

VOBLING – AN INTERSECTION BETWEEN CORPUS AND A MULTIMODAL PLATFORM

Márcio Issamu Yamamoto*
Guilherme Fromm**

- **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.

Introduction

According to Ducrot and Todorov's (1972, p. 12) statement, "the field of Linguistics does not have a unified terminology". Therefore, our proposal is to reflect on the terminology of Linguistics analyzed by a quantitative and qualitative research based on concepts of corpus linguistics (CL). To achieve this goal, we will describe some of the steps taken to the building of a bilingual Vocabulary of Linguistics based on corpus.

* 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>

The study corpus comprises 47 disciplines of Linguistics, organized under two main disciplines: Descriptive and Applied Linguistics.

Thus, in this article, we focus on the following aspects of Terminology practice: (a) corpus compilation, (b) organizing distinctive features for definitions,¹ (c) terminological and encyclopedic definition, and (d) VoBLing as a multimodal platform.

First, the compiled corpus is composed of academic texts encompassing 500 thousand items or tokens in each language for each linguistic discipline.² The same average number of tokens per discipline aimed to meet the principle of corpus balancing and representativeness proposed by corpus linguistics. The quantitative analysis of the corpus was carried out using WordSmith Tools (WST 7.0 and 8.0), a suite of tools which produces word lists and key word lists. WST main tool is Concord, which shows concordance lines (in a KWIC format,³ as we can see in Figure 1) from plain texts and helps users have access to contexts where terms are.

Figure 1 – Biolinguistics – key words in context (KWIC) – partial view

N	Concordance
1	it isthat 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€²24). 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

Source: Authors' elaboration.

Second, after identifying the context from the concordance lines, we proceeded to the organization of distinctive features which followed VoTec's methodology (Fromm,

¹ Here we adopt the expression *distinctive features* by Sager (1990, p. 26). Besides, it can also be named *semantic features* as in Pavel e Nolet (2001, p. 18).

² *Token* is the number of words within a text. If a text is 500 words long, it contains 500 tokens.

³ Key Words In Context: the selected word (or words) to be analyzed appears, in a different color, in the central position of the screen, creating a column of the same word in all lines.

2007; Fromm; Lisboa, 2024), an online terminological database, that allows users to organize distinctive features in terminological card files. Third, the data gathered and organized in VoBLing project files was used to create two types of definitions: (1) a terminological definition and (2) an encyclopedic definition, both types previously selected by the target audience.

Finally, as a research result, we built a **Bilingual Vocabulary of Linguistics** named VoBLing (Yamamoto, 2020) available online.⁴ VoBLing is a multimodal platform that provides users the microstructure elements common to terminological works, including definitions and examples. Furthermore, it provides multisemiotic resources, such as videos and audio for pronouncing entries both in Portuguese and in English.⁵ Additionally, it offers a graphical representation of the term's conceptual structure and its position within the field of Linguistics. It is necessary to mention that this article is part of a PhD thesis, which means procedures contained in the whole research itself are partly described here.

Compilation of the Linguistics Corpus

To start with, it is important to explain that this comparable corpus compilation comprises 29 disciplines of Descriptive Linguistics (DL) and 18 disciplines of Applied Linguistics (AL), totaling 47 disciplines of Linguistics, our main subject field, in Portuguese and in English, composed of academic texts.⁶ Besides that, we also compiled a bilingual comparable corpus of handbooks of Linguistics to have a selection of texts, aimed to students, where definitions could have a chance to appear in simpler contexts. It served its purpose, and in total, we had a corpus of about 2 million tokens altogether. This compilation process was more laborious, because most of the handbooks were only available in hard copies, and we had to undergo a process of optical character recognition (OCR) to obtain electronic versions of them in order to be analyzed by WST. In the end, the files had to be converted into TXT format, so we could achieve better performance using WST.

This compilation of academic texts was carried out by undergraduate and graduate students of English and Portuguese Language and Translation courses over a period of approximately 10 years, from 2010 to 2020, for classroom evaluation purposes (Fromm; Yamamoto, 2021).

As part of the terminological procedure methodology, we focused on naming the various disciplines that collectively form the field of Linguistics. In Figure 2, a

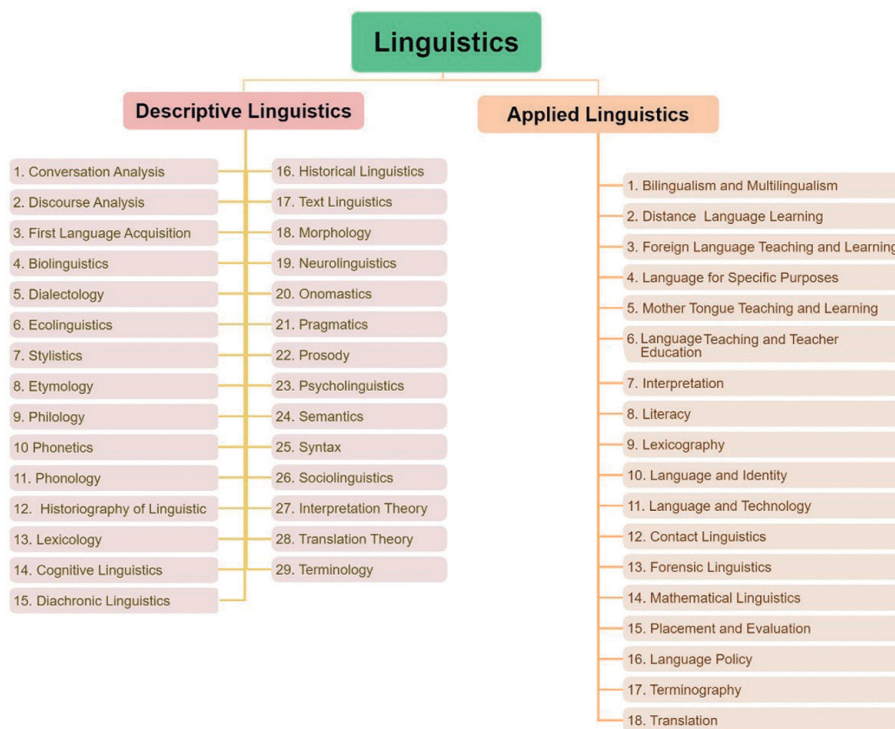
⁴ Available at: <http://vobling.votec.ileel.ufu.br>. Access on: 25 Oct. 2023. VoBLing can be considered a new instance of the VoTec project, according to Fromm and Lisboa (2024).

⁵ The basic tool we use nowadays to write a text like this one, the computer, also shows, through its icons, that the nonverbal language is still in use and recaptures the idea the Egyptian hieroglyphs, for example. Computers allow the interweaving of written and oral texts, including images, tactile elements, space arrangements, and colors.

⁶ PhD Dissertations, MA Thesis, and articles from Linguistics journals.

classification system with the 47 disciplines we studied can be seen, representing our proposition of a Linguistics Taxonomy.⁷ These disciplines were selected and classified based on interviews with specialists and on corpus availability.⁸

Figure 2 – Linguistics classification system



Source: Authors' elaboration

Part of these 47 linguistic disciplines arisen served as a research corpus and were first compiled by Fromm's undergraduate students, then later by graduate students and, at the end, by the researcher that conducted this study to create VoBLing. Altogether, the size of this corpus is 46.4 million tokens, and not 47 million tokens, due to a smaller corpus compiled for Mathematical Linguistics in Portuguese. Since there were not enough articles from this subdiscipline of Linguistics, we could retrieve only 220.245

⁷ We understand that each of these disciplines have a study object, e.g., Phonetics studies the possible sounds in languages, Morphology studies the organization of words in a language and so on.

⁸ In other words, if a discipline exists in the classification system, it means that there are, at least, texts totaling five hundred thousand tokens under it, demonstrating the importance of the discipline. It's relevant, although, to explain that this presented taxonomy is always under change, because new disciplines in Linguistics can suddenly appear from scratches.

items in Portuguese, though we had 507.984 items in English.⁹ Consequently, following the principle of corpus balancing, we reduced the English corpus of Mathematical Linguistics to about 220 thousand items to be on par with the Portuguese one.

Following the compilation process, we proceeded with the cleaning, balancing, and labeling of this corpus. At first, students could retrieve as many tokens as they wanted for a discipline, provided it would have at least 500 thousand items. This number of 500 thousand items originated from Fromm's VoTec research (2007) when it was found that a corpus smaller than that would not provide enough explanatory or definitional contexts from where distinctive features could be extracted to write definitions.¹⁰ In this work, we will call it **standard corpus size**.

That explained, the first step taken towards cleaning the corpus was to eliminate the presentation sections (as the abstract), references, and appendices of each text, leaving only the body of the text itself; The second step involved balancing the corpus, which aimed to standardize the corpus size for each linguistic discipline. This was necessary because some corpora contained more than 1 million items, for example. We also proofread the corpus by using Microsoft Word and its grammar correction tools. It was a very important step since misspelled words or words that were not perfectly processed after OCR affect the word count by WST. Finally, we provided each file with a header, including the title of the original text on top of the file, its website and date of retrieval.

Definition – Organizing Distinctive Features

From the Corpus of Linguistics, described in the previous section, we were able to identify contexts that brought distinctive features of the terminology of Linguistics, identified in concordance lines (WordSmith Tools 7.0 and 8.0; Scott, 2016, 2020). The next step was to organize these features in online terminology card files made available in VoBLing platform.

VoBLing characteristic of organizing distinctive features in record cards is explained by Frame Semantics (Fillmore, 2006) and Frame-based Terminology (Faber Benítez; Marquez Linares; Vega Exposito, 2005) when it comes to using the most frequent semantic features of each term to write its definition.

Fillmore (2006, p. 613, our highlight) defines Frame Semantics as,

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.

⁹ Our experience in compiling information in this discipline revealed that the state of the art in a technical field in one language may not necessarily be equivalent to the state of the art in the same field in another language.

¹⁰ Explanatory contexts provide some ideas, features, characteristics, usage about the term being analyzed. Definatory contexts provide a clear definition of the meaning of the term.

According to Fillmore, **frames** include visual scenes, institutional structures, enactive experiences, human beliefs, actions, experiences, or imaginings. All these scenarios of frames end up being expressed by the speakers' linguistic choices or a group of words and grammar choices they use. These frames are interconnected and activated in the memory by the linguistic material, and they are an essential component for word definitions (Fillmore, 1975, p. 124; 2003, p. 263).

The connection between Frame semantics and Terminology lies in the concept of frames according to Faber Benítez; Marquez Linares; Vega Exposito (2005) when the authors state that

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).

The authors discuss the existence of a system of concepts and one concept which evokes the system itself. When filling out VoBLing card files, it was clear to identify this phenomenon, that is, even though writers were talking about a specific term, e.g.: phonology (see Figure 3), various distinctive features were used for its definition.

Again, Faber Benítez, Marquez Linares and Vega Exposito (2005, p. 4), explain this process by stating that,

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.

VoBLing allows the researchers to copy and paste certain term contexts, previously identified by the Concord tool in WST, from its original files to online record cards, so terms can be analyzed in their original contexts, as shown in Figure 3.¹¹

¹¹ Although the consultant can switch between available languages (English or Portuguese) on the VoBLing query page, the screen layouts in the database are only presented in Portuguese.

Figure 3 – Phonology contexts on 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

Source: Authors' elaboration

When examining the Phonology examples extracted from concordance lines (see Figure 3), it becomes evident that authors employ varying levels of specificity in their definitions of the term as it is shown in column 1 – Exemplo/Example. To cite some examples of various ways to define Phonology, we can consider it as a group of sound rules, or the study of units of sound, or the act of talking about how phonemes function.

In Figure 3 frame, we can also notice three more columns: Conceito (Concept, an attempt of summarizing the provided examples by the researcher, in order to find their core concepts), Fonte (Source) and Ações (Actions, with the possibility of deleting or editing the example). The Concept column comprises pre-summarized distinctive features, created by the researchers, that will aid them creating the final concept and, as a result, formulating the final definition. The Source column designates the original text format of the corpus, which, in this instance, is in PDF format. Lastly, the Actions column allows researchers to either edit or delete the data within the terminological card files.

The next task is to organize the distinctive features presented by the concepts in lines and columns (see Figure 4) — lines are disposed according to the number of examples collected; columns can be added according to the researcher's requirement. Synonymous semes must be allocated in the same columns. Doing so, it shows which concepts are more recurrent, determined by the lexical items employed by authors, converging toward the central concept itself. Semantic features which are less recurrent are constituents of other *frames*, however being part of the same conceptual system.

Figure 4 – VoBLing case file – Phonology

#	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 ...	

Source: Authors' elaboration

In the following section, we will demonstrate how distinctive features were organized to create both a terminological (for the definition) and an encyclopedic definition (for the Note) in the microstructure. We will provide the definition written for Ecolinguistics, as a subdiscipline of Descriptive Linguistics, which falls under the broader domain of Linguistics.

Ecolinguistics – Terminological and Encyclopedic Definition

After organizing the semantic features, the researcher starts the process of writing the definitions based on two distinct patterns. The first pattern is the terminological definition, employing the principles of genus and difference. The second pattern is the encyclopaedic definition. It is worth noting that these two patterns were previously selected based on interviews conducted with potential users, specifically amongst English and Portuguese Language and Literature freshmen at both the Federal University of Uberlândia and the Federal University of Jataí.¹²

Figure 5 – VoBLing – Ecolinguistics definition

INGLÊS

[Go back to search results](#)

ecolinguistics ⓘ

subarea of descriptive linguistics that studies the complex network of relations occurring between environment, languages and people speaking these languages to intervene in the problematics of our living as human beings in a world of diversity

Note: Ecolinguistics is the study of interactions between any given **language** and its environment. It attempts to intervene in the problematics of our living as human beings, to exploit our environment in order to create an extended, sense-saturated ecology that supports our existential trajectories. Ecolinguistics searches to grasp the complexities of **language** raising consciousness about the interdependence between discursive practices and ecological devastation.

Source: Authors' elaboration

¹² Research Ethics Committee approval code CAAE: 80945717.8.0000.5152.

In Figure 5 it is possible to read the highlighted terminological definition written for Ecolinguistics. First, it starts by placing the discipline in a bigger concept system, Descriptive Linguistics, which, in turn, belongs in an even bigger conceptual system, which is Linguistics. This understanding is made possible by the conceptual structure clipping, presented as an image, which will be explained in section 4 of this article. Second, it indicates the object of study for the discipline: *studies the complex network of relations occurring between environment, languages and people speaking these languages*. Finally, it specifies the purpose of the study: *to intervene in the problematics of our living as human beings in a world of diversity*.

In the second part of Figure 5, under **Note**, users can find the encyclopaedic definition, which provides a detailed explanation of the discipline: *study of interactions between any given language and its environment*. It also offers additional details about the relationship between human beings and their linguistic environment and how it is intended to function: *Ecolinguistics searches to grasp the complexities of language raising consciousness about the interdependence between discursive practices and ecological devastation*.

To conclude this section, we reference Faber Benítez, Marquez Linares and Vega Exposito (2005, p. 4) previous quotation which states that “frames can go through several levels of specificity by using hierarchical inheritance”. We apply this exemplified definition because of the connections between the fields of Ecology and Linguistics. In this case, these two main frames become intertwined, giving rise to a third frame, which is the field of Ecolinguistics and its specificity.

VoBLing as a multimodal platform

Before explaining each step taken to build VoBLing (2020) as a multimodal platform, we will briefly discuss the concept of multimodality definition. According to Sabino-Luiz (2023), a multimodal definition refers “to the integration of verbal language with other forms of non-verbal language, both human and non-human, having the potential to show the meaning of the referent in question by using multiple semiotics” (Sabino-Luiz, 2023, p. 1, own translation).¹³ Implementing a multimodal approach can be valuable in guaranteeing a comprehensive grasp of the concepts portrayed in VoBLing. Furthermore, VoBLing, as an electronic vocabulary, brings as advantages its accessibility, multimedia elements such as audio, video, images, and frequent updates. Finally, it’s worth mentioning the significance of custom databases or corpora containing images and audio for multimodal terminology.

Sabino-Luiz (2023, p. 5) explains that in contemporary times, the use of non-verbal language can be a useful and relevant technique for the development of dictionaries,

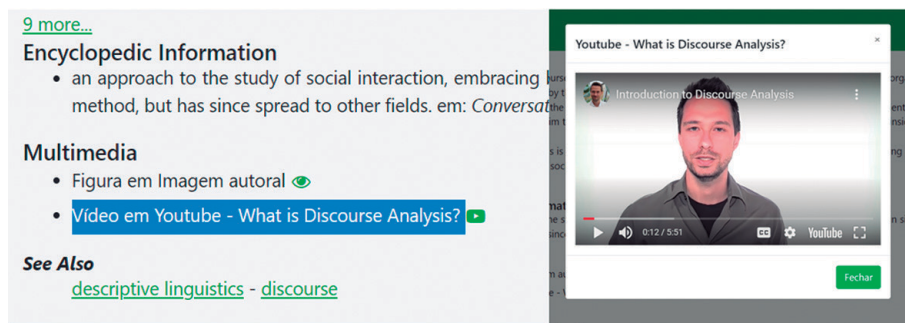
¹³ Original: “O conceito de ‘definição multimodal’ refere-se, por sua vez, à integração da linguagem verbal com outras formas de linguagem não verbal, tanto humanas quanto não-humanas, tendo o potencial de mostrar o significado do referente em questão, utilizando múltiplas semioses”.

provided that it is applied based on theoretical principles rather than merely as an aesthetic choice. Therefore, visual illustrations and also videos have important cognitive and semiotic functions when properly associated with dictionary and vocabulary entries, as they assist the user in understanding the searched lexical or terminological unit. They serve two important cognitive functions in language learning, both complementing and exemplifying the verbal information in the dictionary/vocabulary, showing the user what the defined thing is. However, it is essential that lexicographers and terminographers apply relevant criteria in the selection of multimodal resources in their dictionaries/vocabularies to ensure that these resources contribute to a more comprehensive and accurate understanding of the meanings of lexical units.

Contextualizing the explanation from the previous paragraph to our vocabulary, we can say that, besides the definitions, VoBLing users have access to other learning resources: (1) specialized language videos with English and Portuguese explanation of concepts; (2) examples of language use extracted from corpus; (3) audio files with the entry pronunciation in both languages; (4) clipping of the conceptual structure that shows the term within the field of Linguistics and its disciplines; (5) cross references that pop up when passing the mouse on the hyperlinks, which enables users to read linguistic concepts without having to access other pages; (6) and, by clicking on encyclopedia hyperlinks, users will be provided with more encyclopaedic information on the term being searched. These resources were planned to enhance the comprehension of the entry and offer more modern and responsive features, configuring a multimodal structure.

First, we selected specialized videos from YouTube based on the number of views as a quantitative parameter, which provide explanations about the term being defined in VoBLing. Then, these videos were analyzed and evaluated by the researchers to ensure the reliability and ease of understand for beginner undergraduates. Third, in terms of duration, the videos ranged from five to fifteen minutes. Finally, if these criteria were met, the videos' links were uploaded to VoBLing, allowing users to access them (see Figure 6).

Figure 6 – VoBLing videos – Discourse Analysis



Source: Authors' elaboration

The use of videos as a tool for teaching and learning is highly productive. Videos enable users to get acquainted with the terminology of Linguistics in English, to learn correct pronunciation patterns, and to prepare themselves for reading Linguistics texts in English.

Secondly, VoBLing provides examples of language use based on a corpus (see Figure 7), enabling users to access original excerpts about Linguistics in English. As this corpus primarily consists of texts written in English, it significantly enhances the comprehension and learning of the terminology of Linguistics as well as English as a foreign language. Since most of Brazilian students are beginners in terms of English proficiency, offering them access to these texts would contribute to their language development and deepen their understanding of the English language and literature.

Figure 7 – VoBLing – Applied Linguistics – examples of use

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).

Source: Authors' elaboration

Another advantage of using these corpus examples is that users can explore the historiography of Linguistics within a different context from that of Portuguese speakers, as depicted in Figure 7, example 3. Besides that, direct exposure to texts written in English helps reduce the interference of their mother tongue, as reading these texts contributes to the English language learning process.

Third, concerning the terminology of Linguistics, a significant portion of the vocabulary is derived from Latin, which can be confusing to Portuguese speakers. While the written words are quite similar, the pronunciation may differ. To address this issue, VoBLing provides audio files with English and Portuguese pronunciation, both recorded by native speakers (see Figure 8).¹⁴

¹⁴ The English files were recorded by Fulbright ETAs Gautam Ramesh and Ruben Adery (from Linguistix Pronunciation, <https://www.linguistixpro.com/>). The Portuguese files were recorded by VoBLing first researcher.

Figure 8 – VoBLing entry audio file



Source: Authors' elaboration

As shown in Figure 8, there is an audio icon that users can click to hear the pronunciation of an entry. While this feature is common in many online language dictionaries, however, what sets VoBLing apart is its pronunciation of specific terminology of Linguistics as multiword terms, which is generally not found in standard dictionaries. Multiword terms are often not readily available in well-known online dictionaries. For instance, a search for *linguistic atlas*, yields the following results: (a) In Oxford Learner's Dictionaries,¹⁵ *No exact match found for "linguistic atlas" in English*; (b) On Merriam-Webster.com,¹⁶ a definition was available, but no audio file for pronunciation (see Figure 8); (c) The Cambridge dictionary offers alternate options for *linguistic atlas*,¹⁷ but the term itself is not available on their platform. Their suggestions included: *Search suggestions for linguistic atlas - We have these words with similar spellings or pronunciations: linguistics, linguistic, linguistic science.*

Figure 9 – Merriam Webster search: Linguistic atlas



Source: Merriam Webster

¹⁵ Search available at: <https://www.oxfordlearnersdictionaries.com/spellcheck/english/?q=linguistic+atlas>. Access on: Oct. 26th, 2023.

¹⁶ Search available at: <https://www.merriam-webster.com/dictionary/linguistic%20atlas>. Access on: Oct. 26th, 2023.

¹⁷ Search available at: <https://dictionary.cambridge.org/spellcheck/english-portuguese/?q=linguistic+atlas>. Access on: Oct. 26th, 2023.

Based on the examples provided above, it becomes evident that offering the pronunciation of the multiword terms is valuable. Users often face difficulties locating these pronunciations in standard language dictionaries. At this point, we revisit Sabino-Luiz (2023, p. 16), who emphasizes the benefits of multimodality. The author asserts that, irrespective of the type of electronic dictionary employed, its remarkable potential in multimodality is unquestionable. It can seamlessly combine diverse forms of language, encompassing text, images, sounds, and videos, in order to furnish a more thorough and expansive depiction of the meanings of words and lexical expressions. This openness to other textual modalities allows users to have a more complete and in-depth understanding of word meanings, making the learning process more dynamic and interactive. This versatility means that the electronic dictionary can be updated more easily and frequently, making it more accurate and up-to-date than printed versions.


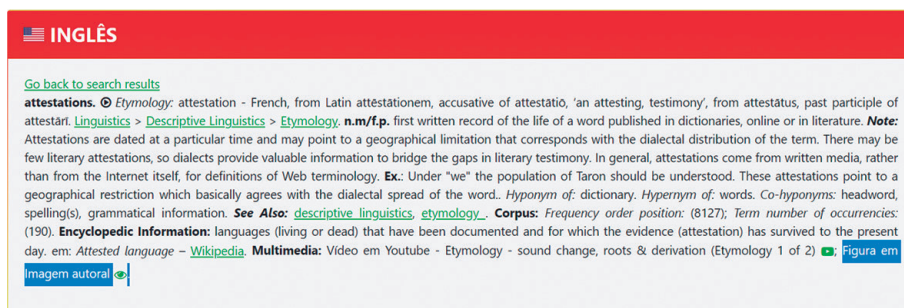
Fourth, VoBLing displays a snippet of the conceptual structure that situates the term within the field of Linguistics and its related disciplines when users click the  icon at the end of the definition.

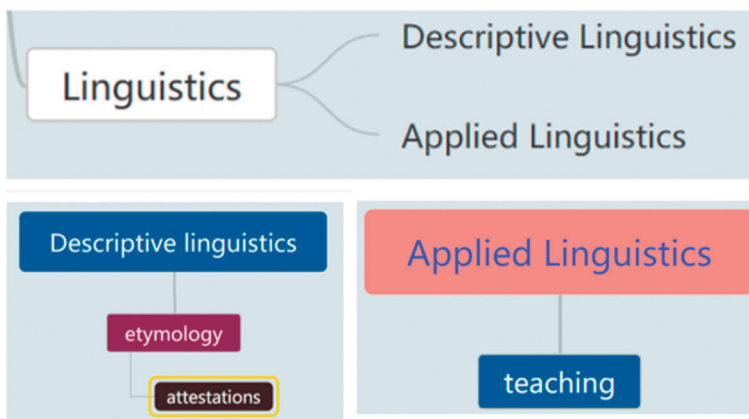
Figure 10 – VoBLing – attestations conceptual structure image



Source: Authors' elaboration

Figure 10 displays the icon for revealing the conceptual structure of *attestation* within the discipline of Etymology. When users click this icon, they can view the image shown in Figure 11.

Figure 11 – VoBLing conceptual structure



Source: Authors' elaboration

The top image illustrates Descriptive and Applied Linguistics as the two primary disciplines within Linguistics. The bottom left image depicts *attestations* as a term linked to Etymology and subordinated to Descriptive Linguistics. The bottom right image displays *teaching* as a term belonging to Applied Linguistics. These images simplify and make clear the learning of concepts and their relationships within the main subject field of Linguistics, making it easy for users to understand.

Fifth, VoBLing provides pop-up cross references. Users can hover their mouse over hyperlinks within the definition or under **Note** to read about linguistic concepts without needing to access other pages.

Figure 12 – VoBLing – descriptive linguistics pop-up cross references

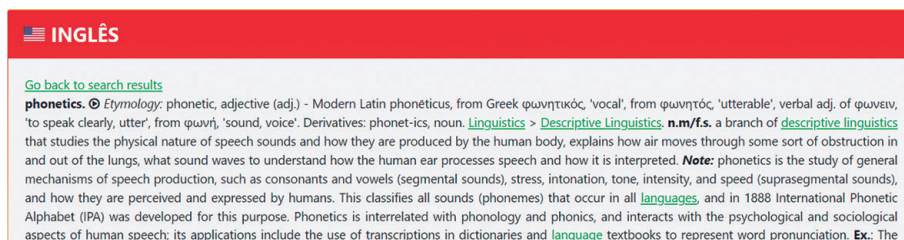
Contraste	
Termo	descriptive linguistics
Definição	subarea of linguistics that describes languages grounded in systematic empirical observation and conceives language as a highly complex symbolic system
Nota	descriptive linguistics is a branch of general linguistics that deals with the fundamental aspects 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

Source: Authors' elaboration

In Figure 12, *descriptive linguistics* is defined based on the semes in *phonetics* definition (as shown in Figure 13). In addition to pop-up cross references, there are also hyperlinks available, enabling users to access other terms that are part of the


microstructure and have already been recorded in the database. Further details will be provided in the explanation under Figure 13.

Figure 13 – VoBLing – phonetics definition and its cross references



INGLÊS

[Go back to search results](#)

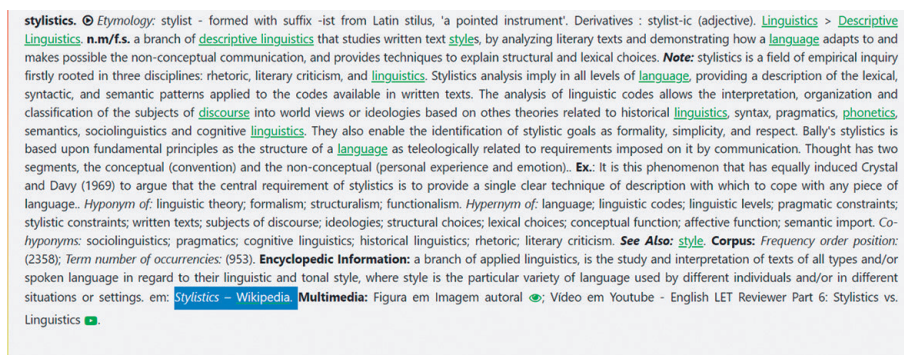
phonetics.  *Etymology:* phonetic, adjective (adj.) - Modern Latin *phoneticus*, from Greek *φωνητικός*, 'vocal', from *φωνήτός*, 'utterable', verbal adj. of *φωνειν*, 'to speak clearly, utter', from *φωνή*, 'sound, voice'. Derivatives: *phonet-ics*, noun. [Linguistics](#) > [Descriptive Linguistics](#). **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




Source: Authors' elaboration

Another feature is depicted in Figure 13: the terms *linguistics* and *descriptive linguistics* appear before the definition of *phonetics*. Users can access the former by clicking on the hyperlinks, while *descriptive linguistics* and *language(s)*, terms within the *phonetics* definition of, will appear as a pop-up window.

Furthermore, there is an encyclopaedic link that connects terms to external sources of information, such as Wikipedia or others (as shown in Figure 14).

Figure 14 – VoBLing – Stylistics link to Wikipedia



stylistics.  *Etymology:* stylist - formed with suffix *-ist* from Latin *stilus*, 'a pointed instrument'. Derivatives : *stylist-ic* (adjective). [Linguistics](#) > [Descriptive Linguistics](#). **n.m/f.s.** a branch of [descriptive linguistics](#) that studies written text [styles](#), by analyzing literary texts and demonstrating how a [language](#) adapts to and makes possible the non-conceptual communication, and provides techniques to explain structural and lexical choices. **Note:** stylistics is a field of empirical inquiry firstly rooted in three disciplines: rhetoric, literary criticism, and [linguistics](#). Stylistics analysis imply in all levels of [language](#), providing a description of the lexical, syntactic, and semantic patterns applied to the codes available in written texts. The analysis of linguistic codes allows the interpretation, organization and classification of the subjects of [discourse](#) into world views or ideologies based on others theories related to historical [linguistics](#), syntax, pragmatics, [phonetics](#), semantics, sociolinguistics and cognitive [linguistics](#). They also enable the identification of stylistic goals as formality, simplicity, and respect. Bally's stylistics is based upon fundamental principles as the structure of a [language](#) as teleologically related to requirements imposed on it by communication. Thought has two segments, the conceptual (convention) and the non-conceptual (personal experience and emotion). **Ex.:** It is this phenomenon that has equally induced Crystal and Davy (1969) to argue that the central requirement of stylistics is to provide a single clear technique of description with which to cope with any piece of language. *Hyponym of:* linguistic theory; formalism; structuralism; functionalism. *Hypernym of:* language; linguistic codes; linguistic levels; pragmatic constraints; stylistic constraints; written texts; subjects of discourse; ideologies; structural choices; lexical choices; conceptual function; affective function; semantic import. *Co-hyponyms:* sociolinguistics; pragmatics; cognitive linguistics; historical linguistics; rhetoric; literary criticism. **See Also:** [style](#). **Corpus:** *Frequency order position:* (2358); *Term number of occurrences:* (953). **Encyclopedic Information:** a branch of applied linguistics, is the study and interpretation of texts of all types and/or spoken language in regard to their linguistic and tonal style, where style is the particular variety of language used by different individuals and/or in different situations or settings. **em:** [Stylistics - Wikipedia](#). **Multimedia:** Figura em Imagem autoral ; Vídeo em Youtube - English LET Reviewer Part 6: Stylistics vs. Linguistics .

Source: Authors' elaboration

In Figure 14, the blue-highlighted link connects to Wikipedia. In addition to the terminological information, images, pop-ups and videos, users also have the option to access external links that connect the term to freely available information on Wikipedia or similar websites. In Figure 14, the first line also displays the etymology of the term, which has been sourced from books and partially reproduced to enrich term definitions.¹⁸

¹⁸ The books used as VoBLing etymology source were *Dicionário Etimológico da Língua Portuguesa* de Nascentes (1955) for *Portuguese* and *Comprehensive Etymological Dictionary of the English Language* by Klein (1971) and

Final comments

After analyzing a substantial corpus of Linguistics, we have observed that the terminology within this field remains fragmented, just as Todorov noted back in the 1970's. However, with the creation of VoBLing, we can propose a potential Linguistics terminology based on frequency, extracted from a representative corpus using corpus linguistics methodology.

By following the steps described in this article, it is evident that this research is feasible, despite its time-consuming nature. The availability of tools like WST and corpus linguistics methodology ensures that results to be of high quality, both quantitatively and in a qualitatively.

The research and the final product demonstrate that the additional components in the entries' microstructures are not only desirable but also necessary for the broader community of scholars and language enthusiasts. Although designed for language students to better understand the terms they read and talk about in their day-by-day at the university, VoBLing is freely available. This means that anyone interested in the field of language, Linguistics and Translation can access it – if the definitions and examples are not enough for a comprehensive understanding of an entry, its multimodal features, which go far beyond the traditional verbal information found in printed dictionaries, vocabularies, and glossaries, can assist those seeking a deeper interpretation of the provided information.

This research and product development pave the way for language professionals to explore new ideas that can be incorporated into future computational tools for lexicographical and terminographical work. This includes not only what is available to the general or specific public of consultants, but also considerations for the database structure. It's important to account for both expected and unexpected users. The continuous interaction of new programming languages, audiovisual features, wiki collaboration, and social medias offers intriguing insights into the future of reference works.

YAMAMOTO, Márcio Issamu; FROMM, Guilherme. VoBLing – uma intersecção entre *corpus* e uma plataforma multimodal. *Alfa*, São Paulo, v. 68, 2024.

- *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*

Origins: A Short Etymological Dictionary of Modern English by Partridge (1966) for English.

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 bilingue 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 bilingue da Linguística; Linguística de Corpus; Terminologia bilingue; Plataforma multimodal.

REFERENCES

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 bilíngue 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.

Received on November 9, 2023

Approved on June 28, 2023