio-interativas de uma comunidade linguística; produto socio-histórico, mutável no tempo caracterizado por um processo dinâmico, gradual e coerente de mutações nos níveis ortográfico, fonético-fonológico, sintático, lexical, morfológico e discursivo

Figura 17. Visualização da aba Conceito Final da página do VoTec.

Preenchidos todos os campos, o pesquisador deve salvar as informações antes de sair do sistema. Essas informações serão enviadas ao administrador do sistema que, após avaliação, as disponibilizará para o acesso público. Do contrário elas poderão ser reeditadas, a pedido do administrador, se necessário for.

Concluída a análise e aprovação do administrador, o termo será visualizado em português e inglês na página do Votec. Acessível ao público em geral e com visualizações alternativas de acordo com a necessidade do consulente.

Tipos de Exibição Normal Descritiva	Português Voltar ao resultado da busca língua. <i>Linguística Histórica.</i> <i>Língua.</i> s.f.s. atividade do espírito humano, que responde às necessidades socio-interativas de uma comunidade linguística; produto socio-histórico, mutável no tempo caracterizado por um processo dinâmico, gradual e coerente de mutações nos níveis ortográfico, fonético-fonológico, sintático, lexical, morfológico e discursivo. Ex.: A língua é um produto social, é uma atividade do espírito humano. Não é, assim, independente da vontade do homem, porque o homem não é uma folha seca ao sabor dos ventos vementes de uma fatalidade desconhecida e cega. Não está obrigado a prosseguir na sua trajetória, de acordo com leis determinadas, porque as línguas seguem o destino dos que a falam, são o que delas fazem as sociedades que as empregam. <i>Hipônimo de:</i> produto social, atividade humana, produto histórico, interação humana, necessidades sócio-interativas, organismo vivo., <i>Hiperônimo de:</i> palavras, regras, frases, nível ortográfico, fonético-fonológico, sintático, lexical, morfológico, discursivo, variações linguísticas., <i>Corpus: Posição na Ordem de Frequência:</i> (26); <i>Nº de Ocorrências do termo:</i> (1964). <i>Informações Enciclopédicas:</i> Língua natural (língua humana ou somente língua) é qualquer linguagem desenvolvida naturalmente pelo ser humano, de forma não premeditada, como resultado da facilidade inata para a linguagem possuída pelo intelecto humano. Vários exemplos podem ser dados Em: <i>língua</i> - Wikipédia
Tipos de Consulta Total Tradutor Modular	English Go back to search results language. <i>Historical Linguistics.</i> n.m/f.s. a system that bears human speech which goes under progression and modifications creating new linguistic forms; grouped in families understood by analogy of factors that work together; can be divided in two main classes: monosyllabic, that is isolating or radical, and polysyllabic or inflectional. Ex.: Labovian linguistics is thus of crucial importance for understanding some of the ways in which languages, at both the I- and E-language levels, are constantly and gradually evolving, and why 'fully convergent grammars' (cf. Hale 1998: 1) are an unlikely (i.e., 'abnormal') outcome of language acquisition (See 52.3 below)., <i>Hyponym of:</i> factors, families, <i>Hypernym of:</i> grammars, linguistics forms, human speech, monosyllabic, polysyllabic. <i>Corpus: Frequency order position:</i> (1); <i>Term number of occurrences:</i> (3919). <i>Encyclopedic Information:</i> Language is the human capacity for acquiring and using complex systems of communication, and a language is any specific example of such a system. The scientific study of language is called linguistics. Any estimate of the precise number of languages in th em: <i>Language</i> - Wikipedia
Consultas Externas Corpus NILC Google Answers.com Wikipedia CORTEC	

02/04/2013 20:33 © 2007 Guilherme Fromm - ICMC Jr.

Figura 18. *VoTec* - Visualização do termo língua na área de Linguística Histórica.

6. Considerações finais

Este artigo apresentou e detalhou os passos metodológicos adotados para a confecção de um dicionário bilíngue português-inglês na área de linguística histórica, direcionado para o público da área de linguística, tradução e outros profissionais que tenham interesse na área. Detalhamos como usar a metodologia da linguística de *corpus* para execução da coleta e tratamento dos *corpora* portanto, valemo-nos da ferramenta WST para leitura e análise dos dados linguísticos, confecção das listas de palavras, de palavras-chave e do concordanciador. Após tais procedimentos, explicitamos como se faz o lançamento de dados específicos na plataforma do *VoTec*, o preenchimento das fichas terminográficas *on-line* e a fase final para que a obra terminográfica seja disponibilizada na Internet. Esperamos que este trabalho seja útil para fortalecer o estudo da linguística histórica, filologia e etimologia nos países lusófonos e que sirva para a difusão e expansão da língua portuguesa, já que o acesso a obras em língua inglesa pode ser facilitado através desse trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

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

COSERIU, E. *Sincronia, diacronia e história*. Trad.: C. A. da Fonseca e M. Ferreira. Rio de Janeiro: Presença; São Paulo: Universidade de São Paulo, 1979.

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.

_____. Terminografia, sociolinguística, linguística de corpus, tradução, tecnologia da informação: convergências. In: MOLLICA, Maria Cecília; GONZALES, Marcos. (Orgs.). *Linguística e ciência da informação: diálogos possíveis*. Curitiba: Appris, 2012, p. 141-158.

SCOTT, M. *WordSmith Tools*. Versão 5. Disponível em: <<http://lexically.net/wordsmith/version5>>. Acesso em: 23-03-2013.

SILVA, Rosa Virgínia Mattos e. *Caminhos da linguística histórica – “ouvir o inaudível”*. São Paulo: Parábola, 2008.

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

VIARO, Mário Eduardo. *Etimologia*. São Paulo: Contexto, 2011.