

**VOTEC – FAZER TERMINOLÓGICO:
A MICROESTRUTURA EM DICIONÁRIOS VIRTUAIS**

Márcio Issamu Yamamoto (UFG/UFU)
issamu2009@gmail.com

RESUMO

Os dicionários são obras que disponibilizam definições para o vocabulário de língua geral ou de especialidade para os consulentes. Nosso questionamento feito foi: os dicionários especializados da linguística trazem uma microestrutura de dicionário ou de enciclopédia? Geralmente, os dicionários de linguística se inserem na segunda opção. São obras que disponibilizam conceitos e definições das entradas, bem como uma microestrutura com definições de abordagem histórica e explicativa (DUBOIS, 1999). A microestrutura, segundo Béjoint (2010), 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 de paradigma definicional. Em um dicionário ou vocabulário, procura-se manter um modelo recorrente dessa estrutura para facilitar o manuseio do usuário, para que haja padronização, e também como traço identitário da obra. Neste trabalho, objetivamos apresentar os passos e a metodologia tomados para a construção da microestrutura no dicionário bilíngue, português-inglês, o *VoTec – Vocabulário Técnico* – disponível *on-line*. Depois dos levantamentos de *corpora*, da construção das listas de palavras e de palavras-chave, os dados são 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, p. 8). O que diferencia a metodologia adotada da tradicional, no que tange às obras terminográficas, é que os dados são armazenados *on-line*. Logo, explicaremos os passos que nos permitem organizar os traços conceituais, para que, ao fim, tenhamos definições construídas, seguindo o padrão GPDE – gênero próximo, diferença específica.

Palavras-chave: *Votec*. Terminologia. Terminografia. Dicionário virtual. Lexicografia.

1. Introdução

Os dicionários são obras que disponibilizam definições para o vocabulário de língua geral ou de especialidade para os consulentes. Nosso questionamento para esta pesquisa foi: os dicionários especializados da linguística trazem uma microestrutura de dicionário ou de enciclopédia? Geralmente, os dicionários de linguística se inserem na segunda opção. São obras que disponibilizam conceitos e definições das entradas, bem como uma microestrutura com definições de abordagem histórica e explicativa (DUBOIS, 1999). A microestrutura, segundo Béjoint (2010, p.

13), é a estrutura disposta na horizontal de uma obra lexicográfica ou terminográfica, *grosso modo*. Em um dicionário ou vocabulário, procura-se manter um modelo recorrente dessa estrutura para facilitar o manuseio do usuário, para que haja padronização, e também como traço identitário da obra.

Neste artigo, objetivamos apresentar os passos e a metodologia tomados para a construção da microestrutura no dicionário bilíngue, português-inglês, o *VoTec– Vocabulário Técnico* – disponível on-line. Posterior aos levantamentos de *corpora*, da construção das listas de palavras e de palavras-chave, os dados são 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, p. 8). O que diferencia a metodologia adotada da tradicional, no que tange às obras terminográficas, é que os dados são armazenados on-line. Logo, explicaremos os passos que nos permitem organizar os traços conceituais, para que ao fim, tenhamos definições construídas, seguindo o padrão GPDE – gênero próximo, diferença específica.

2. Microestrutura e macroestrutura

A microestrutura, segundo Béjoint (2010), 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, etimologia, exemplos, entre outros. Este modelo, como proposto, também pode ser nomeado de paradigma definicional. Em um dicionário ou vocabulário, procura-se manter um modelo recorrente dessa estrutura para facilitar o manuseio do usuário, para que haja padronização, e também como traço identitário da obra.

Para que a definição seja redigida, é necessário que o terminólogo analise os contextos nos quais os traços sêmicos ou distintivos se inserem. No início dos estudos terminológicos, o cientista partia dos contextos para chegar ao termo, o referente. 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 as definições a partir dos traços semânticos nele contido. Este processo, que parte dos conceitos até o referente ou unidade terminológica, é chamado de percurso

onomasiológico; no caso de o pesquisador partir da unidade terminológica para a identificação dos traços distintivos que o constituem, temos o percurso semasiológico (BARROS, 2004).

As definições neste trabalho serão redigidas usando-se o padrão GPDE, gênero próximo e diferença específica. Este padrão implica em partir de conceitos mais abrangentes, a hiperonímia de acordo com as ciências do léxico, para conceitos mais específicos, o de hiponímia, no qual se faz a diferenciação conceitual de processos denominativos distintos. Este passo será desafiador, pois há conceitos que são presentes em mais de uma subárea da linguística histórica, doravante linguística histórica, como, por exemplo, o conceito de metaplasmos, presente na etimologia, filologia e linguística histórica. Outro exemplo é o conceito de língua que, além de ser distintos nessas subáreas, é distinto em toda a Linguística como uma ciência da linguagem.

2.1. 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 dicionário¹⁶⁸. 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 exhaustivamente 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. 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.

¹⁶⁸ Veja mais detalhes disponíveis em Yamamoto (p. 691-705, 2013).

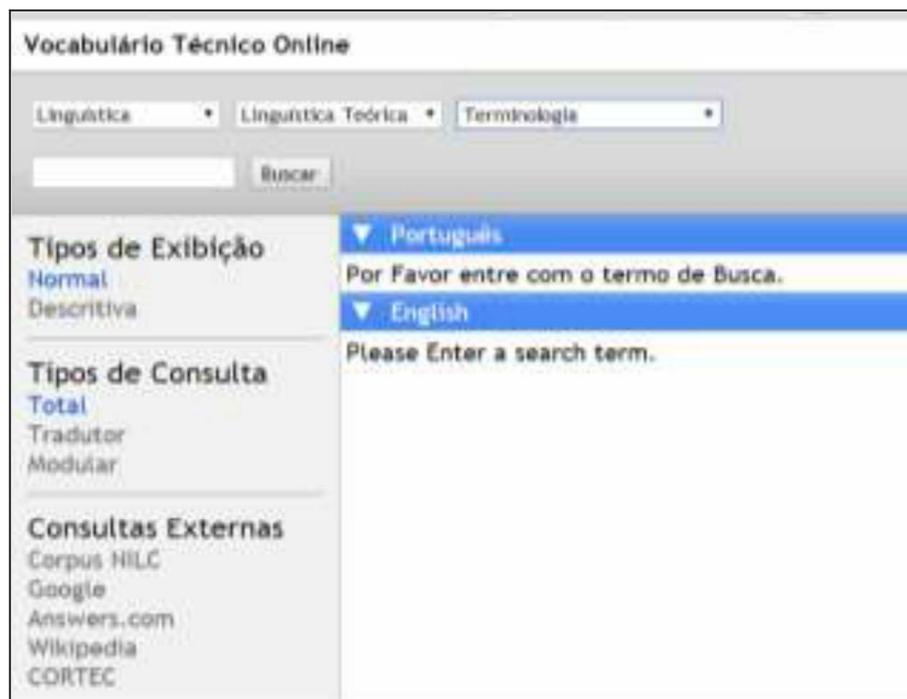


Fig. 1. Visualização da página inicial do VoTec

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.



Fig. 2. 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.



Fig. 3. Visualização da página Cadastro de contextos para *verbo* (visão parcial)

Nesta etapa do procedimento, no espaço chamado Exemplo*, registramos o contexto no qual a unidade terminológica se insere, dados esses provenientes do concordanciador do *WordSmith Tools – WST*. A partir desse exemplo, faz-se um recorte de possíveis contextos definitórios ou explicativos¹⁶⁹ e o transpomos para 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.

¹⁶⁹ 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 contraditórios, os contextos definitórios proporcionam um conjunto completo dos traços conceptuais distintivos do termo.

Contextos Cadastrados	
Exemplo:	Conceito:
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. untransitive) 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
In metonymic extensions to the context in which the P/I occurs, the meaning of the verb is extended to denote certain features of the broader event frame in which the P/I takes place, and without reference to which the correct interpretation of the verb cannot be achieved.	the meaning of the verb is extended to denote certain features of the broader event frame in which the P/I takes place, and without reference to which the correct interpretation of the verb cannot be achieved.
The original separate status of -afenge is further shown by the fact that -afenge combines in Lower Arrente and in ÊrC Arrente (here with a variant shape -afenge) with tense-inflected verbs to indicate that the subject of this subordinate verb is different from that of the main-clause verb.	the subject of this subordinate verb is different from that of the main-clause verb.

Contextos Cadastrados: 3

Fig. 4.

Visualização da página Cadastro de contextos para o termo *verbo* (visão parcial)

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

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. untransitive) 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 is extended to denote certain features of the broader event	the meaning of the verb is extended to denote certain features of the broader event	PDF 15/02/2013

Fig. 5. 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 da unidade terminológica. Nesse caso, linguística > linguística teórica > linguística histórica. Logo abaixo, há o campo categoria gramatical, onde selecionamos a opção substantivo para o termo *verbo*, no canto esquerdo. À direita há o campo Número onde selecionamos a opção singular, plural ou dual (como *lápiz*, 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°- 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*.

Dados	
Gênero:	Neutro *
Sigla/Acrônimo:	
Entrada por Extenso:	verb
Var. Morfossintáticas:	
Acepção N°:	39
Corpus	
Posição na Ordem de Frequência:	3
N° de Ocorrências do termo:	1916

Fig. 6. Visualização da aba *Dados* da página do VoTec

- (ii) **Traços distintivos:** nesse campo da ficha terminográfica são inseridas 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 representa 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
1	three-plac...				
2		denote cer...			
3			subject of...		

Fig. 6. 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: Answers.com

Definição Dicionarizada: The part of speech that expresses existence, action, or occurrence in most languages.
Any of the words belonging to this part of speech, as be, run, or conceive.
A phrase or other construction used as a verb.
[Middle English *verbe*, from Old French, from Latin *verbum*, word, verb]

Hiperônimo de:

Hipopônimo de: part of speech, words, phrase, construction

Fig. 7. Visualização da aba Semântica da página do VoTec

- (iv) **Termo equivalente:** nessa aba, o sistema busca 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.

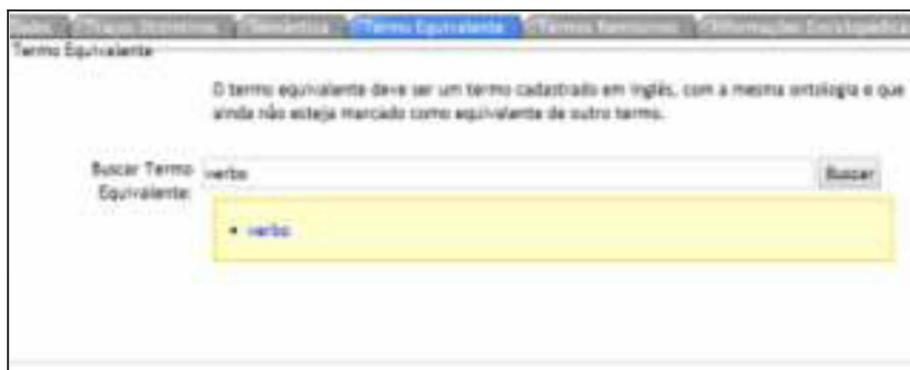


Fig. 8. Visualização da aba Termo Equivalente da página do VoTec

- (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.



Fig. 9. Visualização da aba termos remissivos da página do VoTec

- (vi) **Informações enciclopédicas:** adicionam-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.



Fig. 10. Visualização da aba Informações Enciclopédicas da página do VoTec

- (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 lapidado 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.



Fig. 11. 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.

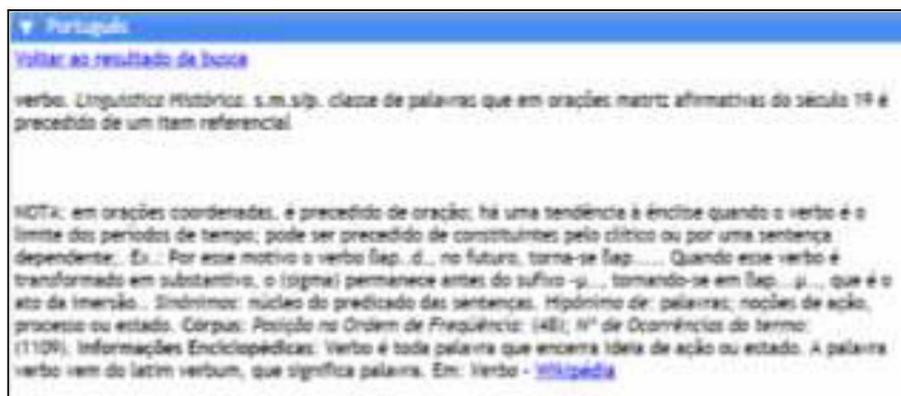


Fig. 12.

VoTec - Visualização do termo *verbo* em português na área de linguística histórica

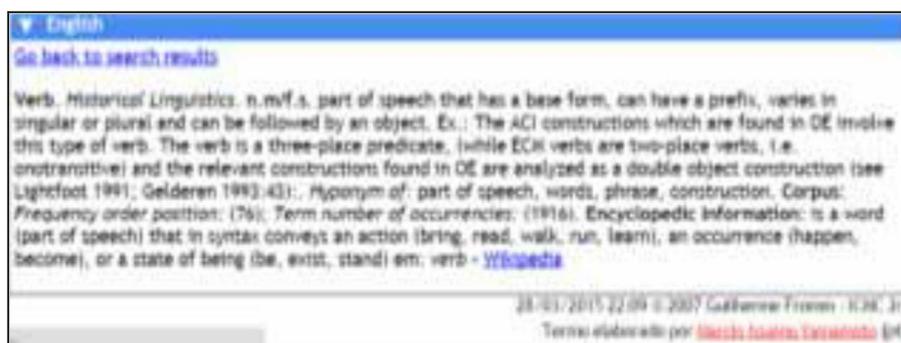


Fig. 12. VoTec - Visualização do termo *verbo* em inglês na área de linguística histórica

3. *Considerações finais*

A partir da pesquisa desenvolvida até o momento e a prática terminológica proporcionada pela plataforma do VoTec, podemos observar alguns indícios de resultados do que é a produção de uma obra terminográfica. Primeiramente foi necessário conhecer o que é o campo ou 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 terminográfica.

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 teoria comunicativa da terminologia permite que entendamos a unidade terminográfica que pode se identificar como unidade lexical ou terminológica, dependendo do seu contexto comunicativo de uso.

A linguística de *corpus* como metodologia, que parte da visão probabilística da língua, possibilita ao seu usuário o manejo de dados concretos, significativos quanto à representatividade e quanto ao uso em *corpora* autênticos.

O VoTec, ambiente de gestão terminológica que permite a construção de vocabulários bilíngues, possibilita a organização, administração e disposição de dados de uma forma prática, dispensando as fichas terminológicas físicas. A partir dos dados reunidos, podemos construir a macro e microestrutura do vocabulário. A macroestrutura segue como a de um dicionário impresso, contudo, cada termo é disposto por páginas individuais. As definições, em contraste, permitem o registro dos dados contidos nos *corpora*. A microestrutura permite a classificação do termo na subárea da linguística histórica, a classificação gramatical, traz exemplos, hipônimos, hiperônimos, posição de frequência no *corpus*, número de ocorrências, e um exemplo enciclopédico.

REFERÊNCIAS BIBLIOGRÁFICAS

AUBERT, Francis Henrik. *Introdução à metodologia da pesquisa terminológica bilíngue*. São Paulo: Humanitas, 1996. [2. ed. 2001, disponível em:

<<http://citrat.fflch.usp.br/sites/citrat.fflch.usp.br/files/u10/cad.%20terminologia%202.pdf>>].

BARROS, Lídia Almeida. *Curso básico de terminologia*. São Paulo: Edusp, 2004.

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

DUBOIS, Jean et alii. *Dicionário de linguística*. 16. ed. São Paulo: Cultrix, 1999.

FROMM, Guilherme. *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, São Paulo.

YAMAMOTO, Márcio Issamu. Linguística histórica e linguística de *corpus*: uma proposta para dicionário bilíngue português-inglês. *Revista Philologus*, Rio de Janeiro: CiFEFiL, ano 19, n. 55 – Suplemento, p. 691-705, 2013. Disponível em:

<<http://www.filologia.org.br/revista/55supl/058.pdf>>.