

VOBLING – UMA INTERSECÇÃO ENTRE *CORPUS* E UMA PLATAFORMA MULTIMODAL

Márcio Issamu Yamamoto*

Guilherme Fromm**

- **RESUMO:** Objetivamos refletir sobre a terminologia da Linguística analisada pela metodologia da Linguística de Corpus como pesquisa quantitativa e qualitativa. Para tanto, apresentamos, em parte, a construção do vocabulário bilíngue (português e inglês) da Linguística, baseado em *corpus*, denominado VoBLing, destinado a alunos iniciantes em Letras. Primeiramente, abordaremos a compilação desse *corpus* comparável, composto por 47 subáreas da Linguística. O registro pretendido foi acadêmico, a partir do qual descrevemos parte da terminologia da Linguística por meio de traços distintivos extraídos de linhas de concordância. As definições são construídas com base na definição terminológica e na definição enciclopédica previamente selecionadas pelo público-alvo. Em segundo lugar, essas funcionalidades foram organizadas em fichas terminológicas *on-line* do VoTec (Fromm, 2007), uma plataforma *on-line* bilíngue de gerenciamento terminológico. Portanto, este vocabulário *on-line* tem um enfoque terminológico e pedagógico, empregando uma abordagem multimodal, para introduzir conceitos linguísticos para alunos iniciantes em línguas. Os usuários têm acesso a definições e a diversos recursos pedagógicos que lhes permitem compreender os conceitos da Linguística e suas subáreas, o que a torna uma plataforma multimodal com potencial para mostrar a definição do termo em questão por meio da semiótica múltipla.
- **PALAVRAS-CHAVE:** VoBLing; Vocabulário bilíngue da Linguística; Linguística de Corpus; Terminologia bilíngue; Plataforma multimodal.

Introdução

Segundo Ducrot e Todorov (1972, p. 12), “o campo da Linguística não possui uma terminologia unificada”. Portanto, nossa proposta é refletir sobre a terminologia da Linguística analisada por uma pesquisa quantitativa e qualitativa baseada em conceitos da Linguística de Corpus (LC). Para tanto, descreveremos alguns dos passos para a

* Universidade Federal de Jataí (UFJ), Jataí, GO, Brasil. marcioiy@ufj.edu.br. ORCID: <https://orcid.org/0000-0001-9792-8187>.

** Universidade Federal de Uberlândia (UFU), Uberlândia – MG – Brasil. guifromm@ufu.br. ORCID: <https://orcid.org/0000-0001-5654-0135>

construção de um Vocabulário Linguístico Bilingue baseado em *corpus*. O *corpus* de estudo é composto por 47 subáreas da Linguística, organizadas em duas áreas principais: Linguística Descritiva e Linguística Aplicada.

Assim, neste artigo, focalizamos os seguintes aspectos da prática terminológica: (a) compilação de *corpus*, (b) organização de traços distintivos para definições,¹ (c) definição terminológica e enciclopédica e (d) VoBLing como plataforma multimodal.

Em primeiro lugar, o *corpus* compilado é composto por textos acadêmicos que abrangem 500 mil itens ou *tokens* em cada idioma para cada subárea linguística.² O mesmo número médio de *tokens* por subárea visava atender ao princípio de equilíbrio e representatividade de *corpus* proposto pela Linguística de Corpus. A análise quantitativa do *corpus* foi realizada por meio do WordSmith Tools (WST 7.0 e 8.0), conjunto de ferramentas que produz listas de palavras e listas de palavras-chave. A principal ferramenta do WST é o Concord, que mostra linhas de concordância (em formato KWIC,³ como podemos ver na Figura 1) a partir de textos sem formatação e ajuda os usuários a terem acesso aos contextos em que os termos se encontram.

Figura 1 – *Biolinguistics* – palavras-chave em contexto (KWIC) – visão parcial

N	Concordance
1	it is that makes us human. ↔(Boeckx& Grohmann 2007: 3-4) Biolinguistics is one vehicle to transport such ideas. ↔
2	appears that his talk of unifying mathematical Platonism and biolinguistics is mere metaphor because it depends on the
3	of such a complex and varied field-in-the-making as biolinguistics is no easy task. ↔We were confronted with
4	↔2002: 23â€24). But ↯of course the purpose of Chomskyan biolinguistics is to account for a supposedly entirely
5	1975), it appears to be an idea whose time has come . ↔ Biolinguistics is not yet a science - it is more a
6	reactions can and have been proposed. ↔The situation for biolinguistics is more parallel to historical developments in
7	do so. ↔We therefore think that the most important task for biolinguistics is to inform linguistic theorizing by putting a
8	many research programs. ↔In that light, a central question for Biolinguistics is which, if any, specialized structures have
9	evolved in the species. ↔4 What is to be done? ↔ Fortunately Biolinguistics is progressively moving from this "folk" (i.e.,
10	I try to deal with some issues that seem to me to be crucial if biolinguistics is to achieve the centrality in the human
11	far appeared, but it will surely become one, and very soon, if biolinguistics is to go on developing. ↯As noted above,
12	The Core Explanandum The central research topic in biolinguistics is a characterization and explanation of the
13	accepted piece; that is, acceptance for publication in Biolinguistics is contingent on complying with the editorial
14	↔The motivation behind much computer simulation in biolinguistics is the belief that language, like the traffic
15	object. ↔Conclusion I have argued that mathematical biolinguistics is based on the perfectly coherent concept of
16	object. Conclusion ↯I have ↯argued that mathematical biolinguistics is based on the perfectly coherent concept of
17	created the field of biolinguistics. ↔As should be obvious , biolinguistics is a supremely interdisciplinary enterprise,
18	biology, and cognitive science. ↔The Handbook of Biolinguistics is intended to be, first and foremost, an
19	when the interdisciplinarity that is characteristic of the field of biolinguistics is making a comeback at the forefront of
20	from a biological perspective. ↔While this core assumption of biolinguistics is not particularly new (Chomsky 1965,
21	of the brain. ↔Far from ignoring biology, the very essence of biolinguistics is to focus on the biological nature of the
22	or some other experiment. ↔But the long-range task of biolinguistics is to pursue, interpret, and reconcile the
23	human mind. ↔The aim of this essay is to show that if a real biolinguistics is intended to be achieved we ought to be

Fonte: Elaboração própria

¹ Aqui adotamos a expressão *traços distintivos* de Sager (1990, p. 26). Além disso, também podem ser denominados *traços semânticos* como em Pavel e Nolet (2001, p. 18).

² *Token* é o número de palavras dentro de um texto, um item. Se um texto for composto por 500 palavras, ele conterá 500 *tokens*.

³ Palavras-chave em contexto: a palavra (ou palavras) a ser analisada aparece, em uma cor diferente, na posição central da tela, criando uma coluna da mesma palavra em todas as linhas.

Em segundo lugar, após identificação do contexto a partir das linhas de concordância, procedeu-se à organização dos traços distintivos que seguiram a metodologia do VoTec (Fromm, 2007; Fromm; Lisboa, 2024), um ambiente *web* de tratamento terminográfico que permite aos usuários organizar traços distintivos em fichas terminológicas. Em terceiro lugar, os dados coletados e organizados nos arquivos do projeto VoBLing foram utilizados para criar dois tipos de definições: (1) uma definição terminológica e (2) uma definição enciclopédica, ambos os tipos previamente selecionados pelo público-alvo.

Por fim, como resultado da pesquisa, construímos um Vocabulário Bilingue de Linguística denominado VoBLing (Yamamoto, 2020) disponível *on-line*.⁴ O VoBLing é uma plataforma multimodal que fornece aos usuários elementos de microestrutura comuns a trabalhos terminológicos, incluindo definições e exemplos. Além disso, disponibiliza recursos multissemióticos, tais como vídeos e áudios com a pronúncia de verbetes tanto em português quanto em inglês.⁵ Além disso, oferece uma representação gráfica da estrutura conceitual do termo e sua posição no campo da Linguística. É preciso mencionar que este artigo faz parte de uma tese de doutorado, o que significa que os procedimentos contidos em toda a pesquisa em si são parcialmente descritos aqui.

Compilação do *Corpus* de Linguística

Para começar, é importante explicar que a compilação desse *corpus* comparável compreende 29 subáreas da Linguística Descritiva (LD) e 18 subáreas de Linguística Aplicada (LA), totalizando 47 subáreas da Linguística, nossa principal área temática, em português e em inglês, composta por textos acadêmicos.⁶ Além disso, também compilamos um *corpus* bilíngue comparável de manuais de Linguística para contarmos com uma seleção de textos voltados para estudantes, nos quais as definições poderiam aparecer em contextos mais simples. O processo serviu ao seu propósito, resultando num *corpus* de cerca de 2 milhões de *tokens* ao todo. Esse processo de compilação foi mais trabalhoso, pois a maioria dos manuais estava disponível apenas em cópias impressas, e tivemos que os passar por um processo de reconhecimento óptico de caracteres (OCR) para obter versões eletrônicas para serem analisados pelo WST. Ao final, os arquivos tiveram que ser convertidos para o formato TXT, para que pudéssemos alcançar um melhor desempenho usando o WST.

A compilação de textos acadêmicos foi realizada por alunos de graduação e pós-graduação dos cursos de Língua Inglesa e Portuguesa e Tradução durante um período de aproximadamente 10 anos, de 2010 a 2020, para fins de avaliação em sala de aula (Fromm; Yamamoto, 2021).

⁴ Disponível em: <http://vobling.votec.ileel.ufu.br>. Acesso em: 25 out. 2023. O VoBLing pode ser considerado uma nova instância do projeto VoTec, segundo Fromm e Lisboa (2024).

⁵ A ferramenta básica que usamos hoje em dia para escrever um texto como este, o computador, também mostra, através de seus ícones, que a linguagem não verbal ainda está em uso e recaptura a ideia dos hieróglifos egípcios, por exemplo. Os computadores permitem o entrelaçamento de textos escritos e orais, incluindo imagens, elementos táteis, arranjos espaciais e cores.

⁶ Teses de doutorado, dissertações de mestrado e artigos de periódicos sobre Linguística.

Como parte da metodologia do procedimento terminológico, focalizamos a nomenclatura das várias subáreas que coletivamente formam o campo da Linguística. Na Figura 2, pode-se observar um sistema de classificação com as 47 subáreas que estudamos, representando nossa proposição de uma Taxonomia Linguística.⁷ Essas disciplinas foram selecionadas e classificadas com base em entrevistas com especialistas e na disponibilidade do *corpus*.⁸

Figura 2 – Sistema de classificação da Linguística



Fonte: Elaboração própria

⁷ Entendemos que cada uma dessas subáreas tem um objeto de estudo, por exemplo, a Fonética estuda os sons possíveis nas línguas, a Morfologia estuda a organização das palavras em uma língua e assim por diante.

⁸ Em outras palavras, se existe uma subárea no sistema de classificação, significa que há, no mínimo, textos totalizando quinhentos mil *tokens* sob ela, demonstrando a importância dessa subárea. É relevante, porém, explicar que essa taxonomia apresentada está em constante mudança, pois novas subáreas da Linguística podem surgir a qualquer momento.

Parte dessas 47 subáreas linguísticas levantadas serviu como *corpus* de pesquisa e foram compiladas primeiramente pelos alunos de graduação de Fromm, depois pelos alunos de pós-graduação e, ao final, por Yamamoto, que conduziu o estudo para criar o VoBLing. Ao todo, o tamanho do *corpus* é de 46,4 milhões de *tokens*, e não 47 milhões de *tokens*, como seria esperado, devido a um *corpus* menor compilado para a subárea de Linguística Matemática em português. Como não havia artigos suficientes nessa subárea, foi possível recuperar apenas 220.245 *tokens* em português, embora tivéssemos 507.984 *tokens* em inglês.⁹ Consequentemente, seguindo o princípio de balanceamento na LC, reduzimos o *corpus* de inglês de Linguística Matemática para cerca de 220 mil *tokens* a fim de equiparar-se ao de português.

Após o processo de compilação, procedeu-se à limpeza, balanceamento e etiquetagem desse *corpus*. No início, os alunos podiam levantar tantos *tokens* quanto quisessem para uma subárea, desde que a quantidade final somasse, ao menos, 500 mil. O número de 500 mil *tokens* foi baseado na pesquisa de Fromm (2007), quando se constatou que um *corpus* menor do que esse não forneceria contextos explicativos ou definitórios suficientes para a extração de traços distintivos necessários na escrita de definições.¹⁰ Neste trabalho, chamaremos de tamanho padrão do *corpus*.

Dito isso, o primeiro passo dado para a limpeza do *corpus* foi a eliminação das seções de apresentação (como o resumo), referências e apêndices de cada texto, deixando apenas o corpo do próprio texto; a segunda etapa envolveu o balanceamento do *corpus*, que visava padronizar o tamanho do mesmo para cada subárea da Linguística. Isso foi necessário porque alguns *corpora* continham mais de 1 milhão de *tokens*, por exemplo. Também revisamos o *corpus* usando o Microsoft Word e suas ferramentas de correção gramatical. Foi um passo muito importante, uma vez que palavras com erros ortográficos ou que não foram perfeitamente processadas após o OCR poderiam afetar a contagem de palavras pelo WST. Finalmente, fornecemos a cada arquivo um cabeçalho, incluindo o título do texto original, seu *site* e data de coleta.

Definição – Organizando traços distintivos

A partir do *Corpus* de Linguística, descrito na seção anterior, foi possível identificar contextos que traziam traços distintivos da terminologia da Linguística, identificados em linhas de concordância (WordSmith Tools 7.0 e 8.0; Scott, 2016, 2020). O próximo passo foi organizar esses recursos em fichas terminológicas *on-line* disponibilizadas na plataforma VoBLing.

⁹ Nossa experiência na compilação de informações nesta subárea revelou que o estado da arte de um campo de especialidade em um idioma pode não ser necessariamente equivalente ao estado da arte do mesmo campo em outro idioma.

¹⁰ Contextos explicativos fornecem algumas ideias, características e modos de uso sobre o termo em análise. Contextos definitórios fornecem uma definição clara do significado do termo.

A capacidade do VoBLing de organizar traços distintivos em fichas terminológicas é explicada pela Semântica de *Frames* (Fillmore, 2006) e Terminologia Baseada em *Frames* (Faber Benítez; Marquez Linares; Vega Exposito, 2005) quando se trata de usar os traços semânticos mais frequentes de cada termo para escrever sua definição.

Fillmore (2006, p. 613, grifo e tradução próprios) define Semântica de *Frames* como:

[...] uma abordagem para descrever os significados de entidades linguísticas independentes (palavras, frases lexicalizadas e uma série de construções gramaticais especiais) apelando para os tipos de estruturas conceituais (*frames*) que fundamentam seus significados e que motivam seu uso.¹¹

De acordo com Fillmore, os *frames* incluem cenas visuais, estruturas institucionais, experiências enativas, crenças humanas, ações, experiências ou imaginações. Todos esses cenários de *frames* acabam sendo expressos pelas escolhas linguísticas dos falantes ou por um conjunto de palavras e escolhas gramaticais que eles usam. Esses *frames* estão interligados e são ativados na memória pelo material linguístico, sendo um componente essencial para a definição de palavras (Fillmore, 1975, p. 124; 2003, p. 263).

A conexão entre a Semântica dos *Frames* e a Terminologia está no conceito de *frames* segundo Faber Benítez, Marquez Linares e Vega Exposito (2005, p. 2, tradução própria), quando os autores afirmam que

Um *frame* tem sido mais amplamente definido como qualquer sistema de conceitos relacionados, de tal forma que um conceito evoca todo o sistema. Nesse sentido, guarda uma evidente afinidade com a terminologia, que também se baseia nessa organização conceitual.¹²

Os autores discutem a existência de um sistema de conceitos e um conceito que evoca o próprio sistema. Ao preencher as fichas terminológicas do VoBLing, ficou clara a identificação desse fenômeno, ou seja, embora os autores estivessem falando de um termo específico, por exemplo: *phonology* (ver Figura 3), vários traços distintivos foram utilizados para sua definição.

Mais uma vez, Faber Benítez, Marquez Linares e Vega Exposito (2005, p. 4, tradução própria) explicam esse processo afirmando que,

¹¹ Original: “an approach to describing the meanings of independent linguistic entities (words, lexicalized phrases, and a number of special grammatical constructions) by appealing to the kinds of conceptual structures (*frames*) that underlie their meanings and that motivate their use” (Fillmore, 2006, p. 613). Todas as traduções aqui apresentadas são de nossa autoria.

¹² Original: “A *frame* has been more broadly defined as any system of concepts related in such a way that one concept evokes the entire system. In this sense, it bears an obvious affinity with terminology, which is also based on such conceptual organization” (Faber Benítez; Marquez Linares; Vega Exposito, 2005, p. 2).

Na construção de uma rede de *frames*, a classificação está envolvida, uma vez que essas redes são divididas em domínios, os domínios em *frames*, e os *frames* podem passar por vários níveis de especificidade usando uma herança hierárquica. Os dados são extraídos por meio da análise de *corpus* para codificar a estrutura proposicional subjacente e definir os papéis semânticos. Os elementos de um *frame* podem ser compartilhados com outros *frames* porque um objeto lexical pode ter vários significados, ou o mesmo significado dicionarizado pode ter diferentes significados sociais (conotativos) entre situações.¹³

O VoBLing permite que os pesquisadores copiem e coleem, para o banco de dados, determinados contextos de termos, previamente identificados pela ferramenta Concord no WST, a partir dos seus arquivos originais para as fichas terminológicas *on-line*, para que os termos possam ser analisados em seus contextos originais, como mostra a Figura 3.¹⁴

Figura 3 – Contextos para *phonology* no VoBLing

Exemplo	Conceito	Fonte	Ações
Phonology is the rules of how a language sounds and how and when certain sounds can be combined.	rules of how a language sounds; how and when certain sounds can be combined	PDF 2 13/10/2019	editar - excluir
Phonology is the study of how a language arranges sounds in predictable patterns; inventory of sounds: 'phonemes'; distribution of sounds: word position, syllable structure, consonant clusters, processes of sound change; spelling : sounds and letters.	the study of how a language arranges sounds in predictable patterns; inventory of sounds: 'phonemes'; distribution of sounds: word position, syllable structure, consonant clusters, processes of sound change; spelling : sounds and letters	PDF 2 20/06/2013	editar - excluir
[...] phonology is the study of units of sound and the patterned variation of their perception and production in the context of other sounds. Such spoken variation is perceived by speakers and is then very often reflected as repeated and patterned spellings or misspellings in their writing.	study of units of sound; the patterned variation of their perception; production in the context of other sounds; perceived by speakers; reflected in their writing	PDF 2 20/06/2013	editar - excluir
Arguably, the main function of phonology is to repair the damage that morphological rules.[...] would cause.	main function of phonology: repair the damage that morphological rules would cause	PDF 2 24/03/2015	editar - excluir
"Phonology is the act of talking about how phonemes function in a language and the relationships among the different phonemes". Under the study of phonology, we have the segmental features and the supra segmental or the prosodic features.	the act of talking about how phonemes function in a language; the relationships among the different phonemes; the study of segmental features and the supra segmental or the prosodic features	PDF 2 18/06/2019	editar - excluir

Fonte: Elaboração própria

Ao examinar os exemplos de *phonology* extraídos de linhas de concordância (ver Figura 3), fica evidente que os autores empregam níveis variados de especificidade em suas definições do termo, como é mostrado na coluna 1 – Exemplo. Para citar alguns

¹³ Original: "In building a frame network, classification is involved since these networks are divided into domains, the domains into frames, and the frames can go through several levels of specificity by using hierarchical inheritance. Data is extracted by means of corpus analysis to encode underlying propositional structure and define semantic roles. The elements of a frame may be shared with other frames because a lexical object can have several meanings, or the same dictionary meaning may have different social (connotative) meanings across situations" (Faber Benítez; Marquez Linares; Vega Exposito, 2005, p. 4).

¹⁴ Embora o consultor possa alternar entre os idiomas disponíveis (inglês ou português) na página de consulta VoBLing, os *layouts* de tela no banco de dados são apresentados apenas em português.

exemplos de várias maneiras de definir *phonology*, podemos considerá-la como um conjunto de regras sonoras, ou o estudo de unidades de som, ou o ato de falar sobre como os fonemas funcionam.

No quadro da Figura 3, também podemos notar mais três colunas: Conceito (uma tentativa de resumir os exemplos fornecidos pelo pesquisador, a fim de levantar seus conceitos centrais), Fonte e Ações (com possibilidade de exclusão ou edição do exemplo). A coluna Conceito é composta por características distintivas pré-resumidas, criadas pelos pesquisadores, que os auxiliarão na criação do conceito final e, conseqüentemente, na formulação da definição final. A coluna Fonte designa o formato de texto original do *corpus*, que, neste caso, está em formato PDF. Por fim, a coluna Ações permite que os pesquisadores editem ou excluam os dados dentro dos arquivos da ficha terminológica.

A próxima tarefa é organizar os traços distintivos apresentados pelos conceitos em linhas e colunas (ver Figura 4) — as linhas são dispostas de acordo com o número de exemplos coletados, as colunas podem ser adicionadas de acordo com a necessidade do pesquisador. Semas sinônimos devem ser alocados nas mesmas colunas. Dessa forma, o quadro demonstra quais conceitos são mais recorrentes, determinados pelos itens lexicais empregados pelos autores, convergindo para um conceito central. As características semânticas menos recorrentes são constituintes de outros *frames*, porém fazendo parte de um mesmo sistema conceitual.

Figura 4 – Arquivo de caso VoBLing – *phonology*.

The screenshot shows a web interface with a navigation bar at the top containing tabs: Dados, Traços Distintivos (selected), Semântica, Termo Equivalente, Termos Remissivos, Informações Enciclopédicas, and Multimídia. Below the tabs is a text input field labeled 'Nova Coluna' with a note '— Não use aspas duplas'. The main content is a table with 5 rows and 6 columns. The columns are labeled #, A, B, C, D, and E. The rows contain the following data:

#	A	B	C	D	E
1		rules of h...	how and wh...		
2	study	how a lang...	inventory..	distributi...	spelling: ...
3	study	units of s...	the patter...	production...	reflected ...
4					main funct...
5		the act of...	the relati...	the study ...	

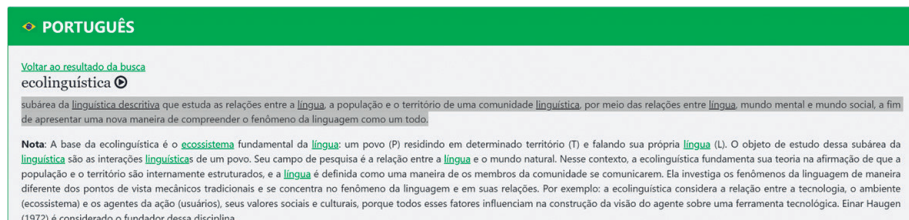
Fonte: Elaboração própria

Na próxima seção, demonstraremos como os traços distintivos foram organizados para criar tanto uma definição terminológica (para o campo Definição) quanto uma definição enciclopédica (para o campo Nota) na microestrutura. Forneceremos a definição elaborada para Ecolinguística, como uma subdisciplina da Linguística Descritiva, que se insere no domínio mais amplo da Linguística.

Ecolinguística – Definição Terminológica e Enciclopédica

Após a organização dos traços semânticos, o pesquisador inicia o processo de escrita das definições com base em dois padrões distintos. O primeiro padrão é a definição terminológica, empregando os princípios de gênero próximo e diferença específica. O segundo padrão é a definição enciclopédica. Vale ressaltar que esses dois padrões foram previamente selecionados a partir de entrevistas realizadas com potenciais usuários, especificamente entre calouros de Língua e Literatura Inglesa e Portuguesa da Universidade Federal de Uberlândia e da Universidade Federal de Jataí.¹⁵

Figura 5 – VoBLing – Definição de Ecolinguística



PORTUGUÊS

[Voltar ao resultado da busca](#)
ecolinguística

subárea da linguística descritiva que estuda as relações entre a língua, a população e o território de uma comunidade linguística, por meio das relações entre língua, mundo mental e mundo social, a fim de apresentar uma nova maneira de compreender o fenômeno da linguagem como um todo.

Nota: A base da ecolinguística é o **ecossistema** fundamental da **língua**: um povo (P) residindo em determinado território (T) e falando sua própria língua (L). O objeto de estudo dessa subárea da **linguística** são as interações **linguísticas** de um povo. Seu campo de pesquisa é a relação entre a **língua** e o mundo natural. Nesse contexto, a ecolinguística fundamenta sua teoria na afirmação de que a população e o território são internamente estruturados, e a **língua** é definida como uma maneira de os membros da comunidade se comunicarem. Ela investiga os fenômenos da linguagem de maneira diferente dos pontos de vista mecânicos tradicionais e se concentra no fenômeno da linguagem e em suas relações. Por exemplo: a ecolinguística considera a relação entre a tecnologia, o ambiente (ecossistema) e os agentes da ação (usuários), seus valores sociais e culturais, porque todos esses fatores influenciam na construção da visão do agente sobre uma ferramenta tecnológica. Einar Haugen (1972) é considerado o fundador dessa disciplina.

Fonte: Elaboração própria

Na Figura 5, é possível ler a definição terminológica em destaque escrita para Ecolinguística. Primeiro, começa por colocar a disciplina em um sistema conceitual maior, a Linguística Descritiva, que, por sua vez, pertence a um sistema conceitual ainda maior, que é a Linguística. Essa compreensão é possibilitada pelo recorte da estrutura conceitual, apresentada como imagem, que será explicitada na seção 4 deste artigo. Em segundo lugar, indica o objeto de estudo da disciplina: *estuda as relações entre a língua, a população e o território de uma comunidade linguística, por meio das relações entre língua, mundo mental e mundo social*. Por fim, especifica o objetivo do estudo: *apresentar uma nova maneira de compreender o fenômeno da linguagem como um todo*.

Na segunda parte da Figura 5, em Nota, os usuários podem encontrar a definição enciclopédica, que fornece uma explicação detalhada da disciplina: *a base da ecolinguística é o ecossistema fundamental da língua: um povo (P) residindo em determinado território (T) e falando sua própria língua (L)*. Oferece também detalhes sobre análises tradicionais de língua e a relação entre os seres humanos e seu ambiente linguístico: *Ela investiga os fenômenos da linguagem de maneira diferente dos pontos de vista mecânicos tradicionais e se concentra no fenômeno da linguagem e em suas relações. Por exemplo: a ecolinguística considera a relação entre a tecnologia, o ambiente (ecossistema) e os agentes da ação (usuários), seus valores sociais e culturais, porque todos esses fatores influenciam na construção da visão do agente sobre uma ferramenta tecnológica*.

¹⁵ Código de aprovação do Comitê de Ética em Pesquisa CAAE: 80945717.8.0000.5152.

Para concluir esta seção, recorreremos à citação anterior de Faber Benítez, Marquez Linares e Vega Exposito (2005, p. 4) ao afirmar que “*frames* podem passar por vários níveis de especificidade usando uma herança hierárquica”. Aplicamos essa definição exemplificada em razão das conexões entre os campos da Ecologia e da Linguística. Nesse caso, esses dois quadros principais se entrelaçam, dando origem a um terceiro quadro, que é o campo da Ecolinguística e sua especificidade.

VoBLing como uma plataforma multimodal

Antes de explicar cada passo dado para construir o VoBLing (2020) como uma plataforma multimodal, discutiremos brevemente o conceito de definição de multimodalidade. Segundo Sabino-Luiz (2023), uma definição multimodal refere-se “à integração da linguagem verbal com outras formas de linguagem não verbal, tanto humanas quanto não-humanas, tendo o potencial de mostrar o significado do referente em questão, utilizando múltiplas semioses” (Sabino-Luiz, 2023, p. 1). A implementação de uma abordagem multimodal pode ser valiosa para garantir uma compreensão abrangente dos conceitos retratados no VoBLing. Além disso, o VoBLing, como vocabulário eletrônico, traz como vantagens sua acessibilidade, seus elementos multimídia (como áudio, vídeo, imagens) e suas atualizações frequentes. Finalmente, vale mencionar a importância de bancos de dados personalizados ou *corpora* contendo imagens e áudio para terminografia multimodal.

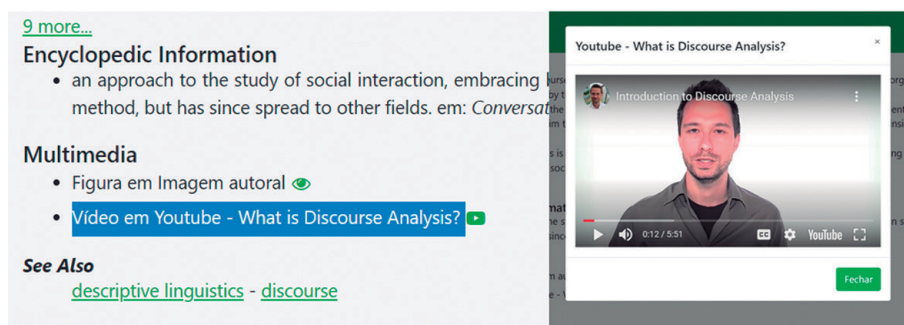
Sabino-Luiz (2023, p. 5) explica que, na contemporaneidade, o uso da linguagem não verbal pode ser uma técnica útil e relevante para o desenvolvimento de dicionários, desde que aplicada com base em princípios teóricos e não meramente como uma escolha estética. Portanto, as ilustrações e também os vídeos apresentam funções cognitivas e semióticas importantes quando devidamente associados aos verbetes de um dicionário ou de um vocabulário, pois auxiliam o usuário na compreensão da unidade lexical ou terminológica pesquisada. Eles cumprem duas funções cognitivas importantes na aprendizagem de línguas, tanto complementando quanto exemplificando as informações verbais no dicionário, mostrando ao usuário o que é a coisa definida. No entanto, é essencial que lexicógrafos e terminógrafos apliquem critérios relevantes na seleção de recursos multimodais em seus dicionários/vocabulários para garantir que esses recursos contribuam para uma compreensão mais abrangente e precisa dos significados das unidades lexicais.

Contextualizando a explicação do parágrafo anterior em relação ao nosso vocabulário, podemos dizer que, além das definições, os usuários do VoBLing têm acesso a outros recursos de aprendizagem: (1) vídeos especializados com explicações de conceitos em inglês e português; (2) exemplos de uso da linguagem extraídos do *corpus*; (3) arquivos de áudio com a pronúncia do verbo em ambos os idiomas; (4) recorte da estrutura conceitual que mostra o termo dentro do campo da Linguística e suas subáreas; (5) remissivas, que aparecem ao passar o mouse sobre os *hiperlinks*,

o que permite ao usuário acesso aos conceitos linguísticos sem precisar pesquisar outras páginas; (6) e, ao clicar em *hiperlinks* enciclopédicos, os usuários receberão mais informações enciclopédicas sobre o termo pesquisado. Esses recursos foram planejados para aprimorar a compreensão do verbete e oferecer recursos mais modernos e responsivos, configurando uma estrutura multimodal.

Primeiramente, selecionamos vídeos especializados do YouTube, com base no número de visualizações como parâmetro quantitativo, que fornecem explicações sobre o termo que está sendo definido no VoBLing. Em seguida, esses vídeos foram analisados e avaliados pelos pesquisadores para garantir a confiabilidade e facilidade de entendimento para estudantes iniciantes. Em terceiro lugar, em termos de duração, os vídeos deveriam variar de cinco a quinze minutos. Finalmente, se esses critérios fossem atendidos, os *links* dos vídeos eram enviados para o VoBLing, permitindo que os usuários os acessassem (veja a Figura 6).

Figura 6 – Vídeos VoBLing – *Discourse Analysis*.



Fonte: Elaboração própria

O uso de vídeos como ferramenta de ensino e aprendizagem é altamente produtivo. Os vídeos permitem que os usuários se familiarizem, no caso desse exemplo, com a terminologia da Linguística em inglês, aprendam padrões corretos de pronúncia e se preparem para ler textos de Linguística em inglês.

Em segundo lugar, o VoBLing fornece exemplos de uso da linguagem com base em um *corpus* (veja a Figura 7), permitindo que os usuários acessem trechos originais sobre Linguística em inglês (neste caso). Como esse *corpus* consiste principalmente de textos escritos em inglês, ele melhora significativamente a compreensão e o aprendizado da terminologia da Linguística, bem como do inglês como língua estrangeira. Como a maioria dos alunos brasileiros é iniciante em termos de proficiência em inglês, oferecer acesso a esses textos contribui para seu desenvolvimento linguístico e aprofundamento de sua compreensão da língua e literatura inglesas.

Figura 7 – VoBLing – *Applied Linguistics* – exemplos de uso

Examples

1. Applied linguistics is divided into several distinct areas that would be of interest to students and others who want to know what practical applications linguistics has.
2. Nevertheless, in contexts where AL does remain the major academic domain within which language-related practices are theorized, and where AL constitutes the main arena for explicit discussion of research methodology, [...]
3. The term 'applied linguistics' appears first to have been used in the late 1940s when the discipline embraced the teaching and learning of second and foreign languages (Johnson and Johnson 1998), but since then the discipline has expanded to cover a wider range of sub-disciplines or 'subfields' as they are called by Bachman and Cohen (1998).
4. First, it should be noted that most existing accounts of applied linguistics and language teaching history emphasize the development of theories in the abstract rather than paying much attention to practice (including contexts for the production of theories), and the false impression can thereby be conveyed that history has everywhere been the same.
5. The main distinguishing characteristic of AL (Applied Linguistics) is its concern with professional activities whose aim is to solve 'real-world' language-based problems, which means that research touches on a particularly wide range of issues-psychological, pedagogical, social, political and economic as well as linguistic. As a consequence, AL research tends to be interdisciplinary (Hudson 1999).

Fonte: Elaboração própria

Outra vantagem do uso desses exemplos advindos de *corpus* é que os usuários podem explorar a historiografia da Linguística dentro de um contexto diferente daquele dos falantes de português, como ilustrado na Figura 7, exemplo 3. Além disso, a exposição direta a textos escritos em inglês ajuda a reduzir a interferência de sua língua materna, pois a leitura desses textos contribui para o processo de aprendizagem da língua inglesa.

Em terceiro lugar, no que diz respeito à terminologia da Linguística, uma parcela significativa do vocabulário é derivada do latim, o que pode ser confuso para os falantes do português. Embora as palavras escritas sejam bastante semelhantes, a pronúncia pode diferir. Para resolver esse problema, o VoBLing fornece arquivos de áudio com pronúncia em inglês e português, ambos gravados por falantes nativos (consulte a Figura 8).¹⁶

Figura 8 – Arquivo de áudio de entrada do VoBLing



Fonte: Elaboração própria

Como mostrado na Figura 8, há um ícone de áudio no qual os usuários podem clicar para ouvir a pronúncia de uma entrada. Embora esse recurso seja comum em muitos dicionários de idiomas *on-line*, o que diferencia o VoBLing é sua pronúncia de terminologia específica da Linguística com termos multpalavras, o que geralmente não é encontrado em dicionários padrão. Termos multpalavras, muitas vezes, não estão disponíveis em dicionários *on-line* bem conhecidos. Por exemplo, uma busca por

¹⁶ Os arquivos em inglês foram gravados pelos ETAs da Fulbright Gautam Ramesh e Ruben Adery (da Linguistix Pronunciation, <https://www.linguistixpro.com/>). Os arquivos em português foram gravados pelo primeiro pesquisador do VoBLing.

linguistic atlas produz os seguintes resultados: (a) No Oxford Learner’s Dictionaries:¹⁷ *Nenhuma correspondência exata encontrada para “linguistic atlas” em inglês;* (b) No Merriam-Webster.com,¹⁸ uma definição estava disponível, mas não apresentava arquivo de áudio para pronúncia (ver Figura 8); (c) O dicionário Cambridge oferece opções alternativas para *linguistic atlas*,¹⁹ mas o termo em si não está disponível em sua plataforma. Suas sugestões incluem: *Sugestões de busca para atlas linguísticos – Temos essas palavras com grafias ou pronúncias semelhantes: linguística (substantivo), linguística (adjetivo), ciência linguística.*²⁰

Figura 9 – Busca Merriam Webster: *Linguistic atlas*



The image shows a screenshot of the Merriam-Webster website. At the top, there is a navigation bar with links for 'GAMES & QUIZZES', 'THESAURUS', 'WORD OF THE DAY', 'FEATURES', and 'SHOP'. The Merriam-Webster logo and 'SINCE 1828' are on the left. A search bar contains the text 'linguistic atlas'. Below the search bar, there are two buttons: 'Dictionary' and 'Thesaurus'. The main content area displays the word 'linguistic atlas' in a large font, followed by 'noun'. There is a 'Save Word' button with a bookmark icon. Below this is the 'Definition of linguistic atlas' section, which states: 'a publication containing a set of maps on which speech variations are recorded — called also *dialect atlas*'. The 'First Known Use of linguistic atlas' section indicates '1917, in the meaning defined above'. At the bottom of the screenshot, it says 'Fonte: Merriam Webster'.

Com base nos exemplos fornecidos acima, torna-se evidente a importância de oferecer a pronúncia dos termos multipalavras. Os usuários, muitas vezes, enfrentam dificuldades para localizar essas pronúncias em dicionários de idioma padrão. Nesse ponto, revisitamos Sabino-Luiz (2023, p. 16), que enfatiza os benefícios da multimodalidade. O autor afirma que, independentemente do tipo de dicionário eletrônico empregado, seu notável potencial na multimodalidade é inquestionável. Ele pode combinar perfeitamente diversas formas de linguagem, englobando texto, imagens, sons e vídeos, a fim de fornecer uma representação mais completa e expandida dos significados das palavras e expressões lexicais. Essa abertura para outras modalidades textuais permite que os

¹⁷ Pesquisa disponível em: <https://www.oxfordlearnersdictionaries.com/spellcheck/english/?q=linguistic+atlas>. Acesso em: 26 out. 2023.

¹⁸ Pesquisa disponível em: <https://www.merriam-webster.com/dictionary/linguistic%20atlas>. Acesso em: 26 out. 2023.

¹⁹ Pesquisa disponível em: <https://dictionary.cambridge.org/spellcheck/english-portuguese/?q=linguistic+atlas>. Acesso em: 26 out. 2023.

²⁰ Original: “*Search suggestions for linguistic atlas – We have these words with similar spellings or pronunciations: linguistics, linguistic, linguistic science*”.

usuários tenham uma melhor compreensão dos significados das palavras, tornando o processo de aprendizagem mais dinâmico e interativo. A versatilidade faz com que o dicionário eletrônico possa ser atualizado com mais facilidade e frequência, tornando-o mais preciso e atualizado do que as versões impressas.


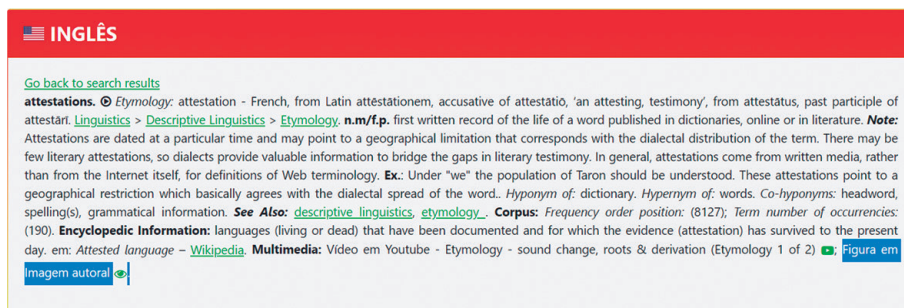
Quarto, VoBLing exibe um trecho da estrutura conceitual que situa o termo dentro do campo da Linguística e suas disciplinas relacionadas quando os usuários clicam no ícone  ao final da definição.

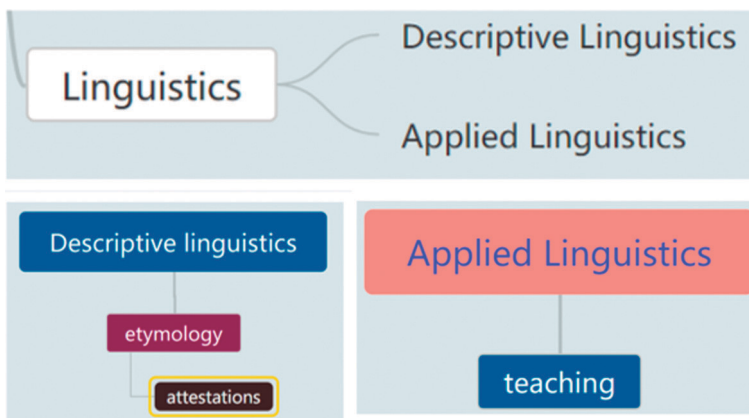
Figura 10 – VoBLing – imagem da estrutura conceitual do termo *attestations*



Fonte: Elaboração própria

A Figura 10 exibe o ícone para revelar a estrutura conceitual de *attestations* dentro da subárea de Etimologia. Quando os usuários clicam nesse ícone, eles podem visualizar uma imagem como mostrada na Figura 11.

Figura 11 – Estrutura conceitual do VoBLing



Fonte: Elaboração própria

A imagem superior ilustra a Linguística Descritiva e a Linguística Aplicada como as duas áreas primárias dentro da Linguística. A imagem inferior esquerda retrata *attestations* como um termo ligado à Etimologia e subordinado à Linguística Descritiva. A imagem inferior direita exhibe *teaching* como um termo pertencente à Linguística Aplicada. Essas imagens simplificam e deixam claro o aprendizado de conceitos e suas relações dentro do campo temático principal, que é a Linguística, facilitando a compreensão pelos usuários.

Em quinto lugar, o VoBLing fornece remissivas *pop-up*. Os usuários podem passar o *mouse* sobre *hiperlinks* dentro da definição ou em Nota para ler sobre conceitos linguísticos sem precisar acessar outras páginas.

Figura 12 – VoBLing – remissiva *pop-up* para o termo *descriptive linguistics*

Contraste x

Termo	descriptive linguistics
Definição	subarea of linguistics that describes languages grounded in systematic empirical observation and conceives language as a highly complex symbolic system
Nota	descriptive linguistics is a branch of general linguistics that deals with the fundamental aspects of languages. It studies linguistic forms and the regularities of their distributions, describing written and spoken texts, based on native speaker introspection, rejecting the normative prescription of one specific style. It documents language descriptions such as text types and authorship. Descriptive linguistics bears subdivisions such as historical linguistics, forensic linguists, phonetics, phonology, lexis, syntax, semantics, pragmatics, discourse and text analysis

Fonte: Elaboração própria

Na Figura 12, *descriptive linguistics* é definida baseada nos semas da definição de *phonetics* (como mostra a Figura 13). Além das remissivas *pop-up*, há também *hiperlinks* disponíveis, permitindo que os usuários acessem outros termos que fazem parte da microestrutura e já foram registrados no banco de dados. Mais detalhes serão fornecidos na explicação na Figura 13.

Figura 13 – VoBLing – Definição de *phonetics* e suas remissivas

🇺🇸 **INGLÊS**

[Go back to search results](#)

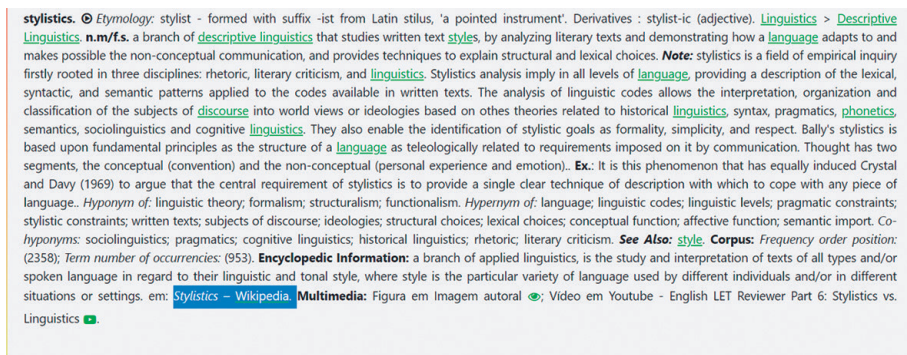
phonetics. *Etymology:* phonetic, adjective (adj.) - Modern Latin *phoneticus*, from Greek *φωνητικός*, 'vocal', from *φωνητός*, 'utterable', verbal adj. of *φωνειν*, 'to speak clearly, utter', from *φωνή*, 'sound, voice'. Derivatives: *phonet-ics*, noun. [Linguistics](#) > [Descriptive Linguistics](#). **n.m/f.s.** a branch of [descriptive linguistics](#) that studies the physical nature of speech sounds and how they are produced by the human body, explains how air moves through some sort of obstruction in and out of the lungs, what sound waves to understand how the human ear processes speech and how it is interpreted. **Note:** phonetics is the study of general mechanisms of speech production, such as consonants and vowels (segmental sounds), stress, intonation, tone, intensity, and speed (suprasegmental sounds), and how they are perceived and expressed by humans. This classifies all sounds (phonemes) that occur in all [languages](#), and in 1888 International Phonetic Alphabet (IPA) was developed for this purpose. Phonetics is interrelated with phonology and phonics, and interacts with the psychological and sociological aspects of human speech; its applications include the use of transcriptions in dictionaries and [language](#) textbooks to represent word pronunciation. **Ex.:** The

Fonte: Elaboração própria

Outra característica está representada na Figura 13: os termos *linguistics* e *descriptive linguistics* aparecem antes da definição de *phonetics*. Os usuários podem acessar suas definições ao clicá-los, enquanto *descriptive linguistics* e *language(s)*, termos dentro da definição de *phonetics*, são disponibilizados como uma janela *pop-up*.

Além disso, há um *link* enciclopédico que conecta termos a fontes externas de informação, como a Wikipédia ou outras (como mostrado na Figura 14).

Figura 14 – VoBLing – *Link* do termo *stylistics* para a Wikipédia



Fonte: Elaboração própria

Na Figura 14, o *link* destacado em azul se conecta à Wikipédia. Além das informações terminológicas, imagens, *pop-ups* e vídeos, os usuários também têm a opção de acessar *links* externos que conectam o termo a informações disponíveis gratuitamente na Wikipédia ou *sites* similares. Na Figura 14, a primeira linha também exhibe a etimologia do termo, que foi resgatada de livros e parcialmente reproduzida para enriquecer as definições de termos.²¹

Comentários finais

Após a análise de um *corpus* substancial da Linguística, observamos que a terminologia dentro desse campo permanece fragmentada, assim como Todorov observou na década de 1970. No entanto, com a criação do VoBLing, podemos propor uma terminologia linguística em potencial baseada na frequência, extraída de um *corpus* representativo e utilizando a metodologia da Linguística de *Corpus*.

Seguindo os passos descritos neste artigo, fica evidente que esta pesquisa é viável, apesar do tempo dispendido. A disponibilidade de ferramentas como o WST e a metodologia de Linguística de *Corpus* garantem que os resultados sejam de alta qualidade, tanto quantitativa quanto qualitativamente.

²¹ Os livros utilizados como fontes etimológicas do VoBLing foram *Dicionário Etimológico da Língua Portuguesa* de Nascentes (1955), para o português, e *Comprehensive Etymological Dictionary of the English Language* de Klein (1971) e *Origins: A Short Etymological Dictionary of Modern English* de Partridge (1966), para o inglês.

A pesquisa e o produto final demonstram que os componentes adicionais nas microestruturas dos verbetes não são apenas desejáveis, mas também necessários para uma comunidade mais ampla de estudiosos e entusiastas da linguagem. Embora projetado para que os estudantes de línguas entendam melhor os termos com que trabalham no seu dia a dia na universidade, o VoBLing está disponível gratuitamente. Isso significa que qualquer pessoa interessada nos campos da Linguagem, da Linguística e da Tradução pode acessá-lo – se as definições e exemplos não forem suficientes para uma compreensão abrangente de um verbete, suas características multimodais, que vão muito além das informações verbais tradicionais encontradas em dicionários, vocabulários e glossários impressos, podem auxiliar aqueles que buscam uma interpretação mais aprofundada das informações fornecidas.

Esta pesquisa e desenvolvimento do produto abrem caminho para que profissionais da linguagem explorem novas ideias que podem ser incorporadas em futuras ferramentas computacionais para trabalhos lexicográficos e terminográficos. Isso inclui não apenas o que está disponível para o público geral ou específico de consultantes, mas também considerações sobre a estrutura do banco de dados. É importante levar em conta tanto os usuários esperados quanto os inesperados. A interação contínua de novas linguagens de programação, recursos audiovisuais, colaboração *wiki* e mídias sociais oferecem *insights* instigantes sobre o futuro das obras de referência.

YAMAMOTO, Márcio Issamu; FROMM, Guilherme. Vobling – an intersection between corpus and a multimodal platform. *Alfa*, São Paulo, v. 68, 2024.

- *ABSTRACT: We aim to reflect on the terminology of Linguistics analyzed by Corpus Linguistics methodology as a quantitative and qualitative research. To do so, we partly present the construction of the corpus-based bilingual vocabulary (Portuguese and English) of Linguistics, called VoBLing, designed for beginner languages students. First, we will address the compilation of this comparable corpus, composed of 47 subareas of Linguistics. The register aimed was academic, from which we describe part of the terminology of Linguistics through distinctive features extracted from concordance lines. Definitions are constructed based on the terminological definition and the encyclopedic definition previously selected by the target audience. Secondly, these features were organized in online terminology card files of VoTec (Fromm, 2007), an online bilingual terminology management platform. Therefore, this online vocabulary has both a terminological and a pedagogical focus, employing a multimodal approach, to introduce linguistic concepts to beginner language students. The users have access to definitions and to several pedagogical resources that allow them to understand the concepts of Linguistics and its disciplines, what makes it a multimodal platform having the potential to show the definition of the term in question by using multiple semiotics.*
- *KEYWORDS: VoBLing; Bilingual vocabulary of Linguistics; Corpus linguistics; Bilingual Terminology; Multimodal platform.*

REFERÊNCIAS

DUCROT, O.; TODOROV, T. **Dictionnaire encyclopédique des sciences du langage**. Paris: Seuil, 1972.

FABER BENÍTEZ, P.; MARQUEZ LINARES, C.; VEGA EXPOSITO, M. Framing Terminology: A Process-Oriented Approach. **Meta**, v. 50, n. 4, dez. 2005. Disponível em: <https://doi.org/10.7202/019916ar>. Disponível em: <https://www.erudit.org/fr/revues/meta/2005-v50-n4-meta1024/019916ar.pdf>. Acesso em: 25 out. 2023.

FILLMORE, C. J. An Alternative to Checklist Theories of Meaning. *In: Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 1975. p. 123-131.

FILLMORE, C. J. Double-Decker Definitions: The Role of Frames in Meaning Explanations. **Sign Language Studies**, Volume: 3, Issues: 3, Gallaudet University Press, p. 263-295, 2003. Disponível em: <https://doi.org/10.1353/sls.2003.0008>. Acesso em: 25 out. 2023.

FILLMORE, C. J. Frame Semantics. *In: Encyclopedia of Language & Linguistics*. Elsevier, 2006. p. 613-620. Disponível em: <https://doi.org/10.1016/B0-08-044854-2/00424-7>. Acesso em: 28 jul. 2022.

FROMM, G. **VoTec**: a construção de vocabulários eletrônicos para aprendizes de tradução. 2007. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007. Disponível em: <https://doi.org/10.11606/T.8.2008.tde-08072008-150855>. Acesso em: 28 jul. 2022.

FROMM, G.; YAMAMOTO, M. I. Compilação, reciclagem e padronização de um Corpus Colaborativo de Linguística: percursos metodológicos. **Revista de Estudos da Linguagem**, [S. l.], v. 29, n. 3, p. 2041-2078, 2021. Disponível em: <http://dx.doi.org/10.17851/2237-2083.29.3.2041-2078>. Acesso em: 26 jun. 2024.

FROMM, G.; LISBOA, J. V. R. VoTec terminographic environment over the years: brief overview. **Acta Scientiarum. Language and Culture**, v. 45, n. 2, p. e67669, 23 fev. 2024. Disponível em: <https://doi.org/10.4025/actascilangcult.v45i2.67669>. Acesso em: 26 jun. 2024.

PAVEL, S.; NOLET, D. **Handbook of Terminology**. Adapted into English by Christine Leonhardt. Ottawa, Translation Bureau, Terminologie and Standardization Directorate, 2001.

SABINO-LUIZ, M. Explorando a Definição Multimodal: um estudo sobre a integração de elementos multimodais em dicionários impressos e eletrônicos. **Revista GTLex**, Uberlândia, v. 8, n. 1, p. e0810, 2023. Disponível em: <https://doi.org/10.14393/Lex-v8a2022/23-10>. Acesso em: 1 set. 2023.

SCOTT, M. **WordSmith Tools**. Version 7. Stroud: Lexical Analysis Software, 2016.

SCOTT, M. **WordSmith Tools**. Version 8. Stroud: Lexical Analysis Software, 2020.

YAMAMOTO, M. I. **VoBLing**: vocabulário bilíngüe de linguística, português-inglês, direcionado por *corpus*. 2020. 214 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Federal de Uberlândia, Uberlândia, 2020. Disponível em: <http://doi.org/10.14393/ufu.te.2020.682>. Acesso em: 1 set. 2023.

Recebido em 9 de novembro de 2023

Aprovado em 28 de junho de 2024