

# A criação do Vocabulário Terminológico de Conceitos de Língua<sup>1</sup>

The creation of the Terminological Vocabulary of Language Concepts

Joel Victor Reis LISBOA\*<sup>ID</sup>

**RESUMO:** Conceitos como *língua materna*, *primeira língua*, *língua não materna*, *língua adicional*, *língua estrangeira*, *segunda língua*, *língua de acolhimento*, *língua de herança*, entre outros, fazem parte da terminologia utilizada nas áreas de Ensino de Línguas e Políticas Linguísticas. Eles são utilizados para especificar relações entre línguas, indivíduos e sociedade e são basilares para a produção e disseminação de conhecimentos especializados produzidos nessas áreas técnico-científicas. Contudo, esse grupo terminológico, que está em constante atualização conforme o desenvolvimento de pesquisas nas referidas áreas, também é fonte de variação terminológica formal e conceitual, nem sempre no contraste entre diferentes trabalhos, mas muitas vezes no escopo de um mesmo texto. Na literatura especializada é possível verificar que uma mesma unidade terminológica possui definições essencialmente distintas e geralmente não intercambiáveis. Também é possível perceber que, muitas vezes, uma mesma definição é dada para diferentes unidades terminológicas. Objetivando apresentar uma visão panorâmica e sistematizada da terminologia dos conceitos de língua, procedemos à criação do Vocabulário Terminológico de Conceitos de Língua – VTCL (Lisboa, 2024 – <http://vtcl.votec.ileel.ufu.br>), fruto de nossa tese de doutorado. Neste artigo, detalhamos os procedimentos realizados para a criação do vocabulário terminológico em questão, construído a partir de um *corpus* especializado de 325 artigos, 521 dissertações e 117 teses da área de Português como Língua Não Materna, totalizando 23.339.211 *tokens* e 168.227 *types*. Apresentamos as motivações da referida pesquisa, os procedimentos de compilação e tratamento do *corpus*, assim como os procedimentos de identificação e análise das unidades terminológicas, planejamento da macro e microestrutura do vocabulário e criação e preenchimento das fichas terminológicas. Por fim, apresentamos o VTCL em conjunto com seus campos, recursos e funcionalidades. A versão inicial do vocabulário focaliza dez unidades terminológicas (*língua adicional*, *língua de herança*, *língua de sinais*, *língua estrangeira*, *língua materna*, *língua não materna*, *língua oficial*, *língua segunda*, *primeira língua* e *segunda língua*) e totaliza 28 verbetes, demonstrando a polissemia presente no grupo terminológico focalizado. O vocabulário segue em desenvolvimento, com o intuito de abranger os 50 conceitos de língua identificados no *corpus* de estudo.

<sup>1</sup> O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Ele é oriundo da tese de doutorado intitulada Vocabulário Terminológico de Conceitos de Língua, defendida em 19 de agosto de 2024 na Universidade Federal de Uberlândia.

\* Doutorado em Estudos Linguísticos (UFU). Professor Assistente do curso de Letras Inglês da Universidade Federal de Jataí (UFJ). Jataí, GO – Brasil. [joelvictorlisboa@gmail.com](mailto:joelvictorlisboa@gmail.com)

**PALAVRAS-CHAVE:** Terminologia. Terminografia. Linguística de *Corpus*. Vocabulário Terminológico. Conceitos de Língua.

**ABSTRACT:** Concepts such as *mother tongue*, *first language*, *non-native language*, *additional language*, *foreign language*, *second language*, *host language*, *heritage language*, among others, form part of the terminology used in the fields of Language Teaching and Language Policy. They are used to specify relationships between languages, individuals, and society and serve as foundational concepts for the production and dissemination of specialized knowledge developed in these technical-scientific fields. However, this terminological group, which is constantly evolving alongside the development of research in these fields, is also a source of formal and conceptual terminological variation. This variation does not always occur in contrasts between different works but often arises within the scope of a single text. In the specialized literature, it is evident that the same terminological unit can have essentially distinct and generally non-interchangeable definitions. It is also common to observe that, in many cases, the same definition is applied to different terminological units. Aiming to provide a comprehensive and systematic overview of the terminology related to language concepts, we developed the Terminological Vocabulary of Language Concepts – VTCL (Lisboa, 2024 – <http://vtcl.votec.ileel.ufu.br>), the result of our doctoral dissertation. In this paper, we detail the procedures carried out for the creation of the VTCL, which was developed from a specialized corpus consisting of 325 papers, 521 master's theses, and 117 doctoral dissertations in the field of Portuguese for Speakers of Other Languages, totaling 23,339,211 tokens and 168,227 types. Furthermore, we outline the research motivations, describe the corpus compilation and processing procedures, and detail the methods used for identifying and analyzing terminological units. We also discuss the design of the vocabulary's macro and microstructure, as well as the creation and completion of terminological records. Finally, we introduce the VTCL, highlighting its fields, resources, and functionalities. Its initial version focuses on ten terminological units (*additional language*, *heritage language*, *sign language*, *foreign language*, *native language*, *non-native language*, *official language*, *first language*, and *second language* (*língua segunda* and *segunda língua*)) and includes 28 entries, demonstrating the polysemy present in the targeted terminological group. The vocabulary is still under development, with the aim of covering the 50 language concepts identified in the study corpus.

**KEYWORDS:** Terminology. Terminography. Corpus Linguistics. Terminological Vocabulary. Language Concepts.

Artigo recebido em: 03.02.2025

Artigo aprovado em: 29.05.2025

## 1 Introdução

A terminologia das áreas que lidam com as interrelações entre línguas, indivíduos e sociedades, como Ensino de Línguas e Políticas Linguísticas, é permeada por conceitos como **língua materna**, **primeira língua**, **língua não materna**, **língua**

**adicional, língua oficial, língua estrangeira, segunda língua, língua de herança, língua de acolhimento**, dentre muitos outros. Esses conceitos são utilizados para classificar, descrever, representar e especificar os papéis desempenhados pelas línguas, as funções sociais atribuídas a elas e as relações entre diferentes línguas em um mesmo espaço sociolinguístico e no repertório linguístico de indivíduos. Em vista disso, essas unidades terminológicas (doravante UT) são basilares para os desenvolvimentos teóricos dessas áreas técnico-científicas, o que pode ser percebido pelo esforço em defini-las e elucidá-las no escopo de diversas pesquisas que compõem a literatura especializada das áreas supramencionadas.

Esses conceitos de língua possuem caráter funcional e dinâmico, estando em constante processo de reformulação, redefinição e atualização em resposta às exigências e aos avanços nas pesquisas, além de estarem intrinsecamente conectados a contextos sociolinguísticos e identitários que, por natureza, estão sujeitos a transformações. Nesse ínterim, novos conceitos surgem, antigos conceitos são reformulados e outros são considerados não funcionais para determinados contextos e recortes de pesquisa.

A ausência de consenso terminológico, a alta variação formal e conceitual e a própria dinamicidade desses conceitos geram instabilidades, sobreposições e ambiguidades terminológicas. Como observado em pesquisas anteriores (Lisboa, 2020; 2021), em diferentes trabalhos, bem como geralmente no escopo de um mesmo texto, diferentes UT recebem uma mesma definição, sendo que, em outros trabalhos, são definidas em relação de co-hiponímia e oposição.

No Quadro 1, a seguir, é possível verificar a situação supramencionada, em que as UT **língua estrangeira** e **língua segunda/segunda língua** são consideradas sinônimas nos dois primeiros trechos apresentados, partilhando de uma mesma base definicional, sendo que são diferenciadas nos últimos dois conforme o contexto sociolinguístico de aquisição de línguas, sendo, portanto, co-hipônimas e não sinônimas.

Quadro 1 – Exemplo de variação conceitual extraído do *corpus* de estudo.

| CÓDIGO DO ARQUIVO <sup>2</sup> | CONTEXTO EXTRAÍDO   |
|--------------------------------|---|
| D_UFAM_2019_Rocilange Cabral   | “Outra questão que merece atenção é que há a seguinte discussão: se não for Língua Materna, todas as outras são Línguas Estrangeiras”.  |
| D_UC_2017_Ana Rodrigues        | “L2 refere-se a qualquer língua aprendida depois da LM”.  |
| A_EXE_2015_Maria Santos        | “A designação língua não materna (LNM) cobre todas as situações em que a língua não é adquirida no contexto familiar, nela se distinguindo duas grandes subdivisões: língua segunda (L2 ou LS) e língua estrangeira (LE)”.      |
| A_DL_2018_Daisy Cordeiro       | “Se o aprendizado da língua alvo ocorrer fora da comunidade nativa desse idioma, diz-se que se trata de aprendizagem de língua estrangeira (LE). Quando ocorre na comunidade, trata-se de aprendizagem de segunda língua (L2)”. |

Fonte: elaborado pelo autor.

Também é comum na literatura especializada das áreas de Ensino de Línguas e Políticas Linguísticas que uma mesma UT seja definida de maneiras distintas e não intercambiáveis. No Quadro 2, ilustramos essa situação. Nele, é possível observar a mesma UT (**língua estrangeira**) sendo definida de três formas diferentes e não intercambiáveis, o que demonstra a polissemia dessa rede terminológica.

Quadro 2 – Exemplo de variação conceitual extraído do *corpus* de estudo.

| CÓDIGO DO ARQUIVO                | CONTEXTO EXTRAÍDO   |
|----------------------------------|---|
| D_UFSCAR_2014_Aline Franceschini | “Nesta pesquisa, consideramos língua estrangeira (LE) aquela ensinada em contextos de imersão, como estrangeiros aprendendo português no Brasil [...]”. |
| T_USP_2017_Leilane Oliveira      | “Língua estrangeira (PLE) abrange o ensino desenvolvido em contexto de não imersão cultural e idiomática”.  |
| T_UAB_2018_Ana Boléo             | “LE será aquela que não tem estatuto político dentro de determinadas fronteiras territoriais”.  |

Fonte: elaborado pelo autor.

Essas questões dificultam a compreensão, o rigor terminológico e o pleno domínio panorâmico dessa rede terminológico-conceitual por parte de pesquisadores

<sup>2</sup> Os códigos dos arquivos que compõem nosso *corpus* de estudo apresentam o tipo textual (A – artigo, D – dissertação, T – tese), a sigla da instituição ou do periódico científico, o ano de publicação e o primeiro e último nome dos autores. Uma listagem completa dos dados dos arquivos pode ser acessada tanto no menu “Ajuda” do VTCL ([http://vtcl.votec.ileel.ufu.br/help/Ajuda\\_VTCL.pdf](http://vtcl.votec.ileel.ufu.br/help/Ajuda_VTCL.pdf)) quanto nos apêndices de nossa tese de doutorado (Lisboa, 2024).

e docentes, como já observado em pesquisas anteriores, como em Barbosa e Freire (2017) e Ferreira (2021).

O cenário apresentado motivou o desenvolvimento de nossa pesquisa de doutorado (Lisboa, 2024), que objetivou a construção de uma obra terminológica que apresentasse de maneira sistematizada e sintetizada o panorama da rede terminológica dos conceitos de língua, em conjunto com suas inter-relações conceituais. O Vocabulário Terminológico de Conceitos de Língua (VTCL) foi desenvolvido segundo os princípios da Teoria Comunicativa da Terminologia, da Linguística de *Corpus* e da Terminografia.

Neste artigo, descrevemos a criação do VTCL, detalhando desde o desenho do *corpus* até o produto finalizado e disponibilizado para acesso on-line e gratuito. Na Seção 2, introduzimos a discussão sobre os conceitos de língua, apresentamos pressupostos teóricos da Teoria Comunicativa da Terminologia e os princípios da Terminografia utilizados, bem como apresentamos a Linguística de *Corpus* e justificamos sua adoção como metodologia nesta pesquisa. Na Seção 3, descrevemos os procedimentos metodológicos de criação do VTCL, desde a compilação e tratamento do *corpus* aos procedimentos de planejamento da macro e microestrutura e análise. O produto finalizado é apresentado na Seção 4, juntamente com uma descrição de seus recursos e funcionalidades. Por fim, na Seção 5, tecemos nossas considerações finais e perspectivas futuras para o VTCL.

## 2 Fundamentos teórico-metodológicos

*Language nominations* (ou **conceitos de língua**, como designado na área de Ensino de Línguas), segundo Aronin e Singleton (2010), refere-se a termos que especificam e classificam as línguas com base nas funções, valores e *status* atribuídos a elas no repertório linguístico de indivíduos e em comunidades sociolinguísticas. Eles são essenciais para pesquisas em Políticas Linguísticas e Ensino de Línguas, orientando desde a fundamentação e recorte teóricos das pesquisas até a formulação

de abordagens educacionais e currículos apropriados a determinados públicos-alvo (Spinassé, 2006; Grosso, 2010; Aronin; Ó Laoire; Singleton, 2011).

Conceitos de língua são frequentemente criados, reformulados ou redefinidos em resposta às demandas das pesquisas, alcançando diferentes níveis de consenso entre os especialistas (Aronin; Singleton, 2010; Grosso, 2010), como é o caso dos conceitos de **língua de herança** e **língua de acolhimento**, que foram cunhados para especificar com mais precisão as funções e *status* das línguas em contextos e repertórios linguísticos específicos, especialmente quando termos genéricos como **língua estrangeira** e **língua adicional** não atendem às particularidades dos contextos pesquisados (Lisboa, 2024).

A multiplicidade e as variadas possibilidades de sobreposição, extensão e substituição de alguns conceitos por outros limítrofes acabam gerando ambiguidades e ruídos na comunicação especializada, resultando no uso impreciso de conceitos de língua em situações em que mesmo mudanças terminológicas sutis podem gerar alterações significativas na conceitualização dessas UT, como pontuam Candian e Bessa (2021).

Essa problemática terminológica já foi constatada em trabalhos anteriores, como em Spinassé (2006), Friedrich e Matsuda (2010), Hammarberg (2010), Jordão (2014), Lisboa (2020; 2021) e Araújo (2021). Em geral, esses trabalhos pontuam que a terminologia dos conceitos de língua é marcada por ausência de rigor terminológico e de consenso entre os especialistas, o que afeta negativamente os desenvolvimentos científicos das áreas situadas na intersecção entre línguas, indivíduos e sociedade. Esses trabalhos também apontam que a elucidação, desambiguação e definição desses conceitos de língua é uma tarefa altamente complexa, haja vista que eles estão associados a questões linguísticas e extralingüísticas, além da instabilidade terminológica já mencionada.

O domínio da rede terminológico-conceitual dos conceitos de língua é essencial para profissionais e pesquisadores nas áreas de Ensino de Línguas e Políticas Linguísticas. No entanto, estudos de Barbosa e Freire (2017) e Ferreira (2021) mostram

que muitos professores têm dificuldades em diferenciar e aplicar esses conceitos devido à falta de clareza conceitual, o que compromete a adaptação do ensino às necessidades sociolinguísticas de diferentes públicos. Diante disso, uma obra terminográfica focada nos conceitos de língua, voltada para pesquisadores e professores (para os quais muitos desses termos podem ser novos e suas inter-relações conceituais podem não ser evidentes), consiste em um produto de pertinência social e acadêmica.

Para a Teoria Comunicativa da Terminologia, esses conceitos de línguas constituem-se como UT. As UT são elementos linguísticos sistemáticos que adquirem estatuto terminológico na comunicação especializada, podendo ser constituídas por uma única unidade lexical ou por agrupamentos sintagmáticos, desde que atuem como unidade conceitual e estejam associadas a nódulos dentro da estrutura conceitual de um domínio de especialidade (Rey, 1995; Cabré, 1999). O foco desta pesquisa são UT compostas por múltiplas unidades lexicais que, apesar dessa composição, cada uma opera funcionalmente como uma única unidade de significação especializada.

Com função linguística, cognitiva e representativa, as UT cumprem papel fundamental na estruturação, organização e comunicação de conhecimentos especializados, interligando-se sistematicamente a outras UT dentro de uma mesma rede terminológica e formando a estrutura conceitual do domínio especializado ao qual pertencem (Cabré, 2009). Assim, toda UT pode ser definida e situada dentro de um conjunto articulado de relações conceituais. Além disso, as UT não existem isoladamente da língua geral. Por serem parte de línguas naturais, elas estão sujeitas a variações formais e conceituais, bem como a fenômenos como a polissemia (Cabré, 2000), como comprovado nesta pesquisa.

A Teoria Comunicativa da Terminologia foi escolhida como base teórica desta pesquisa devido à sua abordagem altamente descritiva e à consideração da variação e da polissemia na comunicação especializada, características centrais da terminologia

dos conceitos de língua. Além disso, seu foco no estudo das terminologias em efetivo uso na comunicação especializada (*in vivo*) alinha-se à Linguística de *Corpus*, metodologia adotada na construção do VTCL.

Nossa pesquisa visa à construção de uma obra de referência entendida por Barbosa (1995, 2001) como vocabulário. Esse tipo de obra reúne e organiza termos representativos de domínios especializados, servindo como referência para a comunicação técnico-científica. Essas obras adotam como unidades básicas os vocábulos e as UT, apresentando caráter sincrônico (recorte temporal), contemporâneo e sinfásico (comunicação especializada em contexto determinado) (Barbosa, 1995; 2001).

A produção desse tipo de obra se dá no âmbito da Terminografia, área situada no escopo das comunicações especializadas, intimamente vinculada à Terminologia, dedicada à construção e avaliação de métodos e práticas de compilação, descrição, gerenciamento e apresentação de terminologias em obras de referência (Krieger; Finatto, 2004; Almeida; Correia, 2008).

Conforme pontuam Aubert (2001), Barros (2004) e Kilian e Reuillard (2023), a elaboração de uma obra terminográfica monolíngue inicia-se com a seleção do domínio de especialidade, a identificação da problemática terminológica, a determinação do público-alvo e da finalidade da obra, além da escolha de uma base teórica e metodológica, como apresentamos nos parágrafos e seções anteriores em relação ao VTCL. Essas decisões levam ao estabelecimento do tipo de obra de referência a ser criada – no caso desta pesquisa, um vocabulário terminológico – bem como sua macroestrutura, microestrutura e padrões definicionais.

Quanto à macroestrutura, é necessário decidir sobre a inclusão de textos externos (*outside matter*), a extensão da nomenclatura, a organização e lematização das entradas, o tratamento de polissemia e a estruturação do(s) sistema(s) de remissivas (Barros, 2004; Welker, 2004).

Os textos externos (*outside matter*) orientam os usuários sobre o uso da obra de referência e fornecem informações sobre sua elaboração, incluindo, por exemplo, introdução, objetivos, público-alvo e manual de utilização (Welker, 2004). A nomenclatura refere-se à listagem do conjunto de UT que compõem a obra e que são definidas nos verbetes (Barros, 2004). A lematização, segundo Barros (2004), é o processo de determinação da forma-base das UT apresentadas na nomenclatura, que deve ser padronizada. Já o sistema de remissivas é utilizado para estabelecer conexões entre UT que compõem a obra de modo a enriquecer a base informacional dos verbetes por meio do resgate de relações semântico-conceituais, vinculação de variações formais e conceituais e evidenciação de campos semânticos, evitando o isolamento das informações dos verbetes (Barros, 2004). As decisões concernentes aos elementos aqui mencionados estão descritas na Subseção 3.2 referentes ao planejamento estrutural do VTCL.

A microestrutura do VTCL segue a herança terminográfica do VoTec (Fromm, 2007; Fromm; Lisboa, 2024), com algumas modificações, e é organizada da seguinte forma:

Quadro 3 – Microestrutura do VTCL.

|  |
|--|
| Entrada + informações gramaticais + definição aristotélica + notas enclopédicas + abreviaturas/acrônimos/siglas + frequência no <i>corpus</i> + relações conceituais (hiponímia, hiperonímia, co-hiponímia, parassinonímia e antónima) + informações enclopédicas externas (Wikipedia) + multimídia (YouTube) + remissivas para outras acepções + abonações. |
|--|

Fonte: elaborado pelo autor.

Como afirmam Barros (2004) e Welker (2004), a definição aristotélica é considerada ideal na elaboração de produtos terminográficos e consiste em um enunciado definicional iniciado por um **gênero próximo** (um termo mais genérico, geralmente o hiperônimo) e seguido por **diferenças específicas** (traços conceituais que permitem distinguir a UT focalizada de outras relacionadas). No entanto, Rey (1995) aponta que, em certos domínios especializados, a inclusão de informações

enciclopédicas pode ser produtiva para ampliar a disponibilidade de informações e reduzir a probabilidade de ambiguidades terminológicas.

As pesquisas de Cardoso (2017) e Yamamoto (2020) com públicos acadêmicos indicam que as definições aristotélica e enciclopédica foram consideradas as mais úteis pelos informantes. Em vista disso, em conformidade com os vocabulários produzidos nas referidas pesquisas, optou-se por seguir como padrão a definição aristotélica e complementá-la com informações enciclopédicas e terminológicas no campo “Notas” de cada verbete, de modo a enriquecer a microestrutura com dados adicionais sobre as UT e seu emprego no discurso especializado.

A maioria das ocorrências dos conceitos de língua no *corpus* aparece como abreviaturas, acrônimos e siglas. Portanto, foi essencial incluir essas informações nos verbetes para que os usuários conheçam formas diversas utilizadas no lugar das UT. Os dados de frequência são importantes para permitir que o consultante perceba, ao comparar diferentes verbetes, quais UT são mais comuns no *corpus* e quais têm menor frequência.

As relações conceituais também são parte essencial nos verbetes do VTCL, pois demonstram como as UT focalizadas em cada verbete estão sistematicamente vinculadas a outras UT pertencentes à mesma rede terminológico-conceitual. Portanto, foi substancial inserir as relações de hiperonímia, hiponímia, co-hiponímia, parassinonímia e antonímia. A hiperonímia refere-se à relação entre um conceito mais abrangente e superordenado e um mais específico e subordinado a ele, como **fruta** (hiperônimo) e **pera** (hipônimo) ou **animal** (hiperônimo) e **gato** (hipônimo) (Hoffmann, 2014). Co-hiponímia é a relação entre conceitos vinculados a um mesmo hiperônimo, como **carro**, **ônibus** e **caminhão**, co-hipônimos vinculados ao hiperônimo **veículo** (Welker, 2004). Parassinônimos são palavras com significados semelhantes, mas com distribuições contextuais diferentes, isto é, não são completamente intercambiáveis, mesmo tendo definições semelhantes, como, por exemplo, **falar** e **dizer** (Galisson, 1979 *apud* Barros, 2004). A antonímia, por sua vez, refere-se à relação

de oposição entre palavras, na qual suas próprias definições geralmente passam pelo contraste entre elas, como, por exemplo, **vivo** e **morto** ou **limpo** e **sujo** (Hartmann; James, 1998).

Assim como Fromm (2007), optamos por utilizar dados da Wikipedia para direcionar os usuários a fontes públicas e gratuitas com definições e informações adicionais sobre cada UT, de modo a ampliar as informações sobre cada UT do VTCL. Na esteira do trabalho de Yamamoto (2020), também foram inseridos no campo “Multimídia”, sempre que possível, *links* de vídeos do YouTube de até dez minutos que definissem as UT enfocadas em cada verbete, de modo a aumentar ainda mais o poder informacional do VTCL.

Por fim, como afirma Welker (2004), as abonações são trechos autênticos do *corpus*, utilizados para validar a UT, mostrar seu uso no discurso especializado e auxiliar na compreensão pelo consultente. Para as abonações, priorizamos excertos que funcionam como contextos definitórios e explicativos, visando enriquecer a informação sobre cada UT nos verbetes. Segundo Aubert (2001), contextos explicativos fornecem alguns traços conceituais relacionados à UT, enquanto os contextos definitórios oferecem uma definição completa, com todos os traços conceituais essenciais, sendo menos frequentes que os demais tipos de contextos.

Todo o processo de análise na construção do VTCL foi feito à luz da Linguística de *Corpus*, metodologia e abordagem de base empírica e estatística dedicada à análise linguística por meio de *corpora* processáveis por computador e de recursos analíticos de base estatística (Berber Sardinha, 2004), como o WordSmith Tools 8.0 (Scott, 2022), utilizado nas análises desta pesquisa.

Os *corpora* são conjuntos textuais criteriosamente planejados, compilados, tratados computacionalmente e organizados de modo a representar uma amostragem da língua ou do recorte da língua selecionado para análise (Aluísio; Almeida, 2006). No caso das pesquisas em Terminologia, como a nossa, utiliza-se *corpora* especializados, que focalizam gêneros, registros e temáticas específicos (Koester, 2010;

Koteyko, 2014). O *corpus* desta pesquisa é pertencente a essa última categoria, sendo composto por artigos, dissertações e teses da área de Português como Língua Não Materna (doravante PLNM).

A Linguística de *Corpus* e o WordSmith Tools (doravante WST) foram adotados nesta pesquisa devido à possibilidade de análise de grandes volumes de texto com precisão quantitativa e por tornar mais fácil a organização dos dados da pesquisa, permitindo sua reorganização conforme os objetivos e necessidades de cada etapa metodológica. Além disso, a celeridade na identificação de UT e quantificação precisa de frequências, o acesso facilitado a contextos definitórios e explicativos e a promoção de objetividade, empirismo e replicabilidade também justificam a seleção da metodologia e, por conseguinte, do programa de análise lexical supracitados.

Apresentados os fundamentos teórico-metodológicos desta pesquisa relacionados aos conceitos de língua, à Teoria Comunicativa da Terminologia, à Terminografia e à Linguística de *Corpus*, passemos à descrição dos procedimentos metodológicos envolvidos na criação do VTCL.

### **3 A criação do Vocabulário Terminológico de Conceitos de Língua**

Os procedimentos metodológicos de criação do VTCL podem ser divididos em três grupos: (i) compilação e tratamento do *corpus* de estudo; (ii) planejamento do vocabulário; (iii) procedimentos de análise. Cada grupo de procedimentos está descrito nas subseções seguintes.

#### **3.1 Compilação e tratamento do corpus de estudo**

O *corpus* utilizado para a construção do VTCL apresenta as seguintes características:

Quadro 4 – Descrição do *corpus* do VTCL.

| CATEGORIAS                          | ESPECIFICAÇÕES   |
|-------------------------------------|--|
| Língua:                             | Monolíngue (português)   |
| Modo:                               | Escrito (digital/digitalizado)   |
| Tempo:                              | Sincrônico/Contemporâneo (2001-2020)   |
| Seleção:                            | Amostragem/Estático  |
| Conteúdo:                           | Especializado (PLNM e áreas correlatas)  |
| Autoria:                            | Diversos autores (nativos e não nativos)   |
| Finalidade:                         | Estudo   |
| Balanceamento:                      | Não balanceado   |
| Tamanho:                            | Médio (23.339.211 <i>tokens</i> <sup>3</sup> e 168.227 <i>types</i> <sup>4</sup> ) |
| Gênero:                             | Acadêmico  |
| Tipologias textuais e quantitativo: | 325 artigos, 521 dissertações e 117 teses  |
| Amostras textuais:                  | Textos completos   |

Fonte: elaborado pelo autor.

Como é possível verificar no Quadro 4, ele é um *corpus* especializado, monolíngue, sincrônico, contemporâneo e compilado como uma amostragem estática. Ele é composto por textos do gênero acadêmico, especificamente 325 artigos, 521 dissertações e 117 teses da área de PLNM e de áreas correlatas. Os textos são nativos digitais ou digitalizados, escritos em português, possuem autoria diversa (incluindo autores lusófonos e não lusófonos) e foram publicados entre 2001 e 2020. O *corpus* tem a finalidade de estudo, não é balanceado, possui tamanho médio (23.339.211 *tokens* e 168.227 *types*) e as amostras textuais são os textos completos, com exceção de elementos excluídos na etapa de limpeza.

O *corpus* foi compilado à época de nossa pesquisa de mestrado (Lisboa, 2021) de 11 diferentes fontes<sup>5</sup>. Como critérios de inclusão, os textos deveriam ser vinculados à área de PLNM, estar acessíveis gratuitamente *on-line* e conter conceitos de língua em seu corpo textual<sup>6</sup>. Os artigos incluídos no *corpus* são provenientes de 114 periódicos

<sup>3</sup> Total de itens, considerando-se as repetições.

<sup>4</sup> Total de itens, desconsiderando-se as repetições.

<sup>5</sup> Especificamente: (i) Acervo Celpe-Bras; (ii) Biblioteca Digital Brasileira de Teses e Dissertações; (iii) Portal de Periódicos CAPES; (iv) Repositório Científico de Acesso Aberto de Portugal; (v) Revista SIPLE; (vi) SciELO; (vii) *American Organization of Teachers of Portuguese*; (viii) Associação de Professores de Português; (ix) Grupo de Estudos e Pesquisas em Linguística Aplicada; (x) site Foca no Português para Estrangeiros; (xi) site Português para Estrangeiros.

<sup>6</sup> Em cada texto acessado, realizava-se a busca por **língua**, seguida de uma análise para confirmar a presença de menções a conceitos de língua no corpo textual.

científicos de 10 países diferentes. As dissertações e teses são oriundas de 66 instituições de ensino superior distribuídas entre Brasil, Portugal e Espanha.

Conforme os arquivos eram selecionados, seus dados de autoria, local de publicação, ano e título eram registrados em fichas de controle específicas por tipo textual, destinadas a facilitar buscas e evitar duplicações. Simultaneamente, os arquivos recebiam códigos de nomeação que informavam tipo textual, local de publicação, ano e autor, permitindo sua rápida identificação. A Figura 1 ilustra parcialmente as fichas de controle das dissertações e teses. Também é possível observar os códigos atribuídos aos arquivos.

Figura 1 – Fichas de controle de dissertações e teses (visão parcial).

| DISSERTAÇÕES |                                    |   |      |                                |  |
|--------------|------------------------------------|---|------|--------------------------------|--|
| Nº           | Código                             | Instituição   | Ano  | Autoria                        | Título   |
| 1            | D_CEFETMG_2012_Liliane Damazo      | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2012 | Liliane Oliveira Damazo        | A modalização na produção de textos em português como língua estrangeira   |
| 2            | D_CEFETMG_2014_Mariza Lacerda      | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2014 | Mariza Gabriela de Lacerda     | O fazer textual de candidatos ao exame para obtenção de certificação em língua portuguesa para estrangeiros (Celp-Bras): olhares sobre os processos argumentativos |
| 3            | D_CEFETMG_2015_Rafaela Coelho      | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2015 | Rafaela Pascoal Coelho         | Diferentes olhares sobre a formação de professores de Português como Língua Adicional no estado de Minas Gerais  |
| 4            | D_CEFETMG_2017_Mahulikplimi Agossa | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2017 | Mahulikplimi Obed Brice Agossa | O exame Celp-Bras como instrumento de divulgação da cultura brasileira: percepções de candidatos   |
| 5            | D_CEFETMG_2017_Rose Silva          | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2017 | Rose Mara Silva                | As novas tecnologias de informação e comunicação na avaliação de proficiência em português como língua estrangeira: o exame Celp-Bras                              |

| TESES |                                  |   |      |   |  |
|-------|----------------------------------|---|------|---|--|
| Nº    | Código                           | Instituição   | Ano  | Autoria                                   | Título   |
| 1     | T_CEFETMG_2018_Laura Ferreira    | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2018 | Laura Márcia Luiza Ferreira               | Avaliação da proficiência oral: uma análise fatorial e de discriminação dos itens do exame Celp-Bras                                   |
| 2     | T_CEFETMG_2018_Liliane Neves     | Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) | 2018 | Liliane de Oliveira Neves                 | Confiabilidade e comportamento avaliativo na prova oral do exame Celp-Bras: um estudo longitudinal                                     |
| 3     | T_PUCRJ_2003_Adriana Albuquerque | Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)      | 2003 | Adriana Ferreira de Sousa de Albuquerque  | A construção dos atos de negar em entrevistas televisivas: uma abordagem interdisciplinar do fenômeno em PLM com aplicabilidade em PLE |
| 4     | T_UFBA_2015_Daniela Benedini     | Universidade Federal da Bahia (UFBA)                              | 2015 | Daniela Ressurreição Mascarenhas Benedini | O português como herança na Itália: línguas e identidades em diálogo   |
| 5     | T_UFC_2013_Vicente Martins       | Universidade Federal do Ceará (UFC)                               | 2013 | Vicente de Paula da Silva Martins         | Estratégias de compreensão de expressões idiomáticas por não nativos do português brasileiro   |

Fonte: Lisboa (2024, p. 108).

Feita a criação dos códigos e a inserção dos dados na listagem de controle, os arquivos eram salvos em pastas específicas para cada gênero textual e renomeados com os códigos de nomeação criados.

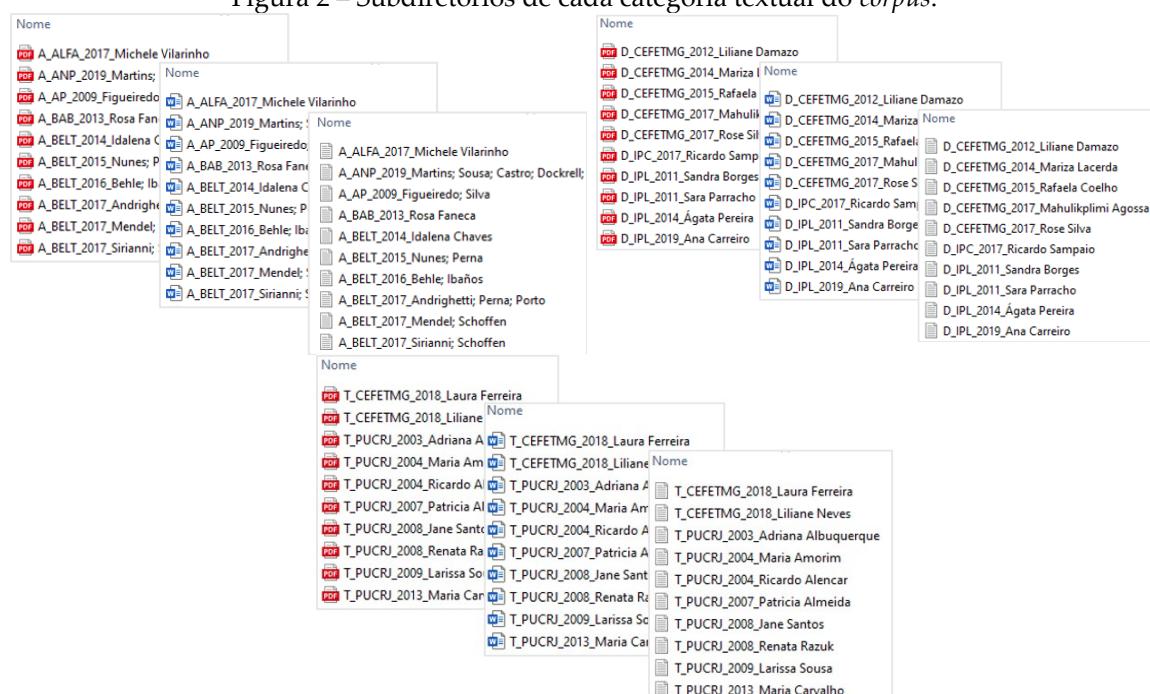
As etapas seguintes envolveram a conversão de arquivos .pdf para .docx para facilitar o procedimento de limpeza dos arquivos. Para isso, utilizamos a versão

*premium* do software iLovePDF<sup>7</sup>. Após cada conversão, os arquivos eram verificados e corrigidos quanto a erros como junção indevida de palavras e problemas de hifenização, espaçamento e decodificação de caracteres.

Após a conversão e correção dos erros, realizamos a limpeza dos arquivos no Microsoft Word, removendo dados não essenciais para a pesquisa, como elementos pré-textuais (capas, contracapas, sumários, listas de figuras, tabelas, quadros e agradecimentos), pós-textuais (referências, apêndices, anexos), certos elementos textuais (quadros, figuras, tabelas, epígrafes e notas de rodapé, com exceção das que continham conceitos de língua) e extratextuais (biodata, cabeçalhos e rodapés).

Após a limpeza, os arquivos .docx eram salvos em diretórios específicos, com o código de nomeação do arquivo original em .pdf. Em seguida, eram convertidos para .txt por meio de cópia e colagem no Bloco de Notas do Windows, mantendo o código. Ao final, contávamos com três pastas, uma para cada tipo textual, com subpastas para os formatos .pdf, .docx e .txt. A Figura 2 ilustra o conteúdo parcial das subpastas de cada categoria textual.

Figura 2 – Subdiretórios de cada categoria textual do *corpus*.



Fonte: Lisboa (2024, p. 113).

<sup>7</sup> Disponível em: <https://www.ilovepdf.com>.

Os arquivos em .pdf e .docx foram mantidos apenas para fins de armazenamento, pois os arquivos .txt são os que foram utilizados para análise. O último passo desse grupo de procedimentos foi a conversão em massa da codificação dos arquivos .txt, originalmente em UTF-8, para Unicode (extensão com menor incidência de erros no processamento pelo WST). A conversão foi realizada por meio do utilitário Text Converter do WST. Feito isso, o *corpus* estava pronto para processamento e análise.

### 3.2 Planejamento estrutural do VTCL

O VTCL foi planejado à luz dos princípios da Teoria Comunicativa da Terminologia e da Terminografia, conforme apresentado na seção de fundamentação teórica, bem como considerando as possibilidades oferecidas pelo banco de dados e página de consulta do VoTec. Vale mencionar que em Fromm e Lisboa (2024), realizamos uma análise do estado da arte sobre o VoTec e seus projetos derivados, como o TermosTeo (Cardoso, 2017)<sup>8</sup> e o VoBLing (Yamamoto, 2020)<sup>9</sup>, para que fosse possível mapear os recursos implementados em projetos anteriores que poderiam ser reaproveitados, além de identificar possibilidades para criação de novos recursos e ajustes e melhorias dos já existentes.

Em relação à macroestrutura do VTCL, foram tomadas as seguintes decisões:

- (i) presença de texto externo (*outside matter*): estabelecemos que o texto externo seria disponibilizado no menu “Ajuda” da página de consulta e contaria com os seguintes dados: introdução à obra; explicação sobre o que são os conceitos de língua; objetivos do VTCL; público-alvo; descrição do *corpus*; seção de clarificação terminológica (destinado ao esclarecimento de siglas e termos presentes nos verbetes, como *s.f.s*, *hiperônimo*, *co-hipônimo* etc.); explicações ilustradas sobre os tipos de exibição (normal e descritivo) e de consulta (total

---

<sup>8</sup> Vocabulário disponível em: <http://teo.votec.ileel.ufu.br/>.

<sup>9</sup> Vocabulário disponível em: <http://vobling.votec.ileel.ufu.br/>.

e modular); manual ilustrado e detalhado de uso, descrevendo recursos, campos e funcionalidades da consulta; detalhamento da microestrutura dos verbetes; lista de dados dos documentos do *corpus*, com vistas a facilitar que o consultante busque por documentos específicos por meio dos códigos apresentados nas abonações veiculadas nos verbetes.

- (ii) abrangência, extensão, lematização e arranjo da nomenclatura: decidimos incluir na versão inicial do VTCL os resultados das análises das 10 UT designativas de conceitos de língua mais frequentes no *corpus* de estudo. Tendo em vista a polissemia identificada, essas análises resultaram em 28 entradas no vocabulário. A nomenclatura é organizada em ordem alfabética, com a utilização de índices numéricos sobrescritos para diferenciar as acepções de uma mesma UT. As UT são veiculadas com iniciais maiúsculas e no singular.
- (iii) tratamento da polissemia: optamos por separar as acepções de uma mesma UT em entradas distintas para melhor representar as relações conceituais de hiperonímia, hiponímia, co-hiponímia, parassinonímia e antonímia. Como cada acepção corresponde a um conceito específico, suas relações com conceitos limítrofes variam, justificando essa decisão.
- (iv) ordem de apresentação das acepções: ao identificar casos de polissemia durante a extração de contextos explicativos e definitórios, definimos que contabilizariamos a frequência de cada acepção associada à UT. Na nomenclatura, como mencionado anteriormente, as acepções seriam organizadas em ordem decrescente de frequência, com a mais frequente marcada pelo número <sup>1</sup> sobrescrito, a segunda pelo número <sup>2</sup> sobrescrito, e assim sucessivamente.
- (v) sistemas de remissivas baseados em *links* e *pop-ups*: projetamos dois sistemas de remissivas para o VTCL. O primeiro, intitulado “Compare”, focaliza a polissemia, direcionando o usuário para outras acepções de uma mesma UT

por meio de *links*. Por exemplo, a UT **Língua Estrangeira** possui três acepções e, consequentemente, três verbetes (**Língua Estrangeira<sup>1</sup>**, **Língua Estrangeira<sup>2</sup>** e **Língua Estrangeira<sup>3</sup>**). Cada verbete contém remissões para outras acepções da mesma UT, permitindo que o usuário acesse o verbete correspondente ao clicar nas UT inseridas nesse campo. Além disso, essas remissivas funcionam como *pop-ups*: ao posicionar o cursor sobre uma UT no campo “Compare”, surge um *pop-up* com sua definição; já ao passar o cursor sobre o próprio campo “Compare”, abre-se um *pop-up* contrastivo exibindo as definições das demais acepções em conjunto, de modo a facilitar a comparação e visualização do panorama polissêmico de cada UT. O segundo sistema de remissivas, também projetado com o uso de *links* e *pop-ups*, focaliza as relações conceituais entre as UT. Ao passar o cursor sobre UT presentes nos campos de hiperonímia, hiponímia, co-hiponímia, parassinonímia e antonímia, surge um *pop-up* com sua definição. Ao clicar em uma UT nesses campos, o usuário é direcionado para seu verbete correspondente, já que todas funcionam como *links*. Para que esses recursos pudessem ser implementados, apenas UT que possuíam verbetes dentro do VTCL poderiam ser inseridas nos campos supramencionados.

Em relação à microestrutura, ela foi organizada da seguinte forma: (i) UT por extenso; (ii) informações gramaticais (categoria, gênero e número); (iii) definição aristotélica; (iv) notas (informações enclopédicas suplementares ou sobre utilização terminológica); (v) abreviaturas/acrônimos/siglas; (vi) frequência no *corpus*; (vii) hipônimos; (viii) hiperônimos; (ix) co-hipônimos; (x) parassinônimos; (xi) antônimos; (xii) informações enclopédicas (extraídas da Wikipedia, de modo que os consultentes possam contrastar as informações do VTCL com outra fonte de acesso público gratuito); (xiii) multimídia (vídeo de até dez minutos oriundo do YouTube que complementa a definição apresentada); (xiv) compare (todas as outras acepções da UT

do verbete); (xv) contextos definitórios e explicativos extraídos do *corpus* junto com os códigos dos arquivos, de modo a ampliar o poder informacional do verbete.

Para tanto, as fichas terminológicas foram criadas com campos específicos para cada uma das informações a serem exibidas nos verbetes. A Figura 3, a seguir, apresenta a estrutura seguida pelas 28 fichas terminológicas do VTCL.

Figura 3 – Modelo de ficha terminológica do VTCL.

| <b>Entrada:</b>                   |                                  |                   |                |                |  |  |
|-----------------------------------|----------------------------------|-------------------|----------------|----------------|--|--|
| <b>AAS:</b>                       | <b>Frequência:</b>               | <b>Categoria:</b> | <b>Número:</b> | <b>Gênero:</b> |  |  |
| <b>Definição:</b>                 |                                  |                   |                |                |  |  |
| <b>Notas:</b>                     |                                  |                   |                |                |  |  |
| <br>                              |                                  |                   |                |                |  |  |
| <b>Hiperônimo de:</b>             |                                  |                   |                |                |  |  |
| <b>Hipônimo de:</b>               |                                  |                   |                |                |  |  |
| <b>Cohipônimo de:</b>             |                                  |                   |                |                |  |  |
| <b>Parassinônimos:</b>            |                                  |                   |                |                |  |  |
| <b>Antônimos:</b>                 |                                  |                   |                |                |  |  |
| <b>Remissivos (compare):</b>      |                                  |                   |                |                |  |  |
| <br>                              |                                  |                   |                |                |  |  |
| <b>Informações Enciclopédicas</b> |                                  |                   |                |                |  |  |
| <b>Artigo:</b>                    |                                  | <b>Fonte:</b>     | Wikipedia      |                |  |  |
| <b>Link:</b>                      |                                  | <b>Definição:</b> |                |                |  |  |
| <br>                              |                                  |                   |                |                |  |  |
| <b>Multimídia</b>                 |                                  |                   |                |                |  |  |
| <b>Youtube:</b>                   |                                  |                   |                |                |  |  |
| <br>                              |                                  |                   |                |                |  |  |
| <b>Contextos</b>                  |                                  |                   |                |                |  |  |
| <b>Código</b>                     | Contexto definitório/explcativo. |                   |                |                |  |  |
| <b>Código</b>                     | Contexto definitório/explcativo. |                   |                |                |  |  |
| <b>Código</b>                     | Contexto definitório/explcativo. |                   |                |                |  |  |
| <b>Código</b>                     | Contexto definitório/explcativo. |                   |                |                |  |  |

Fonte: elaborada pelo autor.

Apesar da microestrutura estabelecida, sabemos que nem sempre é possível obter a totalidade das informações elencadas a partir do *corpus*, bem como não há como assegurar que haja conteúdos externos da Wikipedia e YouTube suficientes e adequados para inserção na microestrutura. Portanto, embora todos os verbetes sigam a mesma estrutura, ligada ao banco de dados do VTCL, os campos não preenchidos nas fichas terminológicas e consequentemente no banco de dados do VTCL não são exibidos aos usuários.

Finalizado o planejamento da macro e microestrutura do VTCL, partimos para os procedimentos de análise, descritos na subseção seguinte.

### 3.3 Procedimentos de análise

O primeiro procedimento de análise foi feito por meio da ferramenta Concord do WST. Utilizamos o recurso Clusters para gerar uma lista de *clusters* (agrupamentos recorrentes de palavras no *corpus*) a partir do termo **língua**, com o objetivo de identificar candidatos a UT designativas de conceitos de língua. As configurações adotadas foram: (i) agrupamentos de 2 a 5 palavras, considerando que as UT-alvo raramente excedem essa extensão, conforme estudos prévios (Lisboa, 2021); (ii) frequência mínima de 5 ocorrências, de modo a garantir diversidade de fontes e contextos suficientes para preenchimento das fichas terminológicas; (iii) extensão máxima de 5 palavras à esquerda (L5) e à direita (R5) do termo **língua**, permitindo recuperar UT variadas fora do padrão **língua+adjetivo** e **língua+preposição+adjetivo**.

A análise resultou em uma lista de 70.732 *clusters* organizados por frequência decrescente, da qual foram selecionadas as 50 UT designativas de conceitos de língua mais frequentes, sendo as 10 primeiras incluídas na versão inicial do VTCL. Para confirmar o caráter terminológico de cada candidato a UT, foram analisadas suas linhas de concordância, já que muitas possuíam estrutura semelhante a agrupamentos sintagmáticos frequentes, mas não terminológicos (como **língua de comunicação**, **língua de socialização**). Além disso, *clusters* que designavam idiomas, famílias linguísticas ou grupos étnicos (como **língua portuguesa**, **língua românica**, **língua indígena**) e aqueles em que o termo **língua** não era a base (como **modalidade de língua**, **status de língua**) foram desconsiderados.

A etapa seguinte consistiu em utilizar novamente a ferramenta Concord para gerar novas linhas de concordância para cada uma das 50 UT selecionadas, incluindo suas formas no plural, com o objetivo de identificar abreviaturas, acrônimos e siglas (AAS) associadas a elas. No total, foram geradas 59.469 linhas de concordância. Para buscar múltiplas expressões simultaneamente no WST, utilizamos o símbolo “/” entre os termos de busca, conforme orienta seu manual (Scott, 2015). Por exemplo, para a UT **língua adicional**, a expressão de busca foi: **língua adicional/línguas adicionais**.

As linhas de concordância geradas foram organizadas de duas formas: (i) pela primeira palavra à direita da expressão (Sort 1: R1) e pelo nome dos arquivos (Sort 2: File Name); (ii) pela primeira palavra à esquerda da expressão (Sort 1: L1) e pelo nome dos arquivos (Sort 2: File Name). Essa organização facilitou a análise das AAS, que geralmente aparecem antes ou depois das UT por extenso. O resultado dos dois primeiros procedimentos de análise está apresentado no Quadro 5, a seguir:

Quadro 5 – Unidades terminológicas e suas respectivas abreviaturas, acrônimos e siglas.

| UT                        | AAS                | UT                          | AAS     |
|---------------------------|--------------------|-----------------------------|---------|
| 1. Língua Estrangeira     | LE   L2            | 26. Língua Gestual          | LG      |
| 2. Língua Materna         | LM   L1            | 27. Língua de Ensino        | -       |
| 3. Segunda Língua         | L2   SL            | 28. Língua de Escolarização | -       |
| 4. Língua Segunda         | L2   LS            | 29. Língua de Origem        | -       |
| 5. Língua Adicional       | LA   L3   LAd   Ln | 30. Língua Internacional    | LI      |
| 6. Língua Oficial         | LO   LOF           | 31. Língua de Instrução     | LI      |
| 7. Língua de Sinais       | LS                 | 32. Língua Comum            | -       |
| 8. Língua Não Materna     | LNM   L2           | 33. Língua Crioula          | -       |
| 9. Primeira Língua        | L1   PL            | 34. Língua Geral            | -       |
| 10. Língua de Herança     | LH                 | 35. Língua Padrão           | -       |
| 11. Língua de Acolhimento | LAc   LA           | 36. Língua Maioritária      | -       |
| 12. Língua Falada         | LFal               | 37. Língua Inicial          | L1   LI |
| 13. Língua Nacional       | LN                 | 38. Língua de Trabalho      | -       |
| 14. Língua Nativa         | L1   LM   LN       | 39. Língua Não Nativa       | LNN     |
| 15. Língua Oral           | LO                 | 40. Língua Viva             | -       |
| 16. Língua Escrita        | LEsc               | 41. Língua Pluricêntrica    | -       |
| 17. Língua Dominante      | -                  | 42. Língua da Escola        | -       |
| 18. Língua Natural        | L1   LN            | 43. Língua de Partida       | LP      |
| 19. Terceira Língua       | L3                 | 44. Língua Parceira         | -       |
| 20. Língua em Uso         | -                  | 45. Língua de Imigração     | -       |
| 21. Língua Majoritária    | -                  | 46. Língua Objeto           | LO      |
| 22. Língua Primeira       | L1                 | 47. Língua Transnacional    | -       |
| 23. Língua Minoritária    | L MIN              | 48. Língua Oral-Auditiva    | -       |
| 24. Língua Veicular       | LV                 | 49. Língua Fonte            | -       |
| 25. Língua Franca         | -                  | 50. Língua Próxima          | -       |

Fonte: elaborado pelo autor com base em Lisboa (2024, p. 116, 119).

É possível perceber no Quadro 5 que algumas UT possuem mais de uma AAS associada a elas, enquanto uma mesma AAS pode se referir a várias UT. Por exemplo, LM e L1 são siglas usadas para **língua materna**, mas L1 também representa outras UT, como **língua primeira**, **primeira língua** e **língua inicial**. Nessa etapa, todas as AAS

associadas às UT foram registradas para garantir que nenhum contexto definitório ou explicativo fosse desconsiderado. Na etapa seguinte, a análise de linhas de concordância permitiu identificar os casos em que, por exemplo, L1 se referia exclusivamente a **língua materna**, possibilitando a limpeza das linhas de concordância finais e a quantificação exata da frequência dessa UT e de suas AAS.

O procedimento seguinte consistiu em gerar novas linhas de concordância para as 10 UT que comporiam a versão inicial do VTCL, agora incluindo suas formas no plural e a(s) AAS correspondente(s). Segundo o manual do WST (Scott, 2015), a barra (/) possibilita a busca simultânea por diferentes expressões de busca, enquanto a duplicação do sinal de igualdade (==) habilita a diferenciação entre letras maiúsculas e minúsculas (*case sensitive*), o que auxiliou na busca exata de AAS, sem confundi-las, por exemplo, no caso das siglas LA e LO, com artigos definidos em espanhol. Tomando como exemplo a UT **língua adicional**, a expressão de busca foi a seguinte: **língua adicional/línguas adicionais==LA== /==L3==/==LAd==/==Ln==**. No total, foram geradas 140.005 linhas de concordância referentes às 10 UT incluídas no VTCL.

Como mencionado anteriormente, foi necessário limpar as linhas de concordância geradas nesta etapa, com o intuito de excluir AAS que faziam referência a outra UT que não a UT sob análise. Só após a limpeza teríamos a frequência exata da UT com suas AAS e as linhas de concordância prontas para o último procedimento de análise, a extração de informações para preenchimento das fichas terminológicas.

A limpeza das 140.005 linhas de concordância geradas nessa etapa foi feita com as seguintes configurações: (i) Sort 1 – Center: organização alfabética das expressões de busca para focalizar as AAS; (ii) Sort 2 – Position in File: ordenação pela posição das AAS no arquivo, pois geralmente as primeiras ocorrências de AAS vêm acompanhadas da UT por extenso; (iii) Sort 3 – File Name: organização alfabética pelo código dos arquivos, visando facilitar a análise de todas as ocorrências de determinada AAS em um mesmo arquivo. Essas configurações agilizaram consideravelmente o

processo de limpeza das linhas de concordância finais. Após a limpeza, o número de linhas de concordância foi reduzido de 140.005 para 87.027.

O último procedimento dessa etapa foi a minuciosa análise das linhas de concordâncias finais das 10 UT selecionadas para compor o VTCL. Essa análise visou a extração de contextos definitórios e explicativos que veiculassem as informações necessárias para a extração de traços conceituais e preenchimento das fichas terminológicas.

Baseando-nos no processo de análise de traços conceituais utilizado pelo VoTec, elaboramos fichas de análise de traços conceituais no Microsoft Word. As fichas eram compostas por três colunas contendo as seguintes informações: código do arquivo, contexto definitório ou explicativo e uma síntese dos traços conceituais identificados nos contextos. A Figura 4, a seguir, apresenta um exemplo de uma das fichas criadas.

Figura 4 – Ficha de análise de traços conceituais da UT *Língua Estrangeira* (visão parcial).

| ARQUIVO                         | CONTEXTO   | TRAÇOS CONCEITUais                             |
|---------------------------------|--|--|
| 1 D_UFAM_2019_Rocilange Cabral  | Outra questão que merece atenção é que há a seguinte discussão: se não for Língua Materna, todas as outras são Línguas Estrangeiras;   | Hiperônimo global (qualquer língua que não LM) |
| 2 D_CEFETMG_2012_Liliane Damazo | Diante disso e, considerando ainda que o Cepe-Bras avalia o desempenho do examinando sem levar em conta onde, quando ou como ele aprendeu a língua, não são consideradas nas análises e discussões as diferenças entre LE e L2 e, por isso, utilizamos, de forma abrangente, o termo LE.   | Hiperônimo global (qualquer língua que não LM) |
| 3 D_PUCRS_2016_Claúdia Pimentel | De acordo com Stern (1991) no passado, o termo LE era usado para contrapor o termo "língua nativa" (L1).   | Hiperônimo global (qualquer língua que não LM) |
| 4 A_DIA_2012_Jorge Pinto        | Estudos em Portugal (cf. Leiria, 2004, 2006) apontam igualmente para esta perspectiva da L2 (língua não materna) como designação de língua segunda (língua do Estado e de ensino, que pode ser adquirida sem recurso à escola) e de língua estrangeira (aprendida em contextos exolíngues, recorrendo principalmente ao ensino formal).                      | Contexto exolíngue                             |
| 5 A_DL_2018_Daisy Cordeiro      | Se o aprendizado da língua alvo ocorrer fora da comunidade nativa desse idioma, diz-se que se trata de aprendizagem de língua estrangeira (LE). Quando ocorre na comunidade, trata-se de aprendizagem de segunda língua (L2).  | Não imersão                                    |
| 6 A_GL_2008_Rosane Amado        | É importante, contudo, lembrarmos que o português ensinado fora dos países lusófonos é uma língua estrangeira (PLE), isto é, os aprendizes não estão envolvidos com a cultura brasileira nem com a língua fora da sala de aula, o que requer abordagens de ensino diferentes.  | Não imersão                                    |
| 7 D_UAV_2008_João Cunha         | Em consequência destas definições, os espaços da Língua Segunda (L2) e da Língua Estrangeira (LE) separam-se claramente. A primeira tem um estatuto oficial e escolar, enquanto que a Língua Estrangeira se confina ao espaço de sala de aula.   | Não oficial                                    |
| 8 D_PUCSP_2014_Bruna Cababe     | Segundo Stern ([1983] 2003, p. 16), é necessário distinguir entre aprendizagem de segunda língua e de língua estrangeira. De acordo com esse pesquisador, a "segunda língua" normalmente tem um status oficial dentro das fronteiras territoriais em que ela tem uma função reconhecida, enquanto que o termo "língua estrangeira" não apresenta esse status | Não oficial                                    |
| 9 D_UAB_2009_Célia Barbeiro     | "O termo língua estrangeira (LE) costuma ser usado para classificar a aprendizagem e o uso de uma língua em espaços onde ela não tem nenhum estatuto sócio-político" (Leiria et alii 2005: 5).   | Sem estatuto sociopolítico                     |

Fonte: Lisboa (2024, p. 123).

Durante a extração de traços conceituais, identificamos casos de polissemia, em que uma mesma UT representava conceitos com traços conceituais compartilhados, mas que não eram completamente intercambiáveis. A Figura 4 ilustra a polissemia identificada para a UT **língua estrangeira**, em que os contextos de 1 a 3 compartilham o traço conceitual "qualquer língua que não a materna"; os de 4 a 6, a restrição "não imersão"; e os de 7 a 9, a "ausência de estatuto de oficialidade". Após tabelar todos os contextos e traços conceituais de determinada UT, reorganizamos os casos de

polissemia, agrupando contextos e traços conceituais relacionados a acepções específicas, como também exemplificado na Figura 4.

No planejamento do VTCL, como mencionado anteriormente, decidimos criar fichas terminológicas com entradas e verbetes separados para cada conceitualização distinta atribuída a uma mesma UT. As entradas foram organizadas em ordem decrescente de frequência, com base na quantificação dos agrupamentos de contextos definitórios e explicativos mencionados no parágrafo anterior. No caso da UT **Língua Estrangeira**, por exemplo, foram planejadas três entradas (**Língua Estrangeira<sup>1</sup>**, **Língua Estrangeira<sup>2</sup>** e **Língua Estrangeira<sup>3</sup>**), organizadas conforme a frequência de ocorrência da acepção no *corpus*.

Após concluir as fichas de análise de determinada UT, procedemos ao preenchimento das fichas terminológicas, elaborando uma para cada entrada do vocabulário. A Figura 5, a seguir, ilustra de forma parcial a ficha terminológica de **Segunda Língua<sup>1</sup>**. Todas as fichas completas desenvolvidas nesta pesquisa estão integralmente disponíveis nos apêndices de nossa tese de doutorado (Lisboa, 2024).

Figura 5 – Ficha terminológica de *Segunda Língua<sup>1</sup>* (visão parcial).

| Entrada: Segunda Língua <sup>1</sup>   |  |                        |  |                  |  |  |
|--|--|------------------------|--|------------------|--|--|
| Abreviatura: L2   SL                   | Nº Ocorrências: 16.293   | Categoria: Substantivo | Número: Singular   | Gênero: Feminino |  |  |
| Definição:                             | Língua adquirida após aquisição da língua materna, independentemente se em contexto formal, informal, imersivo, em não imersão, do estatuto sociopolítico da língua ou da posição que ela ocupa na ordem cronológica de aquisição linguística.   |                        |  |                  |  |  |
| Notas:                                 | Em geral, esse termo é utilizado de forma genérica para cobrir todas as possibilidades de ensino, aprendizagem e/ou uso de línguas para além da materna, especialmente em trabalhos para os quais as especificações de estatuto de oficialidade da língua, de ordem da aquisição da(s) língua(s) de caráter não materno, da situação de imersão ou não e dos contextos de ensino são dispensáveis. |                        |  |                  |  |  |
| Hiperônimo de:                         | <a href="#">Língua Estrangeira<sup>2</sup></a> , <a href="#">Segunda Língua<sup>2</sup></a> , <a href="#">Segunda Língua<sup>3</sup></a> , <a href="#">Segunda Língua<sup>5</sup></a> , <a href="#">Língua Segunda<sup>2</sup></a> , <a href="#">Língua Segunda<sup>4</sup></a> , <a href="#">Língua Segunda<sup>5</sup></a> , <a href="#">Língua Adicional<sup>2</sup></a>                        |                        |  |                  |  |  |
| Hipônimo de:                           |  |                        |  |                  |  |  |
| Co-hipônimo de:                        |  |                        |  |                  |  |  |
| Parassinônimos:                        | <a href="#">Língua Estrangeira<sup>4</sup></a> , <a href="#">Língua Segunda<sup>3</sup></a> , <a href="#">Língua Não Materna</a>   |                        |  |                  |  |  |
| Antônimos:                             | <a href="#">Língua Materna<sup>1</sup></a> , <a href="#">Primeira Língua<sup>1</sup></a>   |                        |  |                  |  |  |
| Remissivos (compare):                  | <a href="#">Segunda Língua<sup>2</sup></a> , <a href="#">Segunda Língua<sup>3</sup></a> , <a href="#">Segunda Língua<sup>4</sup></a> , <a href="#">Segunda Língua<sup>5</sup></a>  |                        |  |                  |  |  |
| Informações Enciclopédicas (Wikipedia) |  |                        |  |                  |  |  |
| Artigo:                                | Segunda Língua   | Fonte:                 | Wikipedia  |                  |  |  |
| Link:                                  | <a href="https://pt.wikipedia.org/wiki/Segunda_língua">https://pt.wikipedia.org/wiki/Segunda_língua</a>  | Definição:             | Uma segunda língua (L2) é qualquer língua aprendida após a primeira língua ou língua materna (L1). |                  |  |  |
| Multimídia                             |  |                        |  |                  |  |  |
| Youtube:                               | <a href="https://youtu.be/eEuwMpeGT_w?si=NwiItvspSaUsM6v">https://youtu.be/eEuwMpeGT_w?si=NwiItvspSaUsM6v</a>  |                        |  |                  |  |  |
| Contextos                              |  |                        |  |                  |  |  |
| A_DL_2018_Mauricio Resende             | “Neste trabalho, o termo ‘L2 (segunda língua)’ está sendo empregado para fazer referência a qualquer língua adquirida/aprendida após a língua materna (L1), seja por meio de exposição natural seja através de instrução explícita”.   |                        |  |                  |  |  |
| A_EMA_2009_Lucia Rottava               | “Para De Angelis, o termo L2 tem sido usado para referir-se a uma L2 ou a uma LE, além da primeira língua (L1)”.   |                        |  |                  |  |  |
| A_LE_2016_Celeste Ribeiro              | “A segunda língua (L2) é caracterizada de forma geral como qualquer língua aprendida após a primeira língua ou língua materna do falante”.   |                        |  |                  |  |  |
| A_RELAL_2018_Aline Sousa               | “Nesse caso, portanto, o termo L2 (ou L2s) é usado para se referir a quaisquer outras línguas não maternas que a pessoa tenha adquirido após a L1”.  |                        |  |                  |  |  |

Fonte: Lisboa (2024, p. 125).

As informações sobre AAS e frequência foram coletadas nas etapas de geração de linhas de concordâncias descritas anteriormente, ao passo que os dados de categoria gramatical, número e gênero foram inseridos durante a análise dos contextos linguísticos. As definições e notas foram elaboradas com base nos traços conceituais previamente extraídos e agrupados. Conforme estipulamos no planejamento do VTCL, o campo “Definição” seguiu o modelo aristotélico, iniciando com o hiperônimo **língua**, seguido dos traços conceituais distintivos da UT. Já o campo “Notas” incluiu informações enclopédicas e terminológicas complementares, ampliando o poder informacional do verbete.

Os campos de hiperonímia, hiponímia, co-hiponímia, parassinonímia e antonímia foram preenchidos apenas após concluir a análise das 10 UT e preenchimento das fichas terminológicas, já que não foi possível determinar essas relações conceituais por meio dos contextos linguísticos devido à falta de consenso entre os especialistas cujas produções compunham o *corpus*. Portanto, essas relações foram estabelecidas com base nas definições construídas e incluem somente as UT do VTCL. Isso, porque na página de consulta do VTCL, cada UT desses campos funciona como nódulo de *pop-up* que veicula sua definição sem a necessidade de cliques ou mudanças de páginas (razão pela qual estão em azul na Figura 5).

Para as informações enclopédicas de outras fontes, priorizamos a Wikipedia devido ao acesso gratuito e à possibilidade de reproduzir conteúdos por meio do *link* fornecido. No entanto, a definição da Wikipedia não está presente em todos os verbetes. Por exemplo, no caso de **Segunda Língua**<sup>1</sup> (Figura 5), a definição extraída da Wikipedia está alinhada aos dados do verbete, mas a mesma definição não aparece em **Segunda Língua**<sup>2</sup>, **Segunda Língua**<sup>3</sup>, **Segunda Língua**<sup>4</sup> e **Segunda Língua**<sup>5</sup>, pois esses verbetes descrevem conceitos diferentes.

No campo “Multimídia”, selecionamos vídeos do YouTube que ampliassem o conteúdo informacional do verbete, priorizando aqueles que definissem a UT. Em vista disso, esse campo não é obrigatório em todas as fichas, já que não foram

encontrados vídeos que definissem todos as acepções presentes no VTCL. Já os contextos e códigos dos arquivos foram extraídos das fichas de análise de traços conceituais (Figura 4). Embora todos os contextos apareçam nas fichas terminológicas, limitamos a exibição a seis contextos na página de consulta, priorizando os mais ricos em traços conceituais, de modo a evitar a sobrecarga de informações nos verbetes.

Esses procedimentos de análise das 10 UT designativas de conceitos de língua mais frequentes do *corpus* resultaram na elaboração de 28 fichas terminológicas, que serviram como base para a criação dos 28 verbetes que compõem o VTCL.

#### 4 Resultados

O VTCL está disponibilizado em um domínio criado especificamente para esse projeto (<http://vtcl.votec.ileel.ufu.br/>). Como especificado anteriormente, o projeto computacional foi criado com base em atualizações feitas a partir do VoTec e projetos derivados, nomeadamente TermosTeo e VoBLing. A Figura 6, a seguir, apresenta uma visão parcial da página exibida aos usuários ao acessarem o VTCL.

Figura 6 – Página inicial do VTCL (visão parcial).



The screenshot shows the homepage of the VTCL. At the top, there is a navigation bar with the logo 'VTCL' and the text 'Vocabulário Terminológico de Conceitos de Língua'. To the right of the logo are links for 'Tela cheia', 'English', and 'Ajuda'. On the left, there is a sidebar with several dropdown menus: 'Tipos de Exibição' (with 'Normal' and 'Descriativa' options), 'Tipos de Consulta' (with 'Total' and 'Modular' options), 'Consultas Externas' (with 'Google' option), and 'Como Citar' (with a link icon). The main content area features a search bar with the placeholder 'Faça uma pesquisa.' and a 'Buscar' button. Below the search bar is a list of terms, each preceded by a small orange square icon. The terms listed are: Língua Adicional<sup>1</sup>, Língua Adicional<sup>2</sup>, Língua de Herança<sup>1</sup>, Língua de Herança<sup>2</sup>, Língua de Sinais, Língua Estrangeira<sup>1</sup>, Língua Estrangeira<sup>2</sup>, Língua Estrangeira<sup>3</sup>, Língua Materna<sup>1</sup>, Língua Materna<sup>2</sup>, Língua Materna<sup>3</sup>, Língua Não Materna, Língua Oficial<sup>1</sup>, Língua Oficial<sup>2</sup>, and Língua Oficial<sup>3</sup>.

Fonte: VTCL – Lisboa (2024).

Ao acessar a página, o consultente tem acesso à lista de UT que compõem o vocabulário em questão, organizada em ordem alfabética. Dez UT foram analisadas para a versão inicial do VTCL, resultando em 28 verbetes, haja vista a polissemia identificada nas análises. Índices numéricos sobrescritos são utilizados para diferenciar acepções, sendo o número <sup>1</sup> indicador da acepção mais frequente no *corpus* de estudo, o número <sup>2</sup> indicador da segunda mais frequente e assim por diante.

Ao posicionar o cursor em cima de cada UT, surge um *pop-up* com sua definição, de modo que o consultente tenha acesso às definições sem necessariamente ter de acessar os verbetes. Cada UT presente nessa lista também funciona como *link*, no qual o usuário pode clicar para ter acesso ao seu verbete.

Nessa mesma página (Figura 6) também são exibidos no canto superior direito os recursos de visualização em tela-cheia, de alteração do idioma da interface (português ou inglês)<sup>10</sup> e de acesso ao texto externo (*outside matter*) do VTCL, disponibilizado em português e inglês e atualizado conforme o idioma de interface selecionado pelo usuário. Há também o recurso “Como Citar”, que apresenta a referência ao VTCL no formato ABNT para auxiliar os usuários que desejem utilizar o conteúdo apresentado ou citar o VTCL em suas produções acadêmicas.

Para ter acesso aos verbetes, os usuários podem clicar em determinada UT presente na nomenclatura apresentada na página inicial ou fazer uma pesquisa na caixa de busca, que apresenta indicações de UT à medida que palavras são digitadas. A Figura 7, a seguir, ilustra a visualização dos verbetes.

---

<sup>10</sup> Vale ressaltar que a mudança de idioma da interface visa apenas facilitar a navegação, mas não altera o idioma das informações dos verbetes, já que o VTCL é uma obra monolíngue em língua portuguesa.

Figura 7 – Verbete Segunda Língua<sup>1</sup> no VTCL (visão parcial).

Fonte: VTCL – Lisboa (2024).

As UT que aparecem na cor laranja funcionam simultaneamente como nódulos de *link* e *pop-up*. Portanto, é possível clicá-las para ser direcionado aos seus respectivos verbetes ou passar o cursor sobre elas para ter acesso a *pop-ups* com suas definições. Também é possível clicar em “Wikipedia” para acessar o artigo completo do qual a definição enciclopédica foi extraída e no símbolo logo após “YouTube” para acessar o vídeo selecionado para ampliar o poder informacional do verbete. Além disso, ao passar o cursor em “Compare”, é aberto um *pop-up* que apresenta, de uma só vez, todas as outras definições das demais acepções da UT do verbete, de modo a facilitar o contraste entre as acepções e melhor visualizar o panorama polissêmico da UT.

É válido mencionar também que todos os verbetes contam com abonações. As abonações são os contextos definicionais e/ou explicativos extraídos do *corpus* de estudo, veiculados com o código do arquivo do qual cada trecho foi extraído. Caso queira ter acesso aos dados completos desses arquivos, o usuário pode copiar o código e buscá-lo no texto externo (*outside matter*) disponibilizado no menu “Ajuda”, localizado no canto superior direito.

Por ter sido originado a partir do VoTec, o VTCL possui flexibilidade na visualização dos verbetes, contando com dois tipos de exibição e dois tipos de consulta. Esses recursos foram reaproveitados do VoTec e estão localizados no canto

esquerdo da página de consulta, conforme é possível verificar nas Figuras 6 e 7. O Quadro 6, a seguir, descreve os modos de exibição e consulta.

Quadro 6 – Modos de exibição e consulta do VTCL.

|                                    |   |
|------------------------------------|---|
| <b>Tipo de Exibição Normal</b>     | As informações de cada campo são dispostas de forma contínua e sequencial no verbete, seguindo o padrão tradicional das obras terminográficas e lexicográficas impressas. |
| <b>Tipo de Exibição Descritiva</b> | Os dados de cada campo são organizados em linhas separadas, com espaçamento entre elas no plano vertical.   |
| <b>Tipo de Consulta Total</b>      | Exibe todos os campos preenchidos e suas informações.   |
| <b>Tipo de Consulta Modular</b>    | Permite ao usuário ajustar os verbetes, ativando ou desativando campos e informações conforme necessário.   |

Fonte: elaborado pelo autor.

No VTCL, o usuário tem quatro opções para adequar a visualização dos verbetes às suas necessidades: (i) Exibição Normal + Consulta Total (padrão); (ii) Exibição Normal + Consulta Modular; (iii) Exibição Descritiva + Consulta Total; (iv) Exibição Descritiva + Consulta Modular. A escolha aplica-se a todos os verbetes da navegação atual, mas as configurações retornam ao padrão (Exibição Normal + Consulta Total) ao atualizar a página de consulta.

## 5 Considerações finais

Neste artigo, apresentamos uma síntese da criação do Vocabulário Terminológico de Conceitos de Língua (VTCL), produto de acesso gratuito *on-line* oriundo de nossa tese de doutorado (Lisboa, 2024). O VTCL foi fundamentado pelos princípios da Teoria Comunicativa da Terminologia, da Terminografia e da Linguística de *Corpus*. Ele se constitui como uma obra de referência que apresenta de maneira sintetizada e sistematizada uma amostragem da rede terminológico-conceitual dos conceitos de língua, bem como indica as relações e fronteiras entre os conceitos analisados.

Sua versão inicial conta com o resultado de análises das 10 unidades terminológicas (UT) designativas de conceitos de língua mais frequentes em um *corpus*

de especialidade de mais de 23 milhões de *tokens*, composto por 325 artigos, 521 dissertações e 117 teses da área de Português como Língua Não Materna, analisado por meio do *software* WordSmith Tools 8.0. Seu banco de dados e sua página de consulta foram construídos a partir de atualizações feitas nos projetos computacionais do VoTec (Fromm, 2007) e seus projetos derivados, nomeadamente TermosTeo (Cardoso, 2017) e VoBLing (Yamamoto, 2020).

Em função da polissemia presente na rede terminológica focalizada, as 10 UT analisadas deram origem a 28 verbetes. Em síntese, os verbetes do VTCL incluem os conceitos de língua, informações gramaticais, definições aristotélicas, notas explicativas, siglas, frequência no *corpus*, relações conceituais (hipônimos, hiperônimos, co-hipônimos, parassinônimos e antônimos), informações enciclopédicas (Wikipedia), recursos multimídia (YouTube), comparações entre acepções, bem como contextos extraídos do *corpus* com identificação dos arquivos.

O VTCL também possui outros recursos adicionais, como diferentes modos de exibição e consulta, interface e texto externo (*outside matter*) em inglês e português, *links* internos entre verbetes, bem como janelas *pop-up* para sínteses e contrastes entre definições. Apesar de inicialmente contar com apenas as 10 UT designativas de conceitos de língua mais frequentes no *corpus* de estudo, os fundamentos teórico-metodológicos estabelecidos permitirão sua expansão para 50 UT identificadas no *corpus* de estudo, além da possibilidade de transformá-lo em uma obra bilíngue ou multilíngue.

Esperamos que o VTCL auxilie pesquisadores e professores de línguas a compreenderem melhor o panorama terminológico e conceitual dos conceitos de língua, incluindo suas variadas possibilidades definicionais e inter-relações conceituais. Esperamos também que seu uso contribua para a precisão terminológica na comunicação especializada e facilite a operacionalização desses conceitos na pesquisa e prática profissional.

## Referências

ALMEIDA, G. M. B; CORREIA, M. Terminologia e *corpus*: relações, métodos e recursos. In: TAGNIN, S. E. O.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 67-94.

ALUÍSIO, S. M.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística. **Calidoscópio**, São Leopoldo, v. 4, n. 3, p. 156-178, 2006. Disponível em: <http://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>.

ARAÚJO, L. B. **Vocabulário terminológico para o ensino de línguas**: proposta metodológica. 2021. Dissertação (Mestrado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2021. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/65375>.

ARONIN, L.; Ó LAOIRE, M.; SINGLETON, D. The multiple faces of multilingualism: language nominations. **Applied Linguistics Review**, [s. l.], v. 2, n. 1, p. 169-190, 2011. DOI <https://doi.org/10.1515/9783110239331.169>

ARONIN, L.; SINGLETON, D. Affordances and the diversity of multilingualism. **International Journal of the Sociology of Language**, [s. l.], v. 2010, n. 205, p. 105-129, 2010. DOI <https://doi.org/10.1515/ijsl.2010.041>

AUBERT, F. H. **Introdução à metodologia da pesquisa terminológica bilingüe**. 2. ed. São Paulo: FFLCH/CITRAT, 2001.

BARBOSA, J. B.; FREIRE, D. J. Formação de professores e ensino de Português como Língua Adicional. **Estudos Linguísticos**, São Paulo, v. 46, n. 2, p. 593-602, 2017. DOI <https://doi.org/10.21165/el.v46i2.1721>

BARBOSA, M. A. Contribuição ao estudo de aspectos da tipologia de obras lexicográficas. **Ciência da Informação**, Brasília, v. 24, n. 3, p. 1-9, 1995. Disponível em: <https://revista.ibict.br/ciinf/article/view/572>

BARBOSA, M. A. Dicionário, vocabulário, glossário: concepções. In: ALVES, I. M. (org.). **A constituição da normalização terminológica no Brasil**. 2. ed. São Paulo: FFLCH/CITRAT, 2001. p. 23-45.

BARROS, L. A. **Curso básico de Terminologia**. São Paulo: Edusp, 2004.

BERBER SARDINHA, T. **Lingüística de Corpus**. Barueri: Manole, 2004.

CABRÉ, M. T. Elements for a theory of terminology: towards an alternative paradigm. **Terminology**, Amsterdam, v. 6, n. 1, p. 35-57, 2000. DOI <https://doi.org/10.1075/term.6.1.03cab>

CABRÉ, M. T. La Teoría Comunicativa de la Terminología: una aproximación lingüística a los términos. **Revue Française de Linguistique Appliquée**, Paris, v. 14, n. 2, p. 9-15, 2009. DOI <https://doi.org/10.3917/rfla.142.0009>

CABRÉ, M. T. **La terminología**: representación y comunicación. Barcelona: Universitat Pompeu Fabra, 1999.

CANDIAN, M.; BESSA, M. C. Português como Segunda Língua Estrangeira Não Materna Adicional para Falantes de Outras Línguas - P2LENMAFOL: uma breve análise de terminologias. **Veredas**, Juiz de Fora, v. 25, n. 2, p. 375-396, 2021. Disponível em: <https://periodicos.ufjf.br/index.php/veredas/article/view/35916>.

CARDOSO, S. A. F. **TermosTeo**: a elaboração de vocabulários monolíngues de termos da Teologia em um estudo conduzido por *corpus*. 2017. Tese (Doutorado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2017. Disponível em: <https://repositorio.ufu.br/handle/123456789/21349>.

FERREIRA, J. G. B. O ensino de língua portuguesa para estrangeiros: implicações da pluralidade de conceitos. **EntreLínguas**, Araraquara, v. 7, n. 6, p. 1-17, 2021. DOI <https://doi.org/10.29051/el.v7iesp.6.15423>

FRIEDRICH, P.; MATSUDA, A. When five words are not enough: a conceptual and terminological discussion of English as a lingua franca. **International Multilingual Research Journal**, Philadelphia, v. 4, n. 1, p. 20-30, 2010. DOI <https://doi.org/10.1080/19313150903500978>

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. DOI <https://doi.org/10.11606/T.8.2008.tde-08072008-150855>

FROMM, G.; LISBOA, J. V. R. VoTec terminographic environment over the years: brief overview. **Acta Scientiarum. Language and Culture**, Maringá, v. 45, n. 2, p. 1-16, 2024. DOI <https://doi.org/10.4025/actascilangcult.v45i2.67669>

GROSSO, M. J. R. Língua de acolhimento, língua de integração. **Horizontes de Linguística Aplicada**, Brasília, v. 9, n. 2, p. 61-77, 2010. DOI <https://doi.org/10.26512/rhla.v9i2.886>

HAMMARBERG, B. The languages of the multilingual: some conceptual and terminological issues. **International Review of Applied Linguistics in Language Teaching**, [s. l.], v. 48, n. 2-3, p. 91-104, 2010. DOI <https://doi.org/10.1515/iral.2010.005>

HARTMANN, R. R. K.; JAMES, G. **Dictionary of Lexicography**. Londres: Routledge, 1998.

HOFFMANN, A. **Sinonímia e hiperonímia**: das relações entre palavras para as relações de sentido. 2014. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal da Fronteira Sul, Chapecó, 2014. Disponível em: <https://rd.uffs.edu.br/handle/prefix/769>.

KILIAN, C. K.; REUILlard, P. C. R. As decisões prévias. In: BEVILACQUA, C. R. et al. (org.). **Como elaborar um dicionário especializado?** Porto Alegre: Zouk, 2023. p. 32-44.

KOESTER, A. Building small specialised corpora. In: O'KEEFFE, A.; McCARTHY, M. (ed.). **The Routledge Handbook of Corpus Linguistics**. Abingdon: Routledge, 2010. p. 66-79.

KOTEYKO, N. Compilation of specialised corpora. In: KOTEYKO, N. **Language and Politics in Post-Soviet Russia**: a corpus assisted approach. Londres: Palgrave MacMillan, 2014. p. 48-64. DOI <https://doi.org/10.1057/9781137314093>

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

LISBOA, J. V. R. **Proposta de harmonização da terminologia designativa de área e subáreas do Português como Língua Não Materna baseada em corpus**. 2021. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2021. DOI <http://doi.org/10.14393/ufu.di.2021.161>

LISBOA, J. V. R. Reflexões iniciais sobre a terminologia designativa de área e subáreas do Português para Falantes de Outras Línguas: um estudo exploratório. **Revista GTLex**, Uberlândia, v. 5, n. 2, p. 283-311, 2020. DOI <https://doi.org/10.14393/Lex9-v5n2a2020-5>

LISBOA, J. V. R. **Vocabulário Terminológico de Conceitos de Língua**. 2024. Tese (Doutorado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2024. DOI <http://doi.org/10.14393/ufu.te.2024.572>

REY, A. **Essays on Terminology**. Tradução e edição de Juan Carlos Sager. Amsterdam: John Benjamins, 1995. DOI <https://doi.org/10.1075/btl.9>

SCOTT, M. **WordSmith Tools Manual**. Stroud: Lexical Analysis Software, 2015. Disponível em: <https://lexically.net/downloads/version6/wordsmith6.pdf>.

SCOTT, M. **WordSmith Tools version 8**. Stroud: Lexical Analysis Software, 2022. Disponível em: [https://www.lexically.net/downloads/version\\_64\\_8/index.html](https://www.lexically.net/downloads/version_64_8/index.html).

SPINASSÉ, K. P. Os conceitos Língua Materna, Segunda Língua e Língua Estrangeira e os falantes de línguas alóctones minoritárias no Sul do Brasil. **Contingentia**, Porto Alegre, v. 1, n. 1, p. 1-8, 2006. Disponível em: <https://seer.ufrgs.br/index.php/contingentia/article/view/3837>.

WELKER, H. A. **Dicionários**: uma pequena introdução à lexicografia. 2. ed. Brasília: Thesaurus, 2004.

YAMAMOTO, M. I. **VoBLing**: vocabulário bilíngue de Linguística, português-inglês, direcionado por *corpus*. 2020. Tese (Doutorado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2020. DOI <http://doi.org/10.14393/ufu.te.2020.682>