

COMO ELABORAR UM DICIONÁRIO ESPECIALIZADO?

A experiência do Grupo  TermiSul

Organização

Cleci Regina Bevilacqua
Denise Regina de Sales
Márcia Moura da Silva
Patrícia Chittoni Ramos Reuillard
Sandra Dias Loguerio

editora

ZO
UK

COMO ELABORAR UM DICIONÁRIO ESPECIALIZADO?

Porto Alegre • 2023 • 1ª edição

Organização

Cleci Regina Bevilacqua
Denise Regina de Sales
Márcia Moura da Silva
Patrícia Chittoni Ramos Reuillard
Sandra Dias Loguercio

editora

ZO
UK

2023 © Cleci Regina Bevilacqua, Denise Regina de Sales, Márcia Moura da
Silva, Patrícia Chittoni Ramos Reuillard e Sandra Dias Loguercio

Projeto gráfico e edição: Editora Zouk

Revisão: Cristiane Krause Kilian

Revisão técnica: Silvana de Fátima Bojanoski

Design da capa: Mateus Moura Godinho

**Dados Internacionais de Catalogação na
Publicação (CIP) de acordo com ISBD**

Elaborado por Odílio Hilario Moreira Junior - CRB-8/9949

C735

Como elaborar um dicionário especializado? [recurso eletrônico] /
organizado por Cleci Regina Bevilacqua, Denise Regina de Sales, Márcia Moura
da Silva, Patrícia Chittoni Ramos Reuillard e Sandra Dias Loguercio - Porto
Alegre, RS : Zouk, 2023.

137 p. ; ePUB.

Inclui bibliografia.

ISBN: 978-65-5778-119-7 (Ebook)

1. Dicionário. I. Bevilacqua, Cleci Regina. II. Sales, Denise Regina de. III.
Silva, Márcia Moura da. IV. Título.

2023-????

CDD 403

CDU 403

direitos desta edição reservados à

Editora Zouk

Av. Cristóvão Colombo, 1343 sl. 203

90560-004 – Floresta – Porto Alegre – RS – Brasil

f. 51. 3024.7554

www.editorazouk.com.br

Dedicamos este livro às nossas mestras, Maria da Graça Krieger e Anna Maria Becker Maciel, fundadoras do Grupo Termisul, pela coragem de iniciar os estudos em Terminologia e Terminografia na UFRGS e por projetá-los nacional e internacionalmente. Nosso agradecimento por tantos anos de dedicação e pelo aprendizado que nos proporcionaram.

Sumário

Prefácio

Ieda Maria Alves

7

Introdução

Cleci Regina Bevilacqua, Denise Regina de Sales, Márcia Moura da Silva,
Patrícia Chittoni Ramos Reuillard e Sandra Dias Loguercio

10

Capítulo 1 – Quando a teoria e a prática se encontram

Cleci Regina Bevilacqua e Cristiane Krause Kilian

14

Capítulo 2 – As decisões prévias

Cristiane Krause Kilian e Patrícia Chittoni Ramos Reuillard

32

Capítulo 3 – Constituição de corpora: critérios de coleta, limpeza e organização

Márcia Moura da Silva e Manuela Arcos Machado

45

Capítulo 4 – Seleção de unidades terminológicas: estratégias de extração e princípios de identificação

Sandra Dias Loguercio e Manuela Arcos Machado

66

Capítulo 5 – A ficha terminológica

Cleci Regina Bevilacqua, Denise Regina de Sales e Márcia Moura da Silva

90

**Capítulo 6 – Busca e identificação de equivalentes em línguas
estrangeiras**

Alexia Gonçalves Pokorski, Ana Letícia Prado de Campos, Iago Marques
Barragan, Marina Canofre dos Santos, Tainara Cecília Silveira Balt

104

Posfácio

Maria José Bocorny Finatto

118

Referências

122

Sobre os autores

129

Chave de respostas das atividades propostas

133

Prefácio

Ieda Maria Alves¹

Neste prefácio, que gentilmente as organizadoras do volume *Como elaborar um dicionário especializado? A experiência do Grupo TERMISUL* me incumbiram de escrever, salientarei, especialmente, aspectos históricos sobre a formação do Grupo Termisul e sobre sua inestimável contribuição para a introdução e o desenvolvimento da Terminologia no Brasil.

Discorrer sobre a história do Termisul significa, também, relemburar o período de consolidação da Terminologia no Brasil e na América Latina, estimulada muito particularmente pela União Latina, entidade sediada na França que atuou de 1954 a 2012 e muito contribuiu para a consolidação e a divulgação dos estudos e das atividades terminológicas no Brasil e nos países de línguas latinas, da América e da Europa.

A criação do grupo, gestada em 1990, consolidou-se em 1991. Correspondeu a um desejo de dirigentes da Universidade Federal do Rio Grande do Sul (UFRGS) de contribuir para a criação do Mercosul, o Mercado Comum do Sul, bloco econômico que engloba Brasil, Argentina, Uruguai e Paraguai. Agregando países vizinhos, porém com distintas línguas oficiais – português, espanhol e guarani (no Paraguai) –, a criação do Mercosul foi decisiva para a criação do Termisul, que, sob a coordenação das profas. Maria da Graça Krieger e Anna Maria Becker Maciel, vem contribuindo, até os dias contemporâneos, para o estudo e a difusão de diferentes terminologias no português brasileiro.

A contribuição do Termisul para as atividades terminológicas e para a tradução de textos técnico-científicos tem sido observada desde o início da formação da equipe, como já atesta o capítulo *Projeto cone sul: pesquisas terminológicas e tradução*, que integra o conjunto de comunicações publicadas nos *Anais do II Simpósio Latino-Americano de Terminologia e I Encontro*

¹ Faculdade de Filosofia, Letras e Ciências Humanas, USP.

Brasileiro de Terminologia Técnico-Científica (1992), realizado em 1990, em Brasília, com apoio do IBICT e da União Latina. Nesse texto, a equipe de professores inicialmente formada apresenta três objetivos gerais para o desenvolvimento de suas atividades – constituição de um núcleo de pesquisas terminológicas na UFRGS; desenvolvimento de linhas de pesquisa nas áreas de Terminologia, Lexicologia, Lexicografia e Tradução; formação de terminólogos –, que têm sido seguidos até os dias atuais.

O primeiro trabalho coletivo do Grupo, o *Dicionário de Direito Ambiental: terminologia das leis do meio ambiente* (1998, reeditado em 2008), resultou da colaboração entre a UFRGS e o Ministério Público Federal, representado pela participação do Dr. João Carlos de Carvalho Rocha, procurador da República. Esse trabalho, de natureza seminal na elaboração de dicionários terminológicos, representa uma salutar parceria que caracteriza os trabalhos terminológicos adequadamente elaborados, em que o saber do terminólogo é complementado pelo saber especializado do especialista da área enfocada.

Outras produções terminológicas, das quais têm participado vários membros do Grupo, seguiram-se a esse dicionário: *Temas de Terminologia* (2001), *Glossário Multilíngue de Direito Ambiental Internacional: terminologia dos tratados* (2004), *Glossário de Gestão Ambiental* (2006). Ao longo dos anos, outras áreas terminológicas, além do Direito Ambiental, também têm sido abordadas – conservação dos bens culturais móveis em papel, identificação e representação de UFEs em distintas terminologias, patrimônio cultural imaterial, assim como estudos sistemáticos sobre a constituição das unidades fraseológicas especializadas.

Por último, mas igualmente importante, menciono a contribuição do volume que a atual constituição do Grupo Termisul apresenta. Com trabalhos que versam sobre as várias fases do trabalho terminológico, os capítulos produzidos representam um importante material, tanto para terminólogos como para tradutores que necessitam de terminologias para suas atividades tradutórias.

Desse modo, pesquisadore(a)s de diferentes gerações seguem dando continuidade ao trabalho pioneiro de Maria da Graça Krieger e de

Anna Maria Becker Maciel, apresentando estudos que atestam que a Terminologia, em seus aspectos teóricos e práticos, é uma disciplina em constante evolução e que se adapta às mudanças e às necessidades do mundo contemporâneo.

Introdução

Este livro surgiu por duas razões fundamentais. Por um lado, buscamos oferecer um guia que reunisse, de forma simples e didática, os conhecimentos para a construção de dicionários especializados e que pudesse, igualmente, proporcionar a reflexão sobre as linguagens especializadas e suas formas de expressão; por outro, objetivamos sistematizar os conhecimentos acumulados na pesquisa em Terminologia e na produção de produtos terminográficos (dicionários, glossários e bases de dados) realizadas há mais de 30 anos pelo grupo Termisul do Instituto de Letras da Universidade Federal do Rio Grande do Sul (UFRGS).

Considerando que este é um material didático que enfoca sobretudo as etapas da elaboração desses produtos, ele foi pensado para tradutores que utilizam a terminologia com fins de busca de soluções tradutórias e também para terminólogos e/ou terminógrafos que se dedicam à organização de produtos terminográficos e que podem desempenhar suas funções em agências de tradução, empresas, órgãos públicos, editoras etc.

Como tradutor ou futuro profissional da tradução e/ou da interpretação¹, você pode se perguntar qual seria a utilidade de saber reconhecer ou mesmo compilar termos (unidades lexicais de sentido especializado ou, tão simplesmente, o léxico temático de uma área). Ora, sabemos que tradutores e intérpretes traduzem/interpretam textos – que, por sua vez, remetem a discursos –, não palavras, sejam essas temáticas ou não, e que, sozinha, uma lista de palavras não dá conta dos textos a serem traduzidos. Sabemos também que as unidades léxicas adquirem sentido nos textos, resultante de uma situação de enunciação, ou seja, de um sujeito se dirigindo a outro sujeito, em um determinado tempo e espaço, com um dado propósito ou motivação, normalmente sobre algum tema específico.

Com os textos e, mais precisamente, com as palavras de especialistas não é tão diferente, senão pela particularidade de que seus textos, em

¹ Ao longo do livro, usamos o termo tradução e/ou tradutores para fazer referência à tradução como uma atividade ampla, que engloba todas as modalidades escritas e orais.

determinada situação de comunicação, refletem um conhecimento científico (ou técnico, laboral, filosófico). Esse conhecimento produzido ao longo do tempo é fruto não de um único sujeito, mas de um grupo relativamente restrito de indivíduos (seus principais interlocutores, colegas de trabalho, colegas de área ou de áreas afins), os quais não necessariamente compartilham o mesmo local de trabalho, podendo estar em lugares (cidades, regiões, países e continentes) diferentes, e que, com base no que já se disse sobre dado assunto – no que já se sabe, já se experimentou ou testou –, elaboram e difundem novas experiências, formulam ou amadurecem ideias, comumente em detrimento ou mesmo oposição a outras, elaboram e discutem propostas, descrevem e difundem técnicas e métodos, entre outras ações. É desse trabalho contínuo feito através da linguagem – matéria-prima da interação humana – que se forjam os termos, a designação e, ao mesmo tempo, a compreensão de um conceito, uma noção ou uma técnica, e de tudo que os sustenta, sejam instrumentos teóricos (abstratos), sejam instrumentos materiais (concretos).

As terminologias se tornam assim “janelas” para o conhecimento especializado, por meio das quais tradutores e intérpretes – assim como redatores, de modo geral – entram em contato com universos de saberes e a linguagem que neles se produz. É por meio desse contato que poderão, por sua vez, traçar caminhos e soluções para suas produções tradutórias. É com o intuito de fornecer informações básicas que auxiliem na organização das terminologias e, portanto, que sirvam de subsídios aos tradutores e intérpretes, que elaboramos este livro.

Em relação à segunda razão mencionada acima, trazemos alguns dados sobre o Grupo Termisul. O grupo foi fundado no final de 1990 por Maria da Graça Krieger e Anna Maria Becker Maciel. Ao longo de toda essa trajetória, contou com a participação de professores do Departamento de Línguas Modernas, de Línguas Clássicas e Vernáculas e de Linguística, Filologia e Teoria Literária. Entre os colegas que participaram estão: Maria Lúcia Machado de Lorenci (*in memeoriam*), Sônia Gehring, Teresinha Fávero, Terumi Koto Bonnet Villalba (*in memeoriam*) e Vera Lúcia do Amaral (*in memeoriam*). Atualmente, o grupo é formado

por Anna Maria Becker Maciel, Cleci Regina Bevilacqua, Cristiane Krause Kilian, Denise Regina de Sales, Gisele de Oliveira Bosquesi, Márcia Moura da Silva, Maria José Bocorny Finatto, Patrícia Chittoni Ramos Reuillard, Patrizia Cavallo, Sandra Dias Loguercio e Silvana de Fátima Bojanoski.

Seu surgimento está atrelado à proposta da Pró-Reitoria de Extensão da UFRGS que, em 1990, reuniu um grupo de professores com a finalidade de realizar atividades de extensão que contribuíssem para a operacionalização e consolidação do Mercosul. O reconhecimento da necessidade de elaboração de produtos lexicográficos levou a Pró-Reitoria de Extensão a convidar a profa. Maria da Graça Krieger a participar da equipe. Esse ponto de partida, atrelado à criação do Mercosul, explica a origem do nome Termisul, Projeto Terminológico Cone Sul.

Ao longo de mais de trinta anos, o grupo tem se dedicado à pesquisa teórica e aplicada em Terminologia e Terminografia, com uma ampla produção de livros, artigos e apresentações de trabalhos, além de dicionários, glossários e bases de dados, mencionados ao longo deste livro. Além disso, foi pioneiro ao inserir disciplinas de Terminologia no Curso de Bacharelado em Letras – Tradução, do Instituto de Letras. Igualmente, propôs a linha de pesquisa Lexicografia, Terminologia e Tradução: relações textuais, hoje denominada Estudos do Léxico e da Tradução, na área dos Estudos da Linguagem no Programa de Pós-Graduação em Letras, abrindo espaço para a produção de teses e dissertações relativas ao léxico especializado, mas também ao léxico geral e sua interface com a tradução e o ensino de línguas.

Nessa caminhada, estivemos sempre acompanhados de alunos de graduação e de pós-graduação. Foram tantos bolsistas de iniciação científica que seria impossível nomeá-los todos aqui. No entanto, gostaríamos de homenagear e agradecer a todos ao mencionar a eterna bolsista do Termisul, Glades Dilelio Noble. Sua dedicação, comprometimento e exemplo foram fundamentais durante muitos anos.

Agradecemos ainda à UFRGS, ao CNPq e à Fapergs pelas bolsas de iniciação científica concedidas às pesquisadoras do grupo ao longo

de sua trajetória e ao CNPq pelo financiamento de alguns dos projetos desenvolvidos.

Este livro foi elaborado por pesquisadoras e estudantes de pós-graduação e graduação que participam ou participavam do projeto no momento da elaboração deste livro e apoia-se na principal marca do grupo desde sua origem: o trabalho coletivo e multicultural, visto na inclusão de diferentes línguas e na conjugação de abordagens e métodos oriundos de trajetórias acadêmicas diversas.

Sua estrutura segue as distintas e sucessivas etapas do trabalho terminográfico. Para tanto, contém um capítulo teórico inicial que sintetiza alguns aspectos gerais da Terminologia e enfoca a Terminologia Linguístico-Textual (TLT), postulada sobretudo por Maria da Graça Krieger e desenvolvida pelos demais membros do grupo em suas pesquisas. O capítulo dois trata das decisões prévias para a elaboração de repertórios terminográficos (delimitação da área; usuários, finalidade da obra, equipe de trabalho e recursos necessários para sua realização); o terceiro aborda a constituição de *corpora*, considerando os critérios de coleta, limpeza e organização; o quarto discorre sobre a identificação e seleção dos termos; o quinto explica a construção da ficha terminológica para registro das unidades selecionadas; e o sexto menciona os procedimentos para a identificação dos equivalentes em línguas estrangeiras. As referências mencionadas nos capítulos se encontram no final do livro.

Pensando no seu uso didático e na sua função pedagógica, caixas explicativas complementam as informações dadas nos textos e, ao final de cada capítulo, há a indicação de leituras complementares (**Para saber mais**), que visam a ampliar os conteúdos oferecidos. Os capítulos dois a seis apresentam, ao final, exercícios de aplicação do tema tratado no capítulo.

Esperamos que este material possa ser útil e possibilite a reflexão sobre as linguagens especializadas e a elaboração de muitos dicionários, glossários e bases de dados especializados. Boa leitura e bom aprendizado!

As organizadoras

Capítulo 1 – Quando a teoria e a prática se encontram

Cleci Regina Bevilacqua
Cristiane Krause Kilian

Neste capítulo, trazemos alguns aspectos teórico-metodológicos que situam e sustentam o fazer terminográfico, a Terminografia. Assim, situamos essa disciplina nos Estudos do Léxico, fazemos referência a algumas teorias da Terminologia e apresentamos os fundamentos básicos propostos pela Terminologia Linguístico-Textual (TLT), além de alguns conceitos que permeiam os capítulos do presente livro.

Os Estudos do Léxico e o lugar da Terminografia

No âmbito dos Estudos do Léxico, como em vários outros domínios do saber, há áreas que se inter-relacionam. Por exemplo, a Lexicologia tem como objeto de estudo as palavras e o conhecimento que os falantes têm sobre elas – sua forma de pronúncia, constituição, significado, flexões e usos. Por sua vez, a Lexicografia dedica-se à elaboração de dicionários relacionados ao léxico geral de uma comunidade linguística, por exemplo, os dicionários *Aurélio*¹ e *Houaiss*² para o português do Brasil. Essas disciplinas se relacionam porque os fenômenos descritos pela Lexicologia, como a formação de palavras e a sinonímia, podem estar representados nos dicionários, e os lexicógrafos precisam pensar na maneira adequada de inseri-los nessas obras a fim de que as informações ali contidas sejam entendidas pelos seus consulentes e os ajudem a resolver suas dúvidas. A

1 FERREIRA, Aurélio B. H. *Dicionário Aurélio da língua portuguesa*. Curitiba: Positivo, 2010. Mais informações em: <https://loja.editorapositivo.com.br/catalogsearch/result/?q=dicion%C3%A9rio+Aur%C3%A9lio>.

2 HOUAISS, Antônio. *Dicionário Houaiss de língua portuguesa*. São Paulo: Objetiva, 2009. Mais informações em: <https://www.companhiadasletras.com.br/livro/9788573029635/dicionario-houaiss-da-lingua-portuguesa>.

Lexicologia e a Lexicografia, junto com a Terminologia e a Terminografia, constituem as Ciências do Léxico, denominação consagrada pelos estudiosos em nosso país.³

Embora o foco aqui seja a Terminografia, não podemos deixar de falar da Terminologia, posto que ambas constituem as duas faces – teórica e aplicada – dos mesmos objetos de estudo, os termos e as Unidades Fraseológicas Especializadas (UFEs) e a definição terminológica.

O que é a Terminologia?

É uma disciplina da Linguística dedicada ao estudo do léxico especializado, ou seja, o conjunto de termos e **fraseologias*** das diferentes áreas do conhecimento, como o Direito, a Física, a História, a Linguística, a Conservação e Restauração de Bens em papel e o Patrimônio Imaterial. Em contraste com o léxico geral utilizado em nosso dia a dia (por exemplo, *casa, cadeira, flor, árvore*), nessas áreas, são utilizadas unidades lexicais com sentidos específicos que vão sendo criadas e estabelecidas pelas **comunidades de saber***⁴ e utilizadas em seus contextos de comunicação específicos. Esse conjunto de unidades conforma a terminologia dessas diferentes comunidades e

3 No Brasil, os pesquisadores das diversas áreas que conformam as Ciências do Léxico fazem parte do Grupo de Trabalho em Lexicologia, Lexicografia e Terminologia (GTLex) da Associação Nacional de Pós-Graduação em Letras (Anpoll). Para saber mais sobre o grupo, consulte: <http://www.letras.ufmg.br/gtlex/>.

4 A proposta do termo e de sua definição foi sugerida por Sandra Loguerio no decorrer da elaboração do presente livro.

* As fraseologias são unidades formadas por duas ou mais palavras que formam um todo de sentido. Podem ser de diversos tipos. Na língua geral, podem ser, por exemplo, colocações (*tomar uma decisão*); expressões idiomáticas (*chutar o balde*); e expressões ou formulações usadas na interação (*Muito prazer!; Feliz aniversário!; Minha opinião é...*).

Nas linguagens especializadas, podem ser colocações especializadas (*conjuguar um verbo*) ou formulações (*esta lei entra em vigor na data de sua publicação; Para dirimir quaisquer questões, as partes elegem o Foro de...*). Dependendo da perspectiva do autor, podem receber diversas denominações: colocações especializadas, unidades fraseológicas especializadas, combinatórias léxicas especializadas, entre outras (Bevilacqua, 2020).

* Conjuntos de sujeitos que compartilham esferas de atividades (científicas, laborais, artísticas etc.) e, a partir destas, desenvolvem saberes. O conhecimento compartilhado por esses sujeitos de saber se produz por meio de e/ou resulta em discursos (escritos e/ou orais) que refletem sua maneira de compreender as coisas e de se expressar, caracterizando, assim, formas de dizer específicas, marcadas por termos e fraseologias. Tal noção, derivada dos conceitos de comunidade interpretativa (Fish, 1980), associada à leitura e aos textos que os sujeitos produzem a partir desta, e de comunidade discursiva (Swales, 1990), associada à interação escrita acadêmico-científica, busca romper com as fronteiras escrito/oral, jargão e termos técnico-científicos, conhecimento prático e teórico, entre outras cisões que tendem a hierarquizar os diferentes saberes.

áreas. Por exemplo, na Linguística, há termos como *categoria gramatical*, *enunciado*, *linguagem*, *língua*, *morfema*, *texto* e *verbo* com definições próprias. Temos também fraseologias como *conjuguar um verbo*, *revisar um texto* e *produzir um enunciado*. Além dos termos e UFEs, algumas perspectivas teóricas se debruçam também sobre as definições terminológicas.

Seu objetivo é, então, recolher, analisar e descrever os termos e UFEs e refletir sobre as definições considerando determinados princípios teóricos com a finalidade de entender sua constituição e funcionamento (Krieger, 2005), visando facilitar a comunicação entre os especialistas, mas também entre os aprendizes de determinada área. Isso porque, quando estamos nos formando em uma determinada profissão ou aprendendo algo novo, parte dos conhecimentos que adquirimos é ou está relacionada à sua terminologia. Além disso, os estudos terminológicos podem auxiliar na divulgação do conhecimento especializado para um público mais amplo, constituindo o que se chama popularização da ciência.

Ao observarmos os parágrafos anteriores, vemos que há referência aos termos Terminologia e terminologia. Temos aqui um caso de variação conceitual, ou seja, um mesmo termo tem dois sentidos: Terminologia (T maiúsculo) é utilizada para referir-se à disciplina linguística que se dedica ao estudo dos termos, às UFEs – na TLT também à definição, como mencionado adiante –, e terminologia (t minúsculo), ao conjunto dos termos e/ou UFE próprios de uma ciência, arte, técnica ou profissão, ou seja, utilizados por diferentes comunidades de saber.

É importante destacar que a Terminologia é multidisciplinar, posto que está integrada por vários aspectos (linguísticos e semióticos*, cognitivos* e comunicativos*), mas é também transdisciplinar, por se dedicar ao estudo do

* A perspectiva linguística permite explicar os termos como elementos da linguagem natural e a ativação de seu caráter especializado em função de seu uso em contextos específicos. A semiótica explica que o conhecimento especializado pode estar representado não só por unidades da linguagem, mas também por outros sistemas simbólicos, não linguísticos (p. ex., fórmulas, símbolos) (Cabré, 1999).

* A perspectiva cognitiva, ou seja, a do conhecimento, permite explicar “como conceituamos a realidade, os tipos de conceitualização que podem ocorrer e a relação dos conceitos entre e com suas denominações”. (Cabré, 1999, p. 122, tradução nossa).

* A perspectiva comunicativa possibilita descrever as diferentes situações e formas de comunicação e, consequentemente, “as características, possibilidades e limites dos diferentes sistemas de expressão de um conceito e de suas unidades” (Cabré, 1999, p.123, tradução nossa).

léxico das diferentes áreas do conhecimento (terminologia do Direito, da Linguística, do Patrimônio Imaterial etc.). Por sua aplicabilidade, estabelece a interface com outras áreas, como a Tradução, o Jornalismo e a Documentação. Por exemplo, a tradução oferece conhecimento sobre textos escritos originalmente em uma língua que o leitor não domina, a língua em que o texto original foi escrito; os textos jornalísticos divulgam notícias relacionadas às novas descobertas científicas que não conhecemos, auxiliando na popularização do conhecimento especializado; e os procedimentos realizados no âmbito da Documentação organizam os descritores (palavras-chave de busca de documentos) de forma semelhante às terminologias. Em todas essas áreas, que trabalham com a linguagem, utilizam-se termos e UFEs, o que requer, portanto, algum conhecimento sobre Terminologia. Finalmente, por utilizar diversos recursos informáticos para extrair termos, UFEs, definições e demais informações constantes nos produtos terminográficos a partir de texto, também estabelece uma inter-relação com a Linguística de *Corpus*.

O que é a Terminografia?⁵

É a face aplicada da Terminologia. Ocupa-se da produção de obras de referência especializadas, que visam ao registro de termos e/ou de UFEs. Para a construção desse tipo de obras, seguem-se determinados parâmetros teórico-metodológicos, os quais comentaremos mais adiante. Como resultado, temos **glossários*** e **dicionários*** em formato impresso ou eletrônico e **bases de dados terminológicos***, que podem ser monolíngues, bilíngues ou plurilíngues, de acordo com os objetivos estabelecidos e as decisões relativas às suas funções e aos seus usuários, conforme indicado no capítulo 2.

* Repertório terminológico de uma área sem pretensão de exaustividade, ou seja, contém um número menor de entradas, podendo oferecer informações gramaticais (gênero e número), definição e equivalentes.

* Repertório terminológico de uma área que abarca um número maior de entradas, podendo incluir informações gramaticais, definições, contextos de uso com sua fonte; equivalentes e seus contextos e remissivas (para sinônimos, por exemplo).

* Repertório terminológico disponível *on-line* que contém um número considerável de entradas; pode conter os mesmos campos de um dicionário, mas, o acesso às informações é facilitado por *hiperlinks*.

⁵ Em tradições como a inglesa e a alemã, utiliza-se também a forma Lexicografia Especializada.

São exemplos os dicionários produzidos pelo grupo Termisul: *Dicionário de Direito Ambiental: a terminologia das leis do meio ambiente* (Krieger et al., 1998), *Glossário de gestão ambiental* (Krieger et al., 2006) e a *Base de Dados Legis*⁶.

Breve panorama histórico da Terminologia

No âmbito da Terminologia, ao longo dos anos, foram se desenvolvendo várias propostas teóricas em decorrência de contextos sócio-históricos específicos. Inicialmente, surgiram teorias mais atreladas a uma visão positivista que defendia o objetivismo da ciência. Posteriormente, surgem perspectivas que seguem uma visão mais textual e comunicativa. Isso ocorreu, em certa medida, pela ampliação do conceito de ciência e do que está compreendido no fazer científico. Consequentemente, passou-se a atribuir um caráter científico a áreas até então não reconhecidas como “ciência” (por exemplo, as Ciências Humanas e Sociais). Além disso, a emergência de novas áreas como a Linguística Textual, a Sociolinguística e a Semiótica permitiu redimensionar várias noções importantes nos estudos linguísticos, como as de *língua*, *linguagem*, *texto* e *variação*. Ganha importância não só a análise dos textos em si, mas também os sujeitos pertencentes às diferentes comunidades de saberes, seus modos de se expressar e produzir conhecimento especializado.

Essas mudanças requerem ainda um olhar mais amplo para as condições de produção e divulgação de conhecimento e, em consequência, do fazer terminográfico, pois é fundamental considerar, além das especificidades das comunicações especializadas e do perfil dos destinatários dos produtos terminográficos, também suas necessidades. Esses fatos fortemente correlacionados geraram e geram demandas e necessidades terminológicas diversas que orientam, conforme Krieger (2000, p. 211), “o manejo social dos léxicos terminológicos”.

6 Disponível em: <http://www.ufrgs.br/termisul/cles/>

Em relação às propostas teóricas, a Teoria Geral da Terminologia (TGT), criada por Eugen Wüster⁷ nos anos 1960, é considerada a proposta fundadora da Terminologia⁸, dando origem à Escola de Viena de Terminologia. Essa proposta refletia uma concepção positivista de ciência em que a linguagem científica deveria ser homogênea para poder expressar claramente as verdades científicas. Nessa visão, os termos não eram concebidos como signos linguísticos – conformados por significante e significado – que fazem parte das línguas naturais, mas eram considerados etiquetas ou denominações que davam nomes aos conceitos; caracterizavam-se mais como unidades de conhecimento e não como expressões linguísticas (Krieger, 2000, 2018), o que define a TGT com um enfoque mais cognitivo e menos linguístico. Defendia-se o ideal de univocidade, ou seja, a um termo correspondia um conceito, pretendendo-se que estes fossem homogêneos, universais e imutáveis em cada área do saber. Conformavam, assim, uma linguagem à parte das línguas naturais que visava à comunicação unívoca entre os especialistas dos diversos países. Tinha, portanto, um caráter **prescritivista e normatizador***, ou seja, buscava-se um uso controlado dos termos para assegurar a precisão na comunicação especializada.

No final dos anos 1990, a partir dos avanços ocorridos no desenvolvimento de novos conhecimentos científicos e tecnológicos, da consequente transmissão, aprendizagem e divulgação desses conhecimentos a um público cada vez mais amplo e da necessidade de partir dos textos e das especificidades dos discursos especializados para identificar e descrever os termos, surgem novas propostas teóricas. Dentre elas, destacamos a Socioterminologia (Gaudin, 1993), a Teoria Comunicativa da Terminologia (TCT)

7 A partir de sua proposta foi criado o Comitê Técnico 37 (CT-37), denominado Terminologia: princípios e coordenação, no âmbito da International Standard Organization (ISO). Hoje, o CT-37 denomina-se *Language and Terminology* e divide-se em vários subcomitês. Mais informações em: <https://www.iso.org/committee/48104.html>.

8 No mesmo período, havia pesquisadores que constituíram outras escolas como a Escola Russa (Lotte, Drenzen) e a Escola de Praga (Vancura, Kopecky e Coda).

* Ao longo do tempo, o conceito de *normalização* em Terminologia passou por mudanças. Anteriormente, a normalização era usada para indicar a relação unívoca entre um conceito e sua denominação, de modo a eliminar a variação denominativa e conceitual, caracterizando-se como uma perspectiva mais prescritiva, normatizadora. Em algumas áreas, essa abordagem prescritiva é necessária, como na Aviação e na Medicina. Alguns autores, como Faulstich (2006), indicam o uso de *normatização* para essa visão mais prescritiva e de *normalização* para os casos em que se busca harmonizar, tornar normais as diferentes variantes de um termo, constituindo em uma abordagem descritiva.

(Cabré, 1999), a Teoria Sociocognitiva da Terminologia (Temmerman, 2000) e, no âmbito do Termisul, a Terminologia Linguístico-Textual (TLT)⁹. Comentamos alguns aspectos gerais dessas propostas e nos detemos na TLT, posto que se origina no Grupo Termisul e sustenta todas as suas pesquisas e produtos terminográficos desde o início de suas atividades.

Essas propostas criticam vários aspectos da TGT, entre eles o enfoque nos conceitos e a descontextualização dos termos dos contextos de uso. Ao contrário, defendem o estudo dos termos em seus contextos, os textos especializados. Isso supõe atentar para os diversos aspectos implicados na comunicação especializada para poder explicar o funcionamento dos termos nesses contextos e determinar seu valor especializado. Em consequência, os termos passam a ser vistos como signos linguísticos, constituídos de significado e significante, e fazem parte das línguas naturais, estando sujeitos aos seus vários processos de formação e de variação, entre outros.

Em algumas propostas, como a TCT (Cabré, 1999), para explicar o caráter complexo e multifacetado do termo, defende-se que a Terminologia é interdisciplinar, isto é, se fundamenta na inter-relação de perspectivas linguísticas, semióticas, cognitivas e comunicativas referidas anteriormente. Desse modo, a perspectiva linguística busca identificar e descrever os termos dentro da linguagem natural no que se refere, por exemplo, a sua constituição, seus limites, seus significados nos textos em que são utilizados e questões relativas à sinonímia e à variação. A perspectiva semiótica permite incluir outros sistemas simbólicos na representação do conhecimento especializado, como os símbolos e fórmulas químicas ou físicas. A perspectiva cognitiva (do conhecimento) elucidada como conceitualizamos a realidade, como se estruturam e se relacionam esses conceitos, de modo a mostrar como a área se constitui conceitualmente. Finalmente, a perspectiva comunicativa explica os contextos de uso em que os termos são

9 Além dessas, há outras propostas, como a Terminologia de Marcos (Faber; Márquez; Vega, 2005), a Terminologia Cultural (Diki-Kidiri, 2007) e a Terminologia Textual (Bourigault; Slodzian, 1999).

utilizados; por exemplo, quem são os participantes da comunicação especializada, qual o seu nível de conhecimento sobre o tema, qual o meio em que ocorre a comunicação, qual a perspectiva abordada etc.

Há também perspectivas que buscam analisar as características linguísticas e discursivas dos textos especializados, pois elas auxiliam na identificação dos termos e das UFEs e explicam seu funcionamento como elementos constitutivos da linguagem. Este é o caso da TLT, que alia esses aspectos à sua prática de mais de 30 anos na elaboração de produtos terminográficos realizada no âmbito do Grupo Termisul, a qual será apresentada na seção a seguir.

A Terminologia Linguístico-Textual (TLT)

A proposta da TLT vem sendo elaborada principalmente por Krieger (1998, 2001, 2004, 2005, 2008, 2018) e por Krieger e Finatto (2004), com a colaboração de pesquisas de outros membros do grupo, como veremos adiante. Conforme as autoras,

Toda essa visão, que leva em consideração as relações entre o funcionamento da linguagem, as especificidades das comunicações especializadas e a gênese do estatuto terminológico de determinadas unidades lexicais, constitui a base dos fundamentos teóricos e metodológicos da Terminografia, que denominamos linguístico-textual (Krieger; Finatto, 2004, p. 57).

Destacamos dessa proposta os seguintes princípios sintetizados a partir da produção das autoras anteriormente citadas e de outras pesquisadoras do Termisul, além da experiência do grupo:

- a terminologia não só auxilia a organizar e representar os conceitos de uma área, mas “é também elemento constitutivo da produção do saber” (Krieger, 2000, p. 211);
- o texto, concebido como um todo de sentido, é o *habitat* natural dos termos e das UFEs e constitui-se como uma moldura cognitiva, na qual essas unidades se articulam;

– o valor especializado dos termos e das UFEs define-se e explica-se por seu uso nos textos especializados, produzidos em situações comunicativas específicas; decorre daí a importância da análise de seus elementos estruturais e discursivos: as condições de produção (por exemplo, finalidade, meio de divulgação, tema e nível de especialização); os sujeitos envolvidos na comunicação especializada (quem produz e quem lê); os diferentes gêneros em que ocorrem (artigo científico, resumos, relatórios, teses, dissertações, leis etc.); suas partes, a distribuição da informação em cada parte e suas características linguísticas específicas;

– as terminologias podem ser multidisciplinares, isto é, podem conformar-se pela junção de saberes de diferentes âmbitos, e híbridas, ou seja, revelam diferentes perspectivas que constituem esses saberes;

– seus objetos de estudo são os termos, a fraseologia especializada e as definições, pois elas projetam o conhecimento especializado;

– as terminologias fazem parte das línguas naturais e revelam igualmente fenômenos como variação e sinonímia;

– as obras terminográficas (dicionários, glossários, bases de dados) também são textos e, como tal, possuem determinadas regras e formas de organização;

– há uma relação intrínseca entre teoria, análise de dados observáveis e soluções metodológicas na elaboração de produtos terminográficos, o que atribui um caráter teórico também à Terminografia, na medida em que oferece elementos teórico-metodológicos para a elaboração de produtos terminográficos.

Conceitos fundamentais implicados na elaboração de obras terminográficas

Trazemos, a seguir, os conceitos dos objetos de estudos no âmbito da TLT e outros que também são necessários no momento de elaborar obras terminográficas e que serão retomados nos capítulos seguintes.

• **Termo:** signo linguístico que possui conteúdo especializado decorrente de seu uso e conformação em determinada comunidade de saber, portanto, remete aos conceitos de uma área e, por essa razão, sua natureza é designativo-denominativa com a função de representar e transmitir conhecimentos específicos das áreas especializadas, sejam elas científicas, técnicas ou tecnológicas ou ainda relacionadas a ofícios (Krieger; Finatto, 2004).

Características dos termos:

– São substantivos simples (*energia, bens, reciclagem*) ou, mais frequentemente, sintagmáticos ou complexos (*recursos naturais, recursos naturais não renováveis, bens culturais, bens culturais imateriais*). Embora menos frequentes, também podem ser verbos (*reciclar, higienizar*), adjetivos (*crônico, grave*) e advérbios (*ambientalmente, politicamente*).

– Podem estar constituídos por formantes greco-latinos (*desinfestação, acervo bibliográfico*) e estar representados por siglas (CIP – Controle Integrado de Pragas), acrônimos (Sigerco – Sistema de informações de Gerenciamento Costeiro) e fórmulas (CO₂ – dióxido de carbono).

– Podem ainda conter variação (*conservação ambiental e conservação do meio ambiente*), sinonímia (*washi e papel japonês*) e neologia (*deltacron*), conceitos explicados mais adiante.

– Podem também sofrer as confluências e/ou transposições do léxico geral para o especializado, como a terminologização e a desterminologização. No primeiro processo, se atribui valor terminológico a uma palavra da língua geral (por exemplo, água como um líquido que bebemos para matar a sede e que passou a ter um valor especializado no Direito Ambiental, pois é considerada um **recurso natural*** e um bem público). No segundo, elimina-se um dos traços que conferia valor especializado aos termos (exemplo, *paranoia* como termo da Psiquiatria para referir-se aos transtornos delirantes e como palavra utilizada cotidianamente para indicar a sensação de perseguição de uma pessoa).

* “Água é um recurso natural de disponibilidade limitada e dotado de valor econômico que, enquanto bem público de domínio do Estado, terá sua gestão definida através de uma Política de Recursos Hídricos nos termos desta lei” (Lei n. 10.350 de 30/12/1994, art. 1º, p. 1).

– Sua identificação deve considerar: a) a pertinência temática, isto é, o pertencimento de um termo a uma área temática com um valor que lhe é próprio; e b) a pertinência pragmática, determinada pela funcionalidade de um termo em determinado produto terminográfico, sendo, em geral, termos de outras áreas que contribuem para a compreensão mais ampla da terminologia repertoriada (Maciel, 2001; Krieger; Finatto, 2004).

• **Unidades Fraseológicas Especializadas (UFEs):** unidades sintagmáticas, formadas por dois ou mais elementos linguísticos que coocorrem frequentemente em função de restrições combinatórias decorrentes das convenções próprias das línguas, dos gêneros textuais e da comunidade de saber que compartilha a mesma especialidade. São, portanto, estáveis e apresentam certa fixação de ordem (por exemplo, em que o nome tem função sintática de objeto direto do verbo), bem como há uma afinidade semântica entre essas palavras. Caracterizam-se como modos de dizer próprios de uma área e fazem referência a ações e processos, no caso das colocações, ou possuem funções específicas nos gêneros textuais nos quais ocorrem, no caso das formulações. Por exemplo, nos textos legislativos, essas UFEs estabelecem relações entre as partes do texto ou entre textos, indicam definições etc. Ambos os tipos representam e transmitem conhecimento especializado da mesma forma que fazem os termos (Bevilacqua, 1996, 2004; Bevilacqua *et al.*, 2013).

No contexto de pesquisa do Termisul, identificamos dois tipos de UFEs¹⁰:

a) as colocações equivalem a sintagmas, ou seja, estão formadas por duas ou mais unidades léxicas. Podem estar formadas por verbo e termo (ex.: *indenizar danos*) ou por uma nominalização e um termo unidos por preposição (*reciclagem de resíduos perigosos*). As colocações denotam processos e ações próprios do âmbito em que são utilizadas. No grupo Termisul, são denominadas Unidades Fraseológicas Especializadas Eventivas, por

10 Exemplo dos dois tipos podem ser vistos na Base Legis, que está disponível em: <http://www.ufrgs.br/termisul/cles/> e na Base Papel, que está disponível em: <http://www.ufrgs.br/termisul/papel/>.

conterem ou derivarem de um verbo (*indenizar* → *indenização* e *reciclar* → *reciclagem*). Nessa proposta, denominamos os verbos e as nominalizações de **núcleos eventivos** (NE), e os termos, de **núcleos terminológicos** (NT).

b) as fórmulas, além de equivalerem a frases ou a partes delas, têm uma função pragmática específica no discurso especializado. Por exemplo, no âmbito jurídico, temos: *esta lei entra em vigor na data de sua publicação e para os efeitos da presente lei*. No primeiro caso, indica que uma lei passa a vigorar e precisa ser respeitada pelos cidadãos; no segundo, aponta para definições constantes na lei e objetos que passam a ser tutelados por ela, como vemos no exemplo seguinte: “**Para efeitos desta Lei**, definem-se: [...] ‘Imóvel Rural’, o prédio rústico, de área contínua qualquer que seja sua localização que se destina à exploração extrativa agrícola [...]” (Lei n. 4.504, de 20 de novembro de 1964).

• **Definição Terminológica (DT):** é um texto que explica os significados de termos ou de expressões de uma técnica, tecnologia ou ciência, ou seja, busca descrever as características dos termos ou expressões, o que os singulariza e, ao mesmo tempo, os relaciona a outros conceitos de uma área (Finatto, 2001, 2003).

A definição pode ser extraída dos próprios textos especializados ou construída pelos terminógrafos. Em ambos os casos, é preciso conhecer como se constitui a área em que se utiliza o termo a ser definido, como se constroem os textos nessa área e quais são os diferentes sujeitos que nela atuam e que podem ter concepções diferentes dos objetos e, portanto, podem construir diferentes definições para um mesmo termo, o que se constitui em variação conceitual. Devem-se levar em conta ainda o destinatário da obra e sua função para selecionar ou construir a definição que atenda, de forma mais adequada, às necessidades do usuário previsto. Exemplo:

Biorremediação: Tecnologia baseada no uso de microrganismos vivos de ocorrência natural como bactérias, leveduras e fungos empregada para degradar substâncias nocivas de solos e águas in situ, transformando-as em substâncias menos tóxicas ou inócuas. **Notas:**

Utilizada para a reabilitação de áreas contaminadas por hidrocarbonetos do petróleo, agrotóxicos, explosivos, entre outros (Krieger *et al.*, 2006, p. 29).

• **Variação terminológica:** engloba a variação denominativa e a variação conceitual. Na variação denominativa, ocorre alteração na forma, ou seja, alteração ortográfica, morfológica, sintática ou lexical para a mesma noção. Por exemplo, *Estação de Tratamento de Esgoto* e *ETE*; *proteção ambiental*, *proteção ao meio ambiente* e *proteção do meio ambiente*; *papel japonês* e *washi*. Já na variação conceitual, há alguma alteração no conteúdo, ou seja, há alteração no significado ou em algum traço semântico. Isso ocorre, conforme comentamos acima, em função das diferentes concepções dos sujeitos que atuam em determinado âmbito do conhecimento. Os termos *agrotóxico* e *defensivo agrícola* remetem à mesma substância química usada na agricultura, com a diferença de que o primeiro destaca seu efeito tóxico à saúde e ao meio ambiente, ao passo que o segundo evoca a defesa das plantações contra pragas (Lazzarin, 2017). Podemos mencionar ainda a variação conceitual referente a uma mesma área de conhecimento quando se utilizam termos distintos dependendo do grau de especialização do usuário. Por exemplo, *lixo*, empregado em contextos mais gerais e cotidianos, e *resíduo*, utilizado em contextos especializados. Assim, os diferentes termos remetem a diferentes modos de perceber a realidade.

• **Neologismo terminológico:** resultado do processo de criação de palavras novas ou que recebem novo significado decorrente das necessidades dos falantes de dada língua em contextos especializados (Alves, 2004). Pode ocorrer por diferentes processos fonológicos, morfológicos, sintáticos, semânticos. Como exemplo de mudança nos traços semânticos, temos o termo *banho*, que designa uma técnica para recuperação das propriedades do papel na área de conservação e restauração de bens em papel. Neste exemplo, *banho* que, na língua geral, significa “Ação ou resultado de molhar o corpo, ou parte dele, para fins de higiene, lazer ou terapêuticos” (*Aulete digital*), ganha novos traços em função do objeto a ser banhado, papel, e não mais o corpo, e de passar a ser uma técnica de uma área específica do

conhecimento e não mais uma ação ou resultado da vida cotidiana. Outro exemplo da mesma área é *foxing*, um empréstimo do inglês incorporado à língua portuguesa para fazer referência à mancha no papel.

• **Equivalência terminológica:** relação estabelecida entre dois ou mais termos que cumprem, em línguas e culturas diferentes, a mesma **função referencial***. Para isso, devem preencher as seguintes condições:

– pertencer à mesma área e temática do termo original; por exemplo, a área do meio ambiente pode se desmembrar em várias temáticas, como gestão ambiental, aquecimento global e preservação ambiental;

– ocorrer no mesmo gênero textual; por exemplo, tratados internacionais, como a **Agenda 21***, e relatórios, como o Relatório de Impacto Ambiental (**Rima**)*;

– ser utilizada no mesmo campo semântico: por exemplo, na preservação ambiental, podem ser campos semânticos as áreas de preservação (incluem termos como área de conservação ecológica, *área de preservação permanente*, *zona de conservação da vida silvestre*) ou os instrumentos de política ambiental (incluem termos como *programa nacional de florestas*, *sistema permanente da avaliação e controle dos agrotóxicos*);

– ser empregada pelos membros de uma mesma comunidade de saber (aspecto comunicativo). Por exemplo, os governos dos 179 países que assinaram a Agenda 21.

Trazemos um exemplo da UFE *combate à poluição* e seus equivalentes em diferentes línguas em seus contextos de uso. Nos contextos, pode-se ver a seleção léxica feita por cada língua para compor a UFE. Enquanto o espanhol e inglês utilizam *combate* e *to combat*, respectivamente, como o português, o francês utiliza *lutte* (luta). O

* A função referencial consiste na “referência que remete a objetos e fenômenos do mundo real, ou mesmo fictício, imaginário” (Nord, 2018, p. 53). Por exemplo, baleia azul faz referência a um animal que de fato existe e que possui determinadas características (tem barbatanas, mede até 39 metros, pesa cerca de 160 toneladas, vive no mar etc.); em contraposição, unicórnio faz referência a um ser mitológico, imaginário, que tem apenas um chifre.

* “Documento elaborado na Conferência do Rio que se constitui no plano de ação a ser implementado pelos governos, agências de desenvolvimento, organizações das Nações Unidas e grupos setoriais independentes em cada área em que a atividade humana afeta o meio ambiente com o objetivo de proteger a natureza e suas riquezas para as gerações futuras no século 21” (Krieger et al., 2008, p. 7).

* “Documento que visa a avaliar as interações da implantação ou da operação de uma atividade real ou potencialmente poluidora com o meio ambiente” (Krieger et al., 2008, p. 2.005).

alemão utiliza *Bekämpfung* (combate/control). Enquanto francês e inglês usam *pollution*, como o português (*poluição*), o espanhol utiliza *contaminación*. Já o alemão utiliza *Verschmutzung*, que, dependendo do contexto, pode ser equivalente a *poluição* ou *contaminação*. Vejamos os exemplos¹¹:

Português: O enquadramento dos corpos de água em classes, segundo os usos preponderantes da água, visa a: [...] diminuir os custos de *combate à poluição* das águas, mediante ações preventivas permanentes. (Lei n. 9.433, de 8 de janeiro de 1997; br).

Alemão: Dem Antrag sind insbesondere folgende Angaben und Unterlagen beizufügen: [...] Maßnahmen der Bau- und der Betriebsphase einschließlich der vorgesehenen Maßnahmen zur Verhütung und *Bekämpfung von Verschmutzungen* sowie der Kontroll- und Überwachungsmaßnahmen, [...]. (Deponieverordnung – DepV – vom 24. Juli 2002)

Espanhol: El presente reglamento establece el régimen de prevención, vigilancia y *combate de la contaminación* en las aguas de mar, puertos, ríos y lagos sometidos a la jurisdicción nacional (Reclamo para el control de la contaminación acuática, Ministério de Defensa, Chile, 06 jan. 1992).

Francês: Au niveau local, une cellule d'experts est constituée auprès de l'autorité principalement en charge de la *lutte contre la pollution* ou auprès du préfet de zone de défense, si celui-ci assure la coordination de la gestion de la crise. (Instruction du 4 mars 2002 relative à la lutte contre la pollution du milieu marin, JORF n°79 du 4 avril 2002).

Inglês: States are bound to prevent and control marine pollution and are liable for damage caused by violation of their international obligations to *combat such pollution*; [...] (United Nations Convention on the Law of the Sea of 10 December 1982).

¹¹ Os dados estão disponíveis na Base Legis: http://www.ufrgs.br/termisul/lib_bases/php/view_entry.php?entry=891&action=view.

Para concluir este capítulo, reiteramos o papel dos produtos terminográficos na comunicação especializada e na organização e divulgação do conhecimento produzido pelas comunidades de saber. Krieger (1998, p. 29) destaca que

A peculiaridade das terminologias, circunscrevendo conceituações nos mais diferentes campos do conhecimento científico e tecnológico, evidencia seu papel na constituição e transmissão dos saberes humanos. A essa funcionalidade primeira soma-se a importância social e política, e mesmo estratégica, da organização e divulgação das terminologias para os contextos de integração, como o Mercosul (Krieger, 1998, p. 29).

A síntese teórica e as informações aqui apresentadas serão retomadas e aprofundadas nos próximos capítulos. É neles que se verá, mais claramente, a relação entre a teoria aqui apresentada e a prática.

PARA SABER MAIS

Terminologia e Terminografia

ALVES, Ieda Maria; MARONEZE, Bruno. Neologia: histórico e perspectivas. *Revista GTLex*, Uberlândia, v. 4, n. 1, p. 5-32, jul./dez. 2018. Disponível em: <https://seer.ufu.br/index.php/GTLex/article/view/55082/28994>. Acesso em: 5 maio 2022.

ARNTZ, Reiner; PICHT, Heribert. *Introducción a la Terminología*. Tradução do alemão de Amelia de Irazazábal *et al.* Madrid: Fundación Germán Sánchez Ruipérez/Pirâmide, 1995.

BEVILACQUA, Cleci Regina; KILIAN, Cristiane Krause. Tradução e Terminologia: relações necessárias e a formação do tradutor. *Domínios de Linguagem*, v. 11, p. 1.707-1.726, 2017. Disponível em: <https://seer.ufu.br/index.php/dominioidelinguagem/article/view/41124/21749>. Acesso em: 16 jun. 2022.

BEVILACQUA, Cleci Regina; REUILLARD, Patrícia C. R. Questões de lexicografia, terminologia e tradução. *Cadernos de Tradução*, Porto Alegre, n. 20, jan./jun. 2007.

CABRÉ, Maria Teresa. Terminología y lingüística: la teoría de las puertas abiertas. *Estudios de Lingüística del Español*, [en línea], 2002, v. 16. Disponível em: <https://raco.cat/index.php/Elies/article/view/195486> Acesso em: 8 abr. 2022.

FREIXA, Judit. *La variació terminològica: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambiente*. Tese (Doutorado em Lingüística Aplicada) – Instituto Universitário de Lingüística Aplicada, Universidade Pompeu Fabra, Barcelona 2002.

KRIEGER, Maria da Graça; ARAÚJO, Luzia. A terminologia em foco. *Cadernos de Tradução*, Porto Alegre, n. 17, out./dez. 2004.

REY, Alain. *La terminologie: noms et notions*. Paris: PUF, 1979.

SAGER, Juan Carlos. *Curso práctico sobre el procesamiento de la terminología*. Tradução do inglês de Laura Chumillas Moya. Madrid: Fundación Germán Sánchez Ruipérez, Pirámide, 1990.

TEMMERMAN, Rita. *Towards new ways of terminology description*. Amsterdam: John Benjamins, 2000.

WÜSTER, Eugene. *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. Tradução do alemão de Anne-Cécile Nokerman. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, 1998.

Frasesologia especializada

L'HOMME, Marie-Claude. Understanding Specialized Lexical Combinations. *Terminology*, v. 6, n. 1, p. 89-110, 2000.

GOUADEC, Daniel. Nature et traitement des entités phraséologiques. In: *Terminologie et phraséologie. Acteurs et aménageurs: Actes de la deuxième Université d'Automne en Terminologie*. Paris: La Maison du Dictionnaire, 1994, p. 167-193.

GRANGER, Sylvie; PAQUOT, Magali. Disentangling the phraseological web. In: GRANGER, Sylvie; MEUNIER, Fanny. *Phraseology: An Interdisciplinary Perspective*. [S.l.]: Benjamins and Philadelphia, 2008. p. 27-49.

MEL'CUK, Igor. *Phraseology in the language, in the dictionary, and in the computer*. Disponível em: http://olst.ling.umontreal.ca/pdf/yop_2012_0003.final.pdf. Acesso em: 18 jun. 2021.

Linguagens e textos especializados

CIAPUSCIO, Guiomar. *Textos especializados y terminologia*. Barcelona: Universitat Pompeu Fabra; Instituto Universitario de Linguística Aplicada, 2003.

FINATTO, Maria José B.; ZILIO, Leonardo (org.). *Textos e termos por Lothar Hoffmann*. Porto Alegre: Palotti, 2015.

KOCOUREK, Rostislav. Textes et termes. *Meta*, v. 36, n. 1, p. 71-76, mars 1991. Numéro Spécial. La Terminologie dans le monde: orientations et recherches. Disponível em: <https://www.erudit.org/fr/revues/meta/1991-v36-n1-meta331/003330ar.pdf>. Acesso em: 10 mar. 2022.

LERAT, Pierre. *Las lenguas especializadas*. Barcelona: Editorial Ariel, 1997.

Capítulo 2 – As decisões prévias

Cristiane Krause Kilian
Patrícia Chittoni Ramos Reuillard

A elaboração de um produto terminográfico – glossário, dicionário, base de dados – requer do terminólogo uma série de decisões prévias. Como vimos no capítulo anterior, tais decisões dizem respeito tanto às bases teóricas da elaboração do produto – a(s) corrente(s) da Terminologia que vão sustentar as escolhas do terminólogo – quanto aos seus aspectos práticos, ou seja, ao tipo de produto terminográfico que ele deseja oferecer ao seu público-alvo considerando suas necessidades.

Um glossário que pretenda ser prescritivo, por exemplo, vai buscar sustentação na Teoria Geral da Terminologia (TGT), pois sua intenção é indicar “como devem ser” os termos de uma determinada área¹, ao passo que a Teoria Comunicativa da Terminologia (TCT) poderá nortear um produto que objetive apenas registrar “como são usados” os termos². Neste último, entre outros aspectos, a microestrutura incluirá um campo para a variação terminológica – denominativa ou conceitual –, ao passo que, no primeiro, a prescrição limitará o escopo do produto a um único termo para cada conceito, sem levar em conta os aspectos comunicativos.

As decisões prévias abrangem vários aspectos e podem ser agrupadas em quatro grandes grupos: **conteúdo** (área e temática), **objetivos** (finalidade, usuários, *corpus* de coleta e tipo de produto), **equipe de trabalho** (profissionais envolvidos) e **recursos** (financeiros, programas e ferramentas informáticas, bases de dados). Detalharemos a seguir cada um deles.

1 Os dicionários produzidos pelas Academias de Letras têm esse caráter, ao indicar a norma que a comunidade deve aceitar.

2 A título de ilustração, ver a base da **Linguagem do Patrimônio Cultural Brasileiro: Conservação dos Bens Culturais Móveis (Base Papel)**, que está disponível em <http://www.ufrgs.br/terminul/papel/>.

Conteúdo: área e temática

A escolha de uma área para a elaboração de um produto terminográfico exige, antes de mais nada, uma análise da bibliografia existente a fim de verificar a necessidade real de sua produção, que pode ser atestada pela inexistência ou pelo esgotamento de obra semelhante, ou pela limitação ou antiguidade dos produtos em circulação. Assim como áreas já consolidadas demandam atualizações periódicas, também as novas carecem, muitas vezes, de um levantamento terminológico que possa auxiliar sua estabilização. Escolhida a área, é preciso verificar sua conformação, que pode ser representada:

- a. por seu caráter interdisciplinar: por exemplo, um estudo na área da Educação certamente abrangerá intersecções com a Psicologia e a História, que contribuirão com sua própria terminologia para o conjunto global dos termos elencados;
- b. pelas influências linguísticas ou culturais que ela sofre: uma área incipiente, como era a informática em meados do século XX, no Brasil, recebeu um grande aporte linguístico do inglês, via estrangeirismos, e isso teve consequências terminológicas; do mesmo modo, alguns países são mais desenvolvidos do que outros em determinados campos e contribuem, conseqüentemente, com mais material bibliográfico, caso das pesquisas brasileiras sobre doenças tropicais³;
- c. pela documentação tanto institucional quanto administrativa e acadêmica, além de sua disponibilização: em alguns países, a pesquisa é bastante desenvolvida, mas o acesso à produção textual é apenas físico, disponibilizado em bibliotecas, e não é gratuito;
- d. pela influência das diretrizes internacionais sobre a área: muitas organizações internacionais, como a ONU, são guardiãs de documentos

³ A esse respeito, ver, por exemplo, <https://www12.senado.leg.br/noticias/materias/2019/10/23/brasil-tem-potencial-para-pesquisa-de-ponta-em-saude-mundial-aponta-debate>. Acesso em: 14 out. 2021.

- referenciais em algumas línguas, o que pode influir na produção de conhecimento dos demais países membros⁴;
- e. por suas particularidades: por vezes, uma única cultura é detentora de determinado saber ou prática e, sobre eles, tem uma terminologia estabelecida⁵;
 - f. pela estabilidade conceitual e terminológica (ou não) da área na(s) língua(s) do levantamento terminológico em questão.

Para melhor compreender a estrutura de uma área, é útil elaborar modelos de organização, como árvores de domínio, mapas conceituais ou fluxogramas⁶.

Uma árvore de domínio representa as relações existentes entre os conceitos, apresentando sua hierarquia e seus termos-chave, que podem, por sua vez, desdobrar-se em outros termos. Trata-se de um passo importante na metodologia do trabalho terminológico, visto que a árvore auxilia na confirmação posterior do pertencimento dos candidatos à área estudada. A figura 2.1 o exemplifica com a árvore de domínio do licenciamento ambiental brasileiro, proposta por Chichorro (2016)⁷:

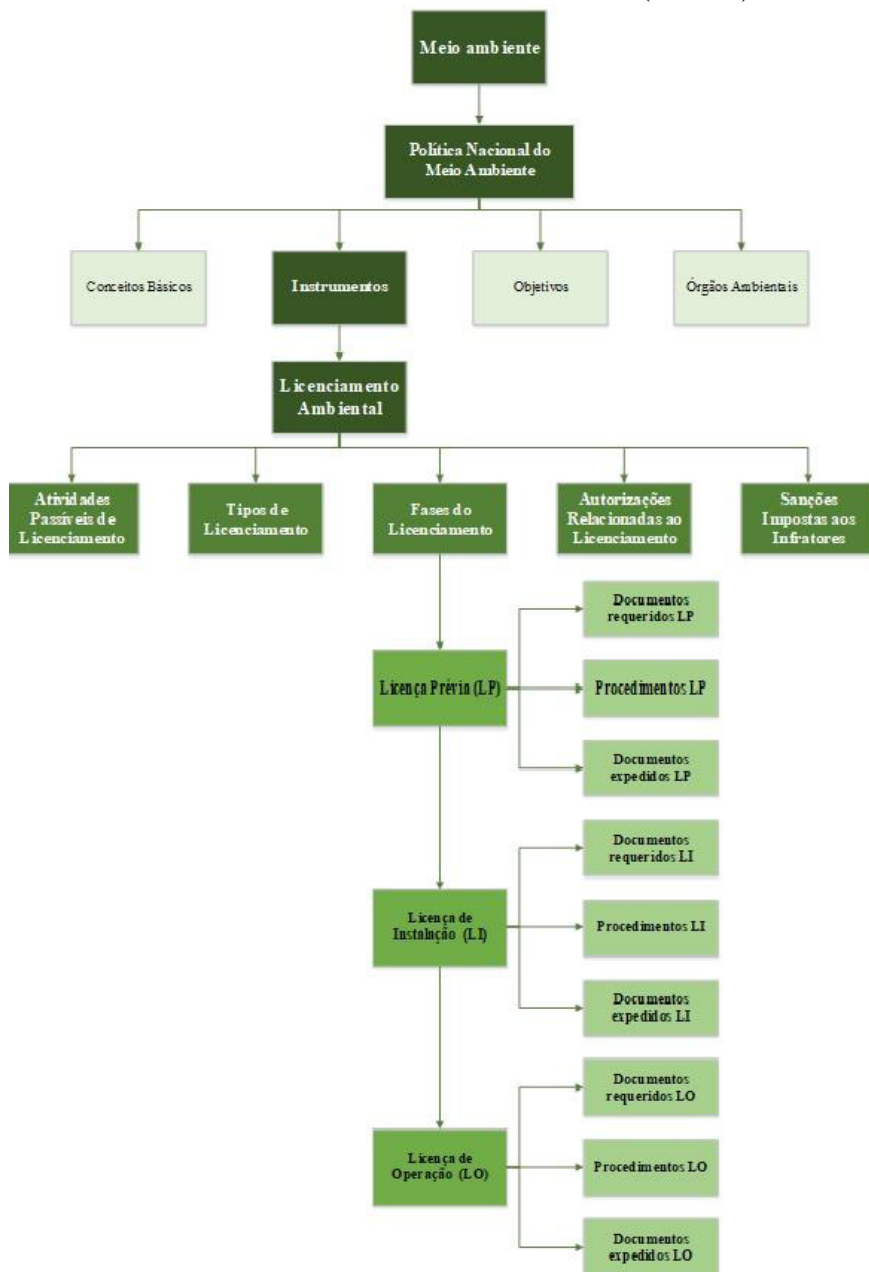
4 As línguas oficiais na ONU são o inglês, o francês, o mandarim, o espanhol, o árabe e o russo.

5 Isso pode ser observado na terminologia do patrimônio imaterial: cada cultura tem práticas culturais específicas, que se consolidam em um vocabulário não compartilhado com outras comunidades. A título de exemplo, veja-se o Bará do Mercado Público de Porto Alegre: <http://www.ipatrimonio.org/porto-alegre-mercado-publico-central/>. Acesso em: 14 out. 2021.

6 Existem programas que auxiliam na elaboração desses modelos. Entre eles, CMap Tools: <https://cmap.ihmc.us/>, Mindup: <https://www.mindmup.com/> e TheBrain: <https://www.thebrain.com/>

7 Disponível em: <https://lume.ufrgs.br/handle/10183/143111?locale-attribute=en>. Acesso em: 27 fev. 2022.

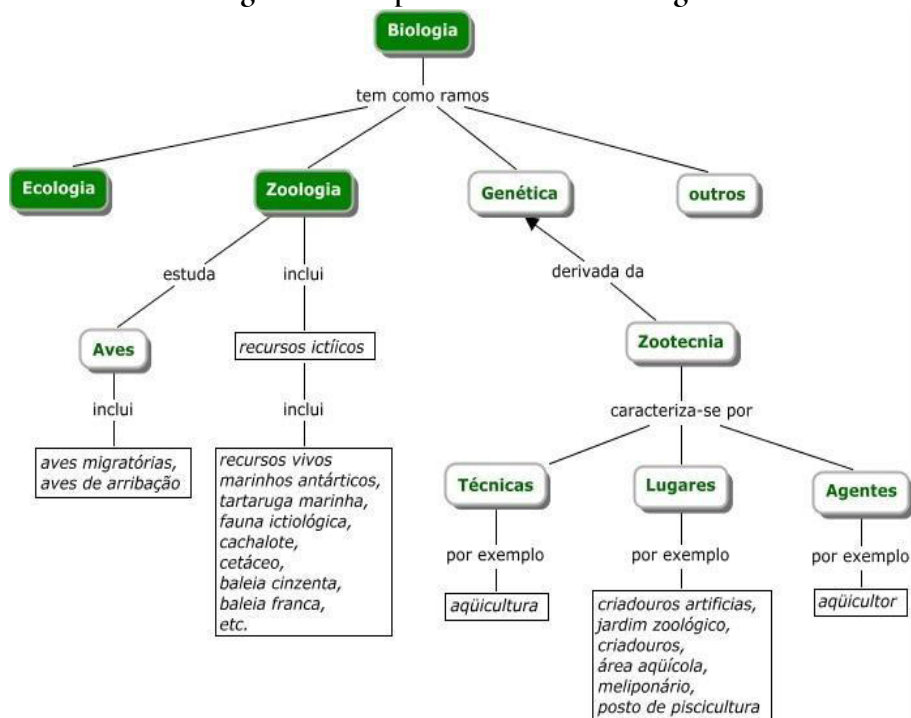
Figura 2.1: Árvore de domínio do Licenciamento Ambiental Brasileiro (recorte)



Fonte: Chichorro, 2016, p. 27.

Com o intuito de abordar de modo sistemático uma área de estudo, também o **mapa conceitual** pode se revelar de grande utilidade. Trata-se de “representações gráficas em forma de diagrama em que os termos referentes aos conceitos ocupam polígonos ou círculos ligados por vetores que identificam seus inter-relacionamentos” (Bevilacqua *et al.*, 2009, p. 817). O mapa permite distinguir os termos da área e suas relações com os termos conexos a ela. Observe-se, a título de ilustração, o mapa conceitual da Biologia, elaborado pelo Grupo Termisul⁸:

Figura 2.2: Mapa conceitual da Biologia



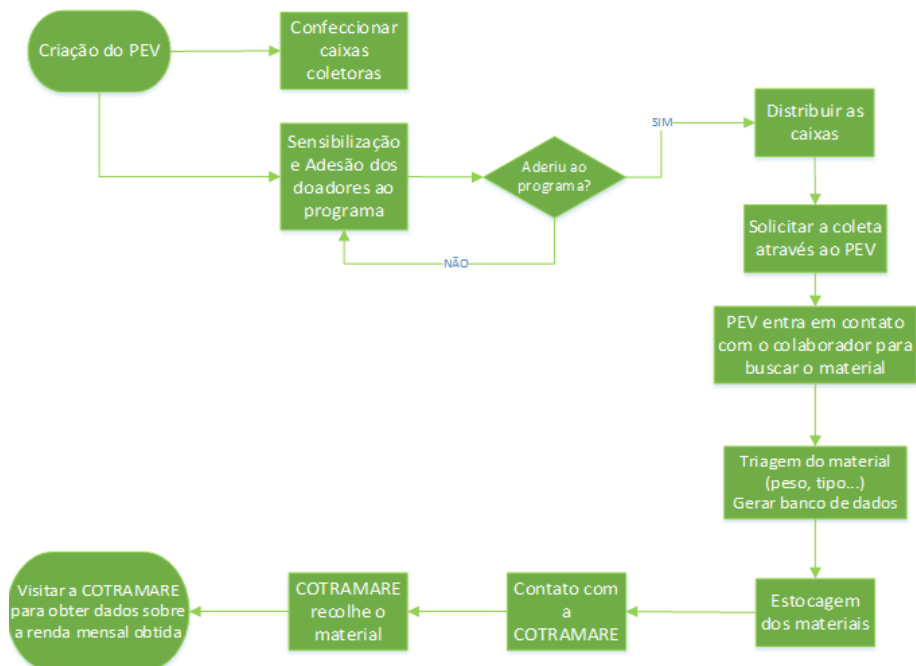
Fonte: Grupo Termisul.

Por fim, também é possível produzir fluxogramas, que indicam passos a serem seguidos, sobretudo quando a área envolve procedimentos de

8 Disponível em <http://www.ufrgs.br/termisul/mapasConceituais.php>.

trabalho. A título de ilustração, segue o fluxograma de procedimentos de coleta de resíduos da Universidade Federal de Campina Grande:

Figura 2.3: Fluxograma de procedimentos de coleta



Fonte: Laboratório de Tecnologias Agroambientais.⁹

A partir dessas informações e da distribuição da área, o terminólogo poderá proceder ao recorte da **temática** para definir o produto terminológico.

Por exemplo, a área do Meio Ambiente pode dividir-se em várias subáreas – Engenharia Ambiental, Ecologia, Direito Ambiental, Desenvolvimento Sustentável, Políticas Públicas, entre muitas outras – que, por sua vez, podem desdobrar-se em incontáveis temáticas, como o

⁹ Disponível em: <http://www.ltablocobx.com.br/2018/09/fluxograma-do-procedimento-de-coleta.html>. Acesso em: 14 out. 2021.

licenciamento ambiental¹⁰, o direito dos tratados internacionais do meio ambiente¹¹ e a gestão ambiental¹².

A temática escolhida dependerá, igualmente, do grau de especialização do usuário pretendido (ver a seção “Equipe de trabalho: profissionais envolvidos” mais abaixo) e da perspectiva de tratamento do tema. Por exemplo, um dicionário sobre o meio ambiente pode, entre outras possibilidades, abordar tanto os aspectos legais da área, apresentando os termos da legislação atinente, quanto a terminologia ambiental propriamente dita, oferecendo termos de áreas como Botânica e Zoologia.

A coleta de todas essas informações sobre a área e sua estruturação, bem como a temática do produto terminográfico almejado, fundamentarão as escolhas futuras do terminólogo: a finalidade, o usuário, o tipo da obra e o *corpus* de coleta. É o que veremos a seguir.

Objetivos: finalidade, usuários, *corpus* de coleta e tipo de produto

Uma série de questionamentos deve ser feita previamente pelo terminólogo a fim de circunscrever a finalidade da obra terminográfica que pretende elaborar. Eles podem ser resumidos pelas seguintes perguntas: para quê e por quê? Para quem? Que tipo de produto?

Primeiramente, é necessário decidir a **função** e a **razão de ser** desse produto: ele auxiliará no estabelecimento de uma nova área ou na consolidação de uma área preexistente? Em outras palavras, buscará sanar uma carência terminológica, apresentando, por exemplo, um glossário em uma área emergente para auxiliar aprendizes, como estudantes de graduação? Ou será um glossário com equivalentes em várias línguas, direcionado a tradutores? Uma ilustração disso é o glossário dos bens culturais móveis, produzido pelo grupo de pesquisa Termisul¹³, que vem preencher uma la-

10 Ver, por exemplo, Chichorro (2016).

11 Ver, por exemplo, Krieger *et al.* (2004).

12 Ver, por exemplo, Krieger *et al.* (2006).

13 A linguagem do patrimônio cultural brasileiro: conservação dos bens culturais móveis. Disponível em <http://www.ufrgs.br/termisul/papel/index.php>. Acesso em: 27 fev. 2022.

cuna em português e auxiliar os tradutores ou redatores em línguas estrangeiras ao oferecer equivalentes em várias línguas. Conforme a função escolhida pelo terminólogo, a apresentação dos termos ou das fraseologias da área demandará soluções distintas.

Em segundo lugar, importa definir os **usuários** preferenciais da obra terminográfica, visto que cada usuário tem necessidades diferentes de consulta: redatores técnicos, tradutores, assessores linguísticos, pesquisadores, aprendizes. Isso se faz necessário porque as demandas variam conforme os consulentes; por exemplo, um tradutor nem sempre precisa do campo classe gramatical, ao passo que uma definição terminológica, mesmo sucinta, lhe é muito útil para compreender um termo em seu contexto de uso; um glossário para especialistas pode prescindir, conforme o caso, do campo definição, enquanto esse campo é primordial para aprendizes de uma área. Mais uma vez, portanto, o terminólogo deverá responder a que usuário se destina a obra, especialista ou não especialista. No caso do primeiro, deverá avaliar quais informações são pertinentes para ele, de modo a não oferecer nem informação em excesso, subestimando seus conhecimentos, nem informação insuficiente, abstraindo suas necessidades; no caso do não especialista, a qualidade das informações fornecidas pode fazer a diferença na remediação de suas lacunas de conhecimento. As respostas a tais perguntas levarão a escolhas distintas referentes à quantidade de informações fornecidas ou aos campos do glossário ou da base de dados.

Outra decisão prévia diz respeito ao **tipo** de obra terminográfica. Glossário, dicionário ou base de dados podem apresentar-se de diferentes formas conforme o público preferencial a que se destinam: impressos ou virtuais, podem ser monolíngues, bi ou trilíngues, multilíngues, ou monolíngues com equivalentes em uma ou mais línguas¹⁴. A caracterização de cada tipo de obra foi apresentada no capítulo 1.

Uma vez delimitadas a área e a temática e as características do produto final, assim como seus usuários, procede-se à construção do *corpus*

14 Ver, por exemplo, Krieger, 2006. 127 p.; ver também a base de dados *Combinatórias léxicas especializadas da linguagem legal*, disponível em: <http://www.ufrgs.br/terminul/cles/>. Acesso em: 27 fev. 2022.

que servirá de fonte de levantamento dos termos da área. *Corpus* é um conjunto de textos autênticos, representativos da variedade ou do uso linguístico que se pretende estudar, selecionados segundo critérios linguísticos e sistematizados de tal forma que possam ser processados por computador. As características de um *corpus* podem variar muito¹⁵, mas importa que o *corpus* construído seja compatível com os objetivos da pesquisa. Poderá ser **comparável**, com textos semelhantes nas línguas envolvidas, ou **paralelo**, ou seja, com textos em dada língua e suas traduções; monolíngue, que oferece a terminologia em uma única língua, ou multilíngue, em língua de partida e língua(s) estrangeira(s). Essa construção vai requerer uma metodologia e critérios de coleta preestabelecidos pelo terminólogo. Por exemplo, em um glossário para especialistas, cabe recolher gêneros textuais acadêmicos, ao passo que um dicionário para amadores de futebol poderá reunir diferentes gêneros jornalísticos, escritos e orais.

Na esteira de Cabré (1993, p. 298-299), defendemos que o *corpus* deve ser pertinente, isto é, representativo da área e, se possível, redigido por um autor qualificado; completo, incluindo todos os aspectos relacionados ao tema de trabalho; atualizado, de modo que a terminologia recolhida reflita a realidade linguística presente no âmbito em questão; e original, quer dizer, escrito por especialistas em sua língua materna, não traduzido.

Equipe de trabalho: profissionais envolvidos

Uma avaliação ingênua poderia levar a pensar que, para elaborar um produto terminográfico, basta que o terminólogo disponha de conhecimentos teóricos sobre Terminologia e Terminografia e um certo domínio da área que pretende repertoriar. De fato, o conjunto desses conhecimentos pode ser o bastante para dar o pontapé inicial no trabalho, mas logo se revelará insuficiente para suprir as lacunas que surgirão ao longo do levantamento terminológico no que diz respeito ao conteúdo. Essas lacunas ocorrem sobretudo em situações de flutuação das denominações, como se

15 Para mais detalhes sobre as propriedades de *corpora*, ver Berber Sardinha (2004).

pode ver no exemplo seguinte: na área da conservação do patrimônio dos bens em papel, empregam-se as unidades sintagmáticas *degradação do papel* e *deterioração do papel* como sinônimos. No entanto, quando empregadas conjuntamente, têm significados distintos¹⁶, só perceptíveis pelo especialista da área. Isso significa que um levantamento terminológico confiável não pode prescindir da presença de um especialista na área em questão, que poderá contribuir com seu conhecimento aprofundado para dirimir as dúvidas que surgirão ao longo do processo, tais como imprecisões ou flutuações conceituais e variação conceitual ou denominativa.

Por essa razão, uma **equipe** bem formada requer não só a presença de terminólogos e pesquisadores da linguagem, mas também a contribuição de um ou mais especialistas da área de pesquisa. Cabe ainda avaliar a pertinência de um projeto interinstitucional, que pode contribuir com diferentes olhares para o mesmo objeto, sem se esquecer dos especialistas em informática e dos estudantes bolsistas de iniciação científica, cuja atuação é imprescindível para as pesquisas acadêmicas.

Recursos informáticos e financeiros

Um trabalho terminológico que se pretenda fiável deve se apoiar também em dois aspectos: a capacidade analítica do terminólogo – responsável pela organização do *corpus* de trabalho, pela metodologia empregada e pela etapa de análise dos resultados – e o levantamento de dados. Para isso, ele pode se servir de uma série de **recursos informáticos** que permitam a investigação linguística de um *corpus* e a extração de candidatos a termo ou UFEs, assim como informações que analisará posteriormente. A cada dia, novas ferramentas são criadas. Apenas a título de ilustração, apresentamos sucintamente a seguir o programa *Sketch Engine*, que oferece um conjunto de ferramentas para análise linguística e que, por ser gratuito por um período de tempo, pode ser empregado por qualquer pesquisador,

16 Segundo Bojanoski (2018), o processo de degradação é químico e é causado por acidez, hidrólise ácida ou oxidação, da tinta ou do papel, ao passo que o processo de deterioração é físico e é causado por agentes externos, forças físicas, roubo e vandalismo, fogo, água, pragas, poluentes, luz, temperatura e umidade incorretas e dissociação.

mesmo sem recursos financeiros para a pesquisa (para mais detalhes, ver a seção “Programas de extração” do Capítulo 3).

* *Corpus* que serve de comparação para o *corpus* de estudo e normalmente deve ter três a cinco vezes o seu tamanho (Tagnin, 2011).

* *Corpus* no qual se baseia a pesquisa a ser desenvolvida pelo pesquisador (Tagnin, 2011).

O gerenciador de *corpus* e de análise textual *Sketch Engine*, criado por Adam Kilgarriff e Pavel Rychly e desenvolvido pela *Lexical Computing Ltd*¹⁷, oferece, em uma plataforma virtual, uma quantidade expressiva de ferramentas: **corpus de referência***, com quase cem línguas, que permite o contraste entre o **corpus de estudo*** e um *corpus* de língua geral (ou referência) ou de uma área específica; *WordList* (lista de palavras que compõem o *corpus*, sua posição e frequência); *Collocates* (listas de palavras que coocorrem com determinada palavra); *KeyWords* (lista de palavras-chave: *single words* e *multi-words*); *Concord* (possibilita a geração dos contextos a partir da busca de uma palavra ou parte dela – radical ou desinência); e *Cluster* (permite a geração de expressões multipalavras a partir de uma palavra-chave de busca)¹⁸.

As ferramentas dos diferentes programas se assemelham em muitos aspectos, mas diferenciam-se principalmente pelos recursos avançados de cada uma delas, que serão empregados ou não conforme o direcionamento da pesquisa. Vale destacar que os dados são extraídos a partir de escolhas feitas pelo pesquisador, que deverá oportunamente avaliar o material coletado. Em outras palavras, o pesquisador estabelece o que buscar e direciona a extração de dados, e as ferramentas escolhidas oferecem resultados que devem ser analisados, em um processo de retroalimentação.

Por fim, não se pode negligenciar a necessidade de **recursos financeiros**, considerando que verbas implicam tanto a escolha entre um programa pago (com um número maior de ferramentas disponíveis) e um programa gratuito ou apenas na versão “demo”, quanto a contratação de profissionais de informática, essenciais para a elaboração e o acompanhamento de bases de dados.

17 Disponível em: <https://www.sketchengine.eu/>.

18 Para uma análise detalhada do programa *Sketch Engine*, recomendamos Fromm *et al.*, 2020.

ATIVIDADES: Pensando nas decisões prévias

Visite páginas como:

<https://www.congressonacional.leg.br/legislacao-e-publicacoes/glossario-legislativo>

<http://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/search/search?query=gloss%C3%A1rio>

https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/index_lex-voc.html

Nelas, escolha um glossário de termos. Procure estabelecer, a partir dos dados apresentados, sua proposta:

1. Quais são sua área e temática?
2. Qual é a finalidade do glossário?
3. Pode-se inferir dados sobre o *corpus*? Que tipos e gêneros textuais ele contém?
4. Como classificá-lo quanto ao tipo (monolíngue, bilíngue, multilíngue etc.)?
5. Que profissionais foram necessários para sua elaboração?
6. Quais os usuários previstos?
7. O produto explicita sua organização?
8. A oferta final de informação atende às necessidades do usuário previsto?

PARA SABER MAIS

BEVILACQUA, Cleci R. Investigación Sistemática en Terminología. In: ÁLVAREZ CATALÁ, Sara; BARITÉ, Mari. (org.). *Teoría y praxis en Terminología*. Montevideo: Ediciones Universitarias; Unidad de Comunicación de la Universidad de la República, 2017. p. 69-90. v. 1.

KRIEGER, Maria da Graça; FINATTO, Maria José B. *Introdução à Terminologia: teoria & prática*. São Paulo: Contexto, 2004.

Capítulo 3 – Constituição de *corpora*: critérios de coleta, limpeza e organização

Márcia Moura da Silva
Manuela Arcos Machado

Para se construir um *corpus*, é preciso seguir uma série de etapas e fazer uso de programas especiais para processá-lo. Neste capítulo, partindo de uma breve reflexão sobre o uso de *corpus* na pesquisa terminológica e na elaboração de material terminográfico, descrevemos os critérios de coleta, limpeza e organização de um *corpus* e apresentamos dois programas, *AntConc* e *Sketch Engine*, utilizados para processar os textos que compõem a base de dados do projeto do Grupo Termisul na área de *Conservação e Restauração de Bens Culturais Móveis em papel (Projeto Papel)*, realizado entre 2019 e 2021. Em uma primeira etapa (2016 a 2019), foram coletados os termos da área e, em uma segunda etapa (2019 a 2021), as UFEs. A elaboração dessa base contou com textos da área em português, espanhol, francês, inglês, italiano e russo. Alguns dos exemplos que trazemos para apoiar nosso texto foram extraídos dessa base.

Uso de *corpora* na elaboração de material terminográfico

Ao longo deste Manual, falaremos em *corpora* de estudo e de referência. *Corpus* é uma coletânea de textos em formato eletrônico, compilado segundo critérios específicos de acordo com o estudo que se pretende realizar. *Corpus* de estudo é o *corpus* no qual se baseia a pesquisa a ser desenvolvida pelo pesquisador. *Corpus* de referência é o que serve de comparação para o *corpus* de estudo e normalmente deve ter de três a cinco vezes o seu tamanho (Tagnin, 2011). Para as pesquisas terminológicas, em geral, o *corpus* de referência deve estar formado por gêneros textuais que representem a língua comum.

Como visto no Capítulo 2, o pesquisador pode trabalhar com dois tipos de *corpus* – o comparável e o paralelo¹. Ainda que existam *corpora* gigantescos, como é o caso do *Corpus of Contemporary American English* (Coca), que é o maior *corpus* de livre acesso do mundo (mais de um bilhão de palavras), o mais importante é construir um *corpus* que siga critérios bem definidos e que dê conta dos objetivos da pesquisa. Aluísio e Almeida (2006, p. 158-159) apresentam uma síntese dos principais critérios defendidos por diferentes teóricos da Linguística de *Corpus*: i) autenticidade (textos escritos em linguagem natural e também por falantes nativos); ii) representatividade (textos representativos da língua ou de uma variedade de língua que se queira investigar para que o *corpus* possa representar seu uso efetivo); iii) balanceamento (equilíbrio de gêneros discursivos, tipos de textos², títulos ou autores, desde que sejam adequados à investigação pretendida e que a escolha tenha sido feita de maneira criteriosa); iv) amostragem (amostras que incluam toda a variação linguística existente); v) diversidade (de gêneros, tipos de textos e sobretudo de tópicos, visto que a variação desses últimos afeta a frequência de muitas palavras); e vi) tamanho (adequado ao tipo de pesquisa e à metodologia adotada).

Entre as várias áreas do conhecimento em que o uso de *corpora* vem se consolidando, está a Terminologia. Maciel (2006, p. 1) aponta que a disciplina acompanhou outros ramos da Linguística ao adotar o uso dessa metodologia em suas pesquisas. Para a autora, a pesquisa terminológica baseada em *corpus* valoriza o contexto sociolinguístico do termo e conseqüentemente do texto enquanto “registro do evento comunicativo real”. Como observam Bojanoski, Michelon e Bevilacqua (2017), o texto se tornou, a partir dos anos 1990, o objeto central de análise da Terminologia. Pela posição que o texto ocupa hoje nas teorias de Terminologia, as autoras defendem que a elaboração de um *corpus* seguindo critérios bem definidos,

1 No site do projeto **Legis** do Grupo Termisul, é possível acessar a base de textos legislativos usados em *corpora* comparáveis, assim como alguns textos alinhados (em língua inglesa e suas traduções para a língua portuguesa, alemã, espanhola e francesa). Disponível em: <http://www.ufrgs.br/termisul/legis.php> (download > legislação ambiental; download > textos alinhados).

2 Ver Marcuschi (2005, 2008) para uma discussão aprofundada sobre gêneros e tipos textuais.

sobretudo no que diz respeito à representatividade, é fundamental ao trabalho terminológico.

Ainda que a Terminologia venha abraçando essa metodologia, Maciel (2006) adverte que conduzir uma pesquisa baseada em *corpus* não significa necessariamente seguir todos os princípios da Linguística de *Corpus*³. Segundo ela, uma das principais vantagens de se construir um *corpus* para um projeto terminológico é a possibilidade de se conduzir uma investigação empírica dos termos ou unidades fraseológicas em uma quantidade considerável de textos especializados. Mas, ainda que seja uma rica fonte de dados e que esses dados evidenciem fatos sobre o padrão de uso que poderiam permanecer imperceptíveis em amostragens menores, o *corpus* não é um “manancial exaustivamente completo” (Maciel, 2006, p. 5).

A autora compartilha com Leech (1991), um dos pioneiros no uso de *corpus* eletrônico do inglês britânico, a noção de que o linguista faz uso da intuição em sua pesquisa, mas vai além e acrescenta à intuição a habilidade de interpretação e o conhecimento do sistema da língua em estudo, seja como falante nativo, falante proficiente ou como linguista. De fato, esses elementos adicionais podem até mesmo indicar ao pesquisador quando é necessário sair dos limites de um *corpus*. No caso do projeto na área de *Conservação e Restauração de Bens Culturais Móveis em papel (Projeto Papel)*, por exemplo, nem sempre foi possível encontrar equivalentes de uma determinada UFE em todas as línguas em seus respectivos *corpora*. Assim, foi necessário interpretar essa ausência de equivalentes para podermos validá-la ou procurar equivalentes fora do *corpus*, em *sites* especializados, como pode ser visto no capítulo 6.

Crítérios de coleta, limpeza e organização

Como já mencionado, a construção de um *corpus* demanda critérios claros para sua adequação. A partir do momento que se definem o recorte da pesquisa (temática, tamanho do *corpus*, composição e número mínimo

³ Para saber mais sobre a Linguística de *Corpus* e seus avanços, ver McEnery e Hardie (2012); Tagnin (2011); Berber-Sardinha (2000, 2004); Baker (1995); Biber (1993); e Leech (1991).

de ocorrência do objeto a ser investigado) e o público-alvo do produto terminográfico⁴, o primeiro passo será identificar a produção de textos na área do estudo. Nesse sentido, seguindo a proposta de Aluísio e Almeida (2006), sugerimos que se procure equilibrar os gêneros discursivos e tipos de textos. Na construção do *corpus* do nosso projeto na área de *Conservação e Restauração de Bens Culturais Móveis (Projeto Papel)*, por exemplo, incluíram-se textos do gênero acadêmico, como livros, manuais, periódicos, trabalhos de conclusão de curso, dissertações, teses, anais de eventos, relatórios e boletins informativos de associações da área. Vale mencionar que, visto estar a área em situação de estruturação no Brasil, encontrou-se um número reduzido de material em língua portuguesa, tendo-se incluído material de áreas afins, como Biblioteconomia, Arquivologia e Museologia. (Bojanoski; Michelin; Bevilacqua, 2017). A seleção dos textos se deu pela busca das palavras-chave *documento*, *documentação*, *conservação*, *papel*, *patrimônio*, *preservação*, *restauração* e *restauo*, tendo-se tido o cuidado de escolher material que proviesse de fontes confiáveis. Nesse caso, textos acadêmicos, *sites* de universidades, instituições de pesquisa e periódicos reconhecidos na área foram considerados fontes possíveis de coleta desse material.

Uma vez que se tenham selecionado os textos que comporão o *corpus*, é preciso que eles sejam salvos e catalogados. Para que possam ser processados pelas ferramentas de maneira adequada, os textos devem ser salvos em formato <.txt>. Entretanto, diferentes programas pedem diferentes codificações. O *AntConc*⁵ por exemplo, requer que os textos sejam salvos em *UTF-8*, mas caso se esteja trabalhando com o programa *ParaConc*, eles devem ser salvos utilizando a codificação *ANSI*⁶. Cada texto é salvo com

4 Ver definição do usuário de material terminográfico no Capítulo 2.

5 Durante a pesquisa, a versão disponível do software era 3.2. Atualmente, a última versão (*AntConc* 4.2) já oferece a possibilidade de leitura de textos em formato .pdf.

6 Há programas específicos disponíveis gratuitamente na *web* para a conversão de textos em <.pdf> para o formato em <.txt> ou ANSI. Um exemplo é o *AntFileConverter*, do mesmo desenvolvedor do *AntConc*, disponível gratui-

* Sugerimos que esse código apresente as duas primeiras letras do idioma dos textos, duas ou três letras que identifiquem o projeto, seguidas da numeração sequencial. No nosso projeto, por exemplo, o código adotado foi ptPPx, para o qual pt correspondia à língua portuguesa, PP a Projeto Papel e x à numeração do texto. Da mesma forma para os *corpora* de outros idiomas, como espPPx (espanhol), frPPx (francês), ingPPx (inglês) etc.

* Outra alternativa para não recuperar ruídos – isto é, palavras ou estruturas que não são relevantes para a pesquisa, mas que são acusadas pela ferramenta como possíveis palavras-chave, apesar de não as serem – durante a extração de palavras-chave é o uso de *stoplists*. Essas listas, geralmente, são de palavras gramaticais, nomes próprios e de lugares. Como resultado, a ferramenta extrai palavras lexicais. Isto é, aquelas que são representativas da temática do *corpus* de estudo e que poderão ser candidatas a termos. Dependendo do objetivo da pesquisa, outras palavras podem ser acrescentadas às *stoplists*.

um **código** que acaba por se tornar a “identidade” desse texto. Feito isso, todos os textos coletados devem passar por um processo de limpeza em que são retirados todos os **elementos extratextuais***, como tabelas, gráficos e imagens, assim como qualquer informação que não seja significativa para o estudo, como agradecimentos, sumários, referências bibliográficas, notas de rodapé

e *links* externos. No caso do *corpus* em língua portuguesa de nosso projeto, 161 textos foram selecionados, totalizando 38.129 *types* (número de palavras diferentes) e cerca de 967.852 *tokens* (número total de palavras).

Em relação à etapa de catalogação⁷, ela é necessária para que se tenha um registro do material utilizado para a construção do *corpus*. Fica a critério do pesquisador se o público externo terá acesso aos catálogos ou se esses serão acessados somente pela equipe de pesquisa. O importante é que esse instrumento registre informações como autor, título, fonte, ano de publicação, gênero textual, código de identificação do texto e, se for o caso, o *link* onde o texto esteja disponível e a data de coleta. Essa catalogação pode ser feita em uma tabela em editor de textos (Word, Documentos Google ou outro) ou em uma planilha em <.xlsx> (Excel ou Planilhas Google). Um exemplo de catálogo do Projeto Papel é apresentada no quadro 3.1.

tamente no site: <https://www.laurenceanthony.net/software/antfileconverter/>.

7 Outra alternativa para controlar o registro dos textos que compõem o *corpus* de estudo é a criação de cabeçalhos (*labels*) com informações do texto. Essa informação, no momento da busca, pode ser excluída com os recursos das ferramentas.

Quadro 3.1 – informações para catalogação do corpus textual

Código	Referências	Disponível em	Acesso em
ptPP001	ROSAS, Fernanda Jenner; MENDES, Débora Assis. Resgatando Olin-da.	Seminário da ABRA-COR, 4, 1988, Grama-do. Anais – v. 1. [S.l]: [ABRACOR], 1988. p. 115-124. [comunicação evento]	
ptPP099	CORADI, Joana Paula; EGGERT-STEINDEL, Gisela. Técnicas básicas de conservação e preservação de acervos bibliográficos. Revista ACB, São José, V. 13, n. 2, jul-dez, 2008.	https://revista.acb.org.br/racb	05 jul. 2017

Fonte: As autoras.

Esses procedimentos são necessários para garantir a adequação da metodologia empregada. Uma vez concluídos, os textos estão prontos para serem processados pelos programas, que facilitam o trabalho do pesquisador, pois disponibilizam uma série de ferramentas que permitem identificar e extrair o objeto do estudo, produzir listas de palavras, buscar colocados (isto é, palavras que costumam aparecer combinadas ao termo pesquisado), verificar frequência dos termos etc., como veremos a seguir.

Programas de extração

A construção de um *corpus* dá ao pesquisador acesso, em um só instrumento, a um grande número de textos, a partir dos quais poderá verificar padrões de uso em uma determinada língua ou linguagem. Para que isso aconteça, é preciso que os textos que foram criteriosamente selecionados, limpos e catalogados sejam processados em programas especialmente criados para serem usados com *corpora*. Há no mercado uma série de

programas que fazem esse trabalho, como, entre outros, *WordSmith Tools*⁸, *AntConc*⁹, *Sketch Engine*¹⁰, *Unitext*¹¹, *TermoStat*¹², *ParaConc*¹³, *AntPConc*¹⁴, estes dois últimos para processamento de *corpora* paralelos. Descreveremos algumas das funcionalidades do *AntConc* e do *Sketch Engine*, dois dos programas usados para processar os *corpora* do nosso projeto.

AntConc e Sketch Engine

O *AntConc* é um programa de livre acesso, desenvolvido pelo linguista britânico Laurence Anthony, que pode ser baixado e instalado em qualquer computador sem haver necessidade de estar conectado à rede. Por ser um programa gratuito, ele é mais limitado que o *Sketch Engine*, mas tem uma interface amigável e intuitiva¹⁵, permitindo diferentes pesquisas por meio de suas ferramentas e índices estatísticos.

O *Sketch Engine* é um gerenciador de *corpora* e um programa de análise textual criado inicialmente em 2003 por Adam Kilgarriff e Pavel Rychly e desenvolvido pela *Lexical Computing Ltd*¹⁶. Por ser uma ferramenta *on-line*, ele está em constante atualização, oferecendo ao usuário *corpora* atualizados e ampliados rotineiramente. Para acessá-lo, é necessário criar uma conta através do site <https://www.sketchengine.eu/>. Um dos aspectos negativos dessa ferramenta é que ela não é totalmente gratuita, mas é possível criar uma conta temporária de 30 dias que permite acesso à maioria dos recursos. Nessa conta *trial*, é permitido carregar um *corpus* textual de

8 Disponível em: <https://www.lexically.net/wordsmith/>

9 Disponível em: <https://www.laurenceanthony.net/software/antconcl/>

10 Disponível em: <https://www.sketchengine.eu/>

11 Disponível em: <https://unitexgramlab.org/pt>

12 Disponível em: <http://termostat.ling.umontreal.ca/>

13 Disponível em: <https://paraconc.com/>

14 Disponível em: <https://www.laurenceanthony.net/software/antpconcl/>

15 Além das instruções de uso apresentadas na própria ferramenta (clique em “help”), há vários vídeos na internet em que o próprio Laurence Anthony dá o passo a passo para a execução das principais funções do *AntConc*. Ver em: <https://www.youtube.com/watch?v=9TsqFVrUYO0>

16 <https://www.lexicalcomputing.com/>

até 1 milhão de palavras. Para analisar *corpora* maiores, é preciso, obrigatoriamente, comprar uma licença.

Apesar disso, o *Sketch Engine* conta com uma ampla lista de aspectos positivos. Destacamos, entre eles, a possibilidade de **lematização*** automática do *corpus* inserido nele (à diferença do *AntConc*) e também

a oferta de *corpora* de referência em diferentes línguas, sem a necessidade de se criar ou de se ter disponível um *corpus* de referência para contrastar com o *corpus* de estudo (como ocorre com o *AntConc*).

Os *corpora* de referência de diferentes línguas ofertados pelo *Sketch Engine* contam com bilhões de palavras e são atualizados regularmente. Também são disponibilizados outros tipos de *corpora*, como alguns paralelos, com textos legislativos traduzidos de diferentes países da Europa, *corpora* especializados de diferentes temas, *corpora* diacrônicos, *corpora* de aprendizes e, inclusive, *corpora* orais.

Vale mencionar que, embora a ferramenta *AntConc* não execute uma lematização automática do *corpus* textual, há disponíveis extensões para o *software* que o fazem. Ainda, existe a possibilidade de lematizar manualmente o *corpus* em .txt. Essa tarefa, no entanto, exige um trabalho longo de codificação. Por fim, outra alternativa no *AntConc* é a possibilidade de realizar buscas com a forma truncada da palavra, utilizando um asterisco. Por exemplo, para identificar todas as ocorrências do verbo *registrar*, pode-se buscar *registr**, o que permitirá chegar às diferentes formas do verbo, como *registra*, *registrou*, *registrará* etc.¹⁷

Além das ferramentas que apresentamos, o *Sketch Engine* conta também com outras de diferentes funções, seja para tarefas de extração terminológica, de análises linguísticas para fins específicos (como ensino de línguas) e também para soluções tradutórias.

* A lematização consiste no processo de juntar todas as formas de uma palavra em sua forma canônica. Um *corpus* lematizado permite que, por apenas uma única busca, recuperem-se todas as formas conjugadas de um verbo, por exemplo. Assim, buscando-se o verbo “registrar”, a ferramenta recuperará todas as suas conjugações: registra, registrou, havia registrado, registrara, registraria, registrado, registrando etc. O mesmo aplica-se a substantivos e adjetivos com suas flexões de gênero e número.

17 Para saber mais sobre buscas de estruturas frasais complexas usando formas truncadas no *AntConc*, ver Arcos e Bevilacqua (2018).

Destacamos abaixo a função de algumas das principais ferramentas desses dois programas. Ambos oferecem ferramentas de busca por palavra (*concordance*); lista de palavras (*wordlist*); busca por expressões complexas (n-gramas); lista de palavras-chave (*keyword list*); colocados (*collocate* no *AntConc*; *Word Sketch* no *Sketch Engine*), porém, o *Sketch Engine* apresenta mais recursos, como é o caso do *Word Sketch difference* e o *parallel concordance*, possibilitando, assim, uma investigação mais minuciosa do *corpus*.

Concordance: permite que o pesquisador procure qualquer palavra no *corpus*. No *AntConc*, quando a palavra é digitada no campo de pesquisa, a ferramenta traz todas as ocorrências dessa palavra, que aparece em cor diferente para identificação imediata, juntamente com o contexto em que se insere (basta clicar em qualquer ocorrência da palavra para que o contexto seja ampliado). A figura 3.1 mostra o resultado da busca pelo termo *acervo* (em azul) com as três palavras à direita em cores diferentes (o número de palavras que acompanham o termo consultado pode ser ajustado, e elas podem também ser destacadas à esquerda do termo). A busca foi feita no *corpus* Papel, ou seja, o *corpus* do projeto.

Figura 3.1 – Concordâncias do termo *acervo* (*Corpus* Papel)

File	Left Context	Hit	Right Context
12 ptPP127.txt	nto e melhoria da qualidade nos serviços de preservação do	acervo	documental o Núcleo de Documentação da UFF. In: Anais...
13 ptPP157.txt	ncipais fatores ambientais que prejudicam e deterioram um	acervo	documental ou bibliográfico. Assim como os danos relaciona
14 ptPP090.txt	estabelecidas com universidades e instituições de guarda de	acervo	documental para viabilizar esta metodologia. Em 2009 o Dep
15 ptPP095.txt	vulgarmente como brocas ou carunchos. Na deterioração de	acervo	documental por insetos, outras famílias também causam dan
16 ptPP111.txt	arquivo refere-se tanto ao local de guarda de um	acervo	documental quanto ao próprio acervo em si. As duas
17 ptPP097.txt	uro. É parte das atribuições destas instituições preservar o	acervo	documental sob sua guarda, ao mesmo tempo em que
18 ptPP142.txt	03, p. 20–21). Sob esta perspectiva, o termo documento ou	acervo	documental tem em seu significado uma estreita relação cor
19 ptPP159.txt	il da UFSM. O desenvolvimento das ações de preservação do	acervo	documental vão ao encontro de um dos objetivos do
20 ptPP153.txt	: de transporte, embalagem e seguro. 9.10 Toda unidade do	acervo	a ser emprestada deverá ser conferida na sua saída
21 ptPP153.txt	: de transporte, embalagem e seguro. 9.10 Toda unidade do	acervo	a ser emprestada deverá ser conferida na sua saída
22 ptPP111.txt	idiciais a profissionais e pesquisadores que irão lidar com o	acervo	a ser tratado, além de não ter apresentado alterações
23 ptPP142.txt	proposta reduziria, em alguma percentagem, o universo do	acervo	a ser tratado, ao mesmo tempo em que o
24 ptPP112.txt	seja uma inundação ou qualquer outro motivo que leve o	acervo	a ser atingido por água, os primeiros procedimentos devem
25 ptPP148.txt	jeve estar adaptado para atender as especificidades de cada	acervo	a ser transportado e a sua localização. Usualmente o

Fonte: AntConc.

Wordlist: a ferramenta gera uma lista de todas as palavras do *corpus*, em ordem alfabética ou por frequência. Conforme a figura 3.2, nota-se, na lista de palavras gerada pelo *AntConc*, que nas primeiras posições aparecem artigos, preposições e conjunções que não são relevantes à pesquisa. O programa permite o uso de *stoplists* para que essas palavras sejam ignoradas na busca e sejam visualizadas apenas as palavras representativas do *corpus* (figura 3.3).

Figura 3.2 – Wordlist do Corpus Papel

	Type	Rank	Freq	Range
8	em	8	14004	161
9	para	9	11275	160
10	se	10	9883	160
11	os	11	9782	158
12	com	12	8820	159
13	as	13	7525	159
14	no	14	7407	157
15	um	15	7292	157
16	é	16	6864	150
17	dos	17	6802	154
18	uma	18	6699	157
19	por	19	6440	159
20	como	20	6337	157
21	na	21	5824	156
22	ou	22	5294	143
23	das	23	4563	153
24	não	24	4562	152
25	ao	25	4472	159
26	papel	26	4132	148
27	ser	27	4013	144
28	conservação	28	3903	150

Search Query Words Case Regex

Fonte: AntConc.

Figura 3.3 – *Wordlist* do *Corpus* Papel com *stoplist*

KWIC		Plot	File	Cluster	N-Gram	Collocate	Word
Types 37748/37748		Tokens 947419/947419		Page Size 100 hits		▼	
	Type	Rank	Freq	Range			
1	papel	26	3996	139			
2	conservação	28	3846	142			
3	documentos	31	3139	107			
4	mais	32	3035	138			
5	preservação	33	2994	116			
6	acervo	37	2388	111			
7	restauração	39	2089	110			
8	materiais	45	1605	128			
9	livros	46	1560	104			
10	acervos	48	1528	99			
11	umidade	52	1384	99			
12	arquivo	54	1298	80			
13	patrimônio	55	1280	77			
14	biblioteca	56	1271	86			
15	forma	57	1267	115			
16	obras	58	1247	112			
17	processo	60	1171	119			
18	material	64	1133	116			

Search Query Words Case Regex

Fonte: AntConc.

N-grams: a ferramenta extratora de n-gramas, também chamada de MWE (*multiword expressions*) ou expressões complexas, permite identificar estruturas de diferentes tamanhos (por exemplo, de 2 a 3 palavras, de 2 a 4 palavras, de 2 a 5 palavras, de 3 a 4 palavras, de 3 a 5 palavras etc.) que apresentam uma ocorrência frequente no *corpus* de estudo. O extrator de n-gramas é bastante útil para identificar expressões complexas típicas do *corpus* de estudo e pode ser empregado como extrator de termos sintagmáticos e unidades candidatas a fraseologias especializadas. Contudo, cabe ressaltar que ele não filtrará somente expressões complexas iniciadas e terminadas por palavras lexicais (por exemplo, *estado de conservação*), mas também estruturas recorrentes iniciadas e terminadas por palavras

gramaticais (por exemplo, *de Conservação e*). A figura 3.4 ilustra n-gramas do *Corpus* Papel do tamanho de 3 a 4 palavras no *Sketch Engine*.

Figura 3.4 – n-gramas do *Corpus* Papel

3–4-grams, word (Items: 14,821 , total frequency: 156,296)

Word	Frequency ?	Word	Frequency ?
1 conservação e restauração	389 ...	18 Conservação e Restauração	176 ...
2 Rio de Janeiro	327 ...	19 a partir da	176 ...
3 de conservação e	299 ...	20 do século XIX	174 ...
4 temperatura e umidade	286 ...	21 e umidade relativa	174 ...
5 a necessidade de	240 ...	22 de conservação e restauração	171 ...
6 estado de conservação	238 ...	23 umidade relativa do	169 ...
7 a fim de	237 ...	24 uma vez que	155 ...
8 de temperatura e	225 ...	25 de Conservação e	155 ...
9 o uso de	224 ...	26 temperatura e umidade relativa	147 ...
10 a partir de	216 ...	27 para a preservação	145 ...
11 e restauração de	216 ...	28 relativa do ar	144 ...
12 de acordo com	213 ...	29 preservação e conservação	143 ...
13 por meio de	207 ...	30 do patrimônio cultural	143 ...
14 obras de arte	203 ...	31 de preservação e	142 ...
15 conservação e restauração de	197 ...	32 e Restauração de	141 ...
16 o processo de	191 ...	33 de bens culturais	139 ...
17 de temperatura e umidade	188 ...	34 do Rio de	138 ...

Fonte: Sketch Engine.

Keyword list: essa lista de palavras-chave é resultado da comparação das palavras do *corpus* de estudo com as de um *corpus* de referência, como é o caso do *BNC* (*British National Corpus*¹⁸), cuja importação pode ser feita para o *AntConc*. Tal comparação permite que o pesquisador identifique palavras que são estatisticamente mais frequentes no *corpus* de estudo. Quanto

18 *Corpus* de livre acesso constituído de textos de inglês britânico falado e escrito com mais de cem milhões de palavras.

maior for a frequência estatística de uma palavra nesse *corpus*, maior será sua especificidade. Em termos simples, o cálculo feito pelo extrator consiste na comparação da proporção de ocorrências de uma palavra no *corpus* de estudo frente à proporção de ocorrências dessa mesma palavra no *corpus* de referência. No trabalho terminológico, essa ferramenta é usada para chegar a candidatos a termos. A figura 3.5 mostra uma lista de palavras-chave do *corpus* em língua inglesa do projeto Papel, em que a palavra *paper* aparece como sendo a mais frequente (10.021 ocorrências), sendo que no *corpus* de referência essa mesma palavra aparece com uma frequência de 155 ocorrências, o que indica sua alta especificidade no *corpus* de estudo.

Figura 3.5 – Keywords do Corpus Papel

The screenshot shows the AntConc interface with the following details:

- Target Corpus:** Name: temp, Files: 109, Tokens: 848699. A list of files from enPP001.txt to enPP016.txt is shown.
- Reference Corpus:** Name: AmE06, Files: 500, Tokens: 1017879. A list of files from AmE06_A01.txt to AmE06_A16.txt is shown.
- Keyword Statistics:** Keyword Types: 2268/24626, Keyword Tokens: 516648/848699, Page Size: 100 hits.
- Table:** A table with columns: Rank, Type, Freq_Tar, Freq_Ref, Range_Tar, Range_Ref, Likelihood, Effect. The top results are:

Rank	Type	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Likelihood	Effect
1	paper	10021	155	105	95	14442.69	0.023
2	papers	2201	55	83	40	3021.97	0.005
3	used	3238	566	102	267	2595.495	0.008
4	cellulose	1599	0	74	0	2522.172	0.004
5	conservation	1622	12	103	7	2431.163	0.004
6	water	2755	406	87	137	2414.967	0.006
7	treatment	1899	114	90	53	2257.723	0.004
8	materials	1647	70	96	40	2097.624	0.004
9	be	8607	4651	109	499	2041.514	0.02
10	or	7465	3790	108	489	1995.089	0.017
11	surface	1629	92	82	50	1963.038	0.004
12	adhesive	1228	1	54	1	1921.669	0.003
13	object	1428	55	64	39	1848.58	0.003
14	ph	1209	13	63	6	1778.467	0.003
15	may	3150	912	102	302	1748.523	0.007
16	parchment	1096	4	37	3	1680.34	0.003
17	samples	1220	33	61	21	1658.889	0.003
18	et	1428	163	41	32	1398.286	0.003
19	fibers	914	4	44	3	1394.666	0.002
20	solution	1095	47	67	32	1391.847	0.003
21	is	11768	8420	109	488	1347.553	0.027
22	funga	821	0	38	0	1294.588	0.002
23	of	33395	30331	109	500	1273.711	0.073
- Search Query:** Words, Case, Regex. Results: 1 to 100 of 2268 hits.
- Sort by:** Likelihood, Invert Order.

Fonte: AntConc.

O *Sketch Engine*, como mencionamos anteriormente, já possui *corpora* de referência acoplados a ele¹⁹, inclusive do português brasileiro; assim, não há necessidade de se buscar um *corpus* de referência externo para gerar uma lista de palavras-chave. Outra vantagem dessa ferramenta é que ela identifica não somente candidatos a termos simples, mas também sintagmáticos (ver figura 3.6).

Figura 3.6 – *Keyword*: extração de candidatos a termos sintagmáticos

Word	Word
1 conservação preventiva ...	14 papel japonês ...
2 umidade relativa ...	15 acervos documentais ...
3 preservação documental ...	16 restauração de bens ...
4 restauração de papel ...	17 restauração de bens culturais ...
5 bens culturais ...	18 pasta mecânica ...
6 obras raras ...	19 polpa química ...
7 edson motta ...	20 plantas arquitetônicas ...
8 arquivo nacional ...	21 suporte de papel ...
9 patrimônio documental ...	22 documentos gráficos ...
10 preservação de acervos ...	23 agentes biológicos ...
11 acervos bibliográficos ...	24 fibras de celulose ...
12 restauração de documentos ...	25 tinta ferrogálica ...
13 política de preservação ...	26 reserva alcalina ...

Fonte: Sketch Engine.

¹⁹ O *corpus* de referência do português oferecido pelo *Sketch Engine*, o *Portuguese Web 2018* (pt-TenTen18), está formado por aproximadamente 4 bilhões de palavras e representa o português brasileiro.

Word Sketch: a ferramenta *Word Sketch* oferece o padrão colocacional de uma palavra pesquisada, reunindo, em uma única página, o sumário do comportamento colocacional do item buscado, organizado por suas relações sintáticas. Por exemplo, se no *Corpus Papel* do Termisul for buscado o termo *acervo*, o *Word Sketch* indicará, por estrutura sintática, as palavras que costumam aparecer juntas com esse termo, considerando não somente a frequência, mas também outros índices matemáticos, como o *Mutual Information* (MI) (Church; Hanks, 1990), responsáveis por identificar estruturas complexas cujas palavras apresentam uma atração semântica entre si. Assim, o *Word Sketch* permitirá identificar para o termo *papel* colocações especializadas formadas por diferentes estruturas morfossintáticas como, por exemplo, nome + adjetivo: *papel japonês*, *papel translúcido*, *papel vegetal*; verbo + nome: *degradar papel*, *enfraquecer papel*, *daniificar papel*; nome + particípio: *papel envelhecido*, *papel reciclado*; nome deverbal + preposição + nome: *restauração de papel*, *deterioração de papel*; e outras estruturas possíveis (ver figura 3.7).

Figura 3.7 – *Word Sketch* do termo *papel*

The screenshot shows the Word Sketch tool interface. At the top, the title is "WORD SKETCH" and the search term is "corpus papel". Below the title, it indicates "papel as noun 4,981x". The interface is divided into four columns, each representing a different collocational pattern:

- papel + adjetivo**: Includes terms like japonês, ácido, translúcido, alcalino, neutro, vegetal, artesanal, industrial, ingre, moderno, transparente, and contemporâneo.
- verbo + papel**: Includes terms like desempenhar, tornar, fabricar, atacar, produzir, utilizar, assumir, degradar, enfraquecer, destruir, danificar, and conferir.
- papel + adjetivo participial**: Includes terms like produzir, fabricar, fazer, manufaturar, utilizar, tratar, envelhecer, colar, and industrializar.
- sintagma preposicional**: Includes terms like ...de papel, ...em papel, papel de substantivo, papel em substantivo, ...sobre papel, papel com substantivo, ...com papel, ...a papel, papel por substantivo, papel para substantivo, papel a substantivo, and ...para papel.

Fonte: Sketch Engine.

Word Sketch difference: segue o mesmo funcionamento do *Word Sketch*, porém é usado para comparar e contrastar o padrão colocacional de duas palavras diferentes. Ou seja, duas palavras são pesquisadas no *corpus* e a ferramenta indicará com quais outros itens lexicais cada uma costuma aparecer. Cada palavra pesquisada tem uma cor – vermelho ou verde – que será aplicada aos colocados, identificando o grau de atração que há entre eles e o item de pesquisa. As palavras indicadas na parte central de fundo cinza são aquelas que costumam ser usadas com ambos os itens pesquisados.

Esse é um recurso bastante útil para a tradução de língua geral, por exemplo, para decidir entre palavras que apresentam valores sinônimos,

mas que não são empregadas nas mesmas combinações de palavras. No caso da terminologia, essa ferramenta pode ser bastante proveitosa para a conceitualização (através da criação de árvores de domínio ou mapas conceituais) de unidades de valor especializado. A figura 3.8 ilustra a comparação entre os verbos *conservar* e *restaurar*, termos que definem a área de estudo do *Corpus Papel*. Observa-se, a partir do padrão colocacional desses verbos, que *conservar* é um termo associado, por um lado, a uma noção de cuidado: *guardar, salvaguardar, prevenir, valorizar*; e, por outro, a uma noção de registro: *transmitir, disseminar, documentar, expor*. Já o verbo *restaurar* associa-se às ações práticas da área: *reparar, encadernar, renovar*.

Figura 3.8 – *Word Sketch difference* dos verbos *conservar* e *restaurar*



Fonte: Sketch Engine.

Também é possível compreender os valores desses verbos e as diferenças conceituais entre eles pela estrutura de colocação participial (segunda coluna da figura 3.8). Enquanto *conservar* é uma prática realizada com *acervos* e *sistemas*, a prática de *restaurar* é realizada com os bens em suporte papel em si, como *folhas*, *jornais*, *mapas*, *artefatos* etc. Com essa análise, pode-se chegar a conclusões de que *conservar* é uma tarefa mais voltada para o âmbito institucional e de gestão, enquanto o ato de *restaurar* envolve procedimentos práticos realizados nos bens materiais.

Parallel concordance: para usar essa ferramenta, é necessário ter *corpora* paralelos disponíveis. É o caso do *corpus* EUR-LEX JUDGMENTS, disponibilizado pelo próprio programa, que oferece textos jurídicos traduzidos. *Corpora* paralelos são bastante úteis para a tradução especializada. A figura 3.9 demonstra o exemplo da busca por um equivalente, em inglês, do termo jurídico *processo*. As concordâncias paralelas oferecem diferentes contextos do termo *case* sendo empregado em sua maioria, além da opção *proceedings*, cabendo ao tradutor analisar e optar pela melhor solução tradutória.

Figura 3.9 – *Parallel concordance* do termo *processo*

The screenshot shows the 'PARALLEL CONCORDANCE' interface for the search term 'processo'. The search results are displayed in a table with two columns: Portuguese text and English text. The results show various contexts where 'processo' is used in Portuguese and its corresponding English equivalents, such as 'proceedings', 'case', and 'report'. The interface includes a search bar, a results count (112,220), and navigation controls.

Portuguese Text	English Text
EXPLORACAO DE UMA SUCURSAL. - PROCESSO C-439/93. Colectânea da Jurispr	Judgment of the Court of 6 April 1995. - Lloyd's Register of Shipping v Soci�t� Campenon Be
� estabelecida. Partes No processo C-439/93, que tem por objecto um	in Case C-439/93,
interpreta�o defendida pela recorrente no processo principal privaria de quase todo o efeito �til	17 Secondly, the interpretation put forward by the appellant in the main proceedings would
s�o reembols�veis. Revestindo o processo , quanto �s partes na causa principal, a	23 The costs incurred by the French and Greek Governments, the United Kingdom and the C
NEGOCIOS - SEXTA DIRECTIVA IVA. - PROCESSO C-20/91. Colect�nea da Jurispru	Judgment of the Court (Third Chamber) of 6 May 1992. - Pieter de Jong v Staatssecretaris var
elo do terreno. Partes No processo C-20/91, que tem por objecto um	in Case C-20/91,
dos factos e do enquadramento jur�dico do processo principal, da tramita�o do processo bem c	10 Reference is made to the Report for the Hearing for a fuller account of the facts and lega
do processo principal, da tramita�o do processo bem como das observa�es escritas	10 Reference is made to the Report for the Hearing for a fuller account of the facts and lega
ira audi�ncia. Estes elementos do processo apenas ser�o adiante retomados na medid	10 Reference is made to the Report for the Hearing for a fuller account of the facts and lega
entac�o do Tribunal. 11 No actual processo , est�o essencialmente em causa duas	11 The present case is concerned essentially with two provisions:

Fonte: Sketch Engine.

Fechamos o cap tulo com a proposta de atividades que permitir o que voc  pratique todas as etapas da constru o de um *corpus* aqui descritas

e seu processamento nas duas ferramentas apresentadas. Esperamos que elas sirvam como um ponto de partida para futuras pesquisas com o uso de *corpus*.

ATIVIDADES: *compilação de corpus de estudo*

Agora você pode praticar um pouco o que abordamos até agora, construindo um pequeno *corpus* e testando as ferramentas de extração. Sugerimos que construa um *corpus* em língua portuguesa e depois experimente construir outro com uma língua estrangeira para observar diferenças e semelhanças entre elas.

1. Escolha uma área de seu interesse e pense em um tópico dentro dessa área para fazer o seu recorte. O nosso projeto, por exemplo, é da área da *Conservação e Restauração de bens móveis em suporte papel (Projeto Papel)*, sendo nosso objetivo identificar os termos e as UFEs dessa área.

2. Agora que já tem a área e o recorte definidos, busque na internet 20 textos de diferentes gêneros sobre o tópico escolhido. Você pode usar a pesquisa avançada do Google, que permite identificar textos de países determinados, escritos em determinado idioma, e que também oferece o filtro de busca por palavras-chave específicas.

3. A seguir, os textos precisam passar por uma limpeza, serem salvos em <.txt> (codificação UTF-8) e catalogados. Veja acima, na seção “Critérios de coleta, limpeza e organização”, o que falamos sobre os elementos que precisam ser retirados dos textos antes de serem salvos. Não se esqueça de criar seu próprio sistema de código para ordenar os textos e incluí-los no seu catálogo.

4. Após compilar o *corpus*, gere sua lista de palavras usando o *AntConc*. A primeira palavra lexical dessa lista deve ser representativa da temática do seu *corpus*, isto é, provavelmente corresponderá a uma das palavras-chave

que você usou para a busca e seleção de textos. Se isso não acontecer, rejeite os critérios de seleção e limpeza dos textos escolhidos para compor seu *corpus*.

5. Se quiser identificar termos e fraseologias especializadas da área de seu interesse sobre a qual você compilou seu *corpus*, você pode fazer uma conta gratuita no programa *Sketch Engine* e compilar seu *corpus* de estudo ali. Em seguida, gere a *keyword list* e identifique os principais termos que aparecem nessa lista (para saber mais sobre identificação de candidatos a termo e sobre o valor terminológico da unidade lexical, leia o capítulo 4).

6. Depois de identificar alguns termos, use a ferramenta *Word Sketch* para observar se esses termos formam fraseologias especializadas. Para identificá-las, pesquise o termo no *Word Sketch* (sugerimos que, se seu *corpus* tiver menos de 1 milhão de palavras, escolha um corte de frequência que fique entre 2 e 5 ocorrências). Em seguida, observe e selecione as estruturas verbais e nominalizadas com as quais esse termo aparece registrado, isto é, [verbo + nome(termo)] e [nome deverbal + de + nome(termo)]. Essas serão as estruturas candidatas a unidades fraseológicas especializadas da linguagem que seu *corpus* de estudo representa. Para saber mais sobre identificação e extração de fraseologias especializadas, leia, também, o capítulo 4.

Você também pode identificar os candidatos a termos do seu *corpus* com a ferramenta *Keyword list*, do *AntConc*; contudo, para isso, você precisará ter um *corpus* de referência para contrastar com o seu de estudo. Da mesma forma, uma vez identificados alguns termos, você também pode buscar pelas fraseologias usando a ferramenta *Clusters/N-grams* do *AntConc*.

PARA SABER MAIS

ARCOS, Manuela; BEVILACQUA, Cleci R. UFE eventivas na área da conservação e restauração de bens culturais móveis em suporte papel: identificação e análise. *Debate Terminológico*, [Porto Alegre], n. 18, p. 4-18, 2020. Disponível em: <https://seer.ufrgs.br/riterm/article/view/98700>. Acesso em: 17 set. 2021.

ARCOS, Manuela; BEVILACQUA, Cleci R. Metodologias para a extração e identificação de unidades fraseológicas especializadas eventivas em corpora textuais. *Guavira Letras*, v. 27, p. 75-95, 2018. Disponível em: <http://websensors.net.br/seer/index.php/guavira/issue/view/34/showToc>. Acesso em: 7 jun. 2022.

BAKER, Mona. Corpora in Translation Studies. An overview and suggestions for future research, *Target*, v. 7, n. 2, p. 223-243, 1995.

BAKER, Mona. Corpus Linguistics and Translation Studies: implications and applications. In: BAKER, Mona; FRANCIS, Gill; TOGNINI-BONELLI, Elena (org.). *Text and Technology: in honour of John Sinclair*. Amsterdam: John Benjamins, 1993. p. 233-250.

BERBER SARDINHA, Antônio Paulo. *Linguística de Corpus*. Barueri: Manola, 2004.

BERBER SARDINHA, Antônio Paulo. Histórico e problemática. *D.E.L.T.A.*, v. 16, n. 2, p. 323-367, 2000.

BIBER, Douglas. Representativeness in corpus design. *Literary and Linguistic Computing*, v. 5, n. 4, p. 243-257, 1993.

MARCUSCHI, Luiz Antônio. *Produção textual, análise de gêneros e compreensão*. São Paulo: Parábola Editorial, 2008.

MARCUSCHI, Luiz Antônio. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, Ângela P.; MACHADO, Anna Raquel; BEZERRA, Maria Auxiliadora (org.). *Gêneros textuais e ensino*. 4. ed. Rio de Janeiro: Lucerna, 2005.

MCENERY, Tony; HARDIE, Andrew. *Corpus Linguistics: method, theory and practice*. Edinburgh: Cambridge University Press, 2012.

TAGNIN, Stella E. O. Glossário de linguística de corpus. In: *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011. p. 357-361.

Capítulo 4 – Seleção de unidades terminológicas: estratégias de extração e princípios de identificação

Sandra Dias Loguercio
Manuela Arcos Machado

A seleção de unidades terminológicas, sejam termos ou fraseologias, confunde-se com a própria definição dos objetos da Terminologia e, desse modo, com as fronteiras da disciplina e seu estabelecimento. Se entendemos que as unidades terminológicas são unidades lexicais de sentido e uso especializados, resultantes de uma conceitualização (uma maneira de compreender as coisas), mas também do jeito com que uma dada comunidade de saber se expressa – daí serem associadas a gírias ou jargões profissionais –, e não formas artificiais de linguagem, na prática, estabelecer o que é da linguagem comum e o que adquire traços de especialidade está longe de ser evidente e assunto esgotado.

Neste capítulo, abordaremos assim estratégias de extração e seleção, que contam com a ajuda de programas de processamento automático da linguagem, mas sobretudo discutiremos princípios que norteiam a identificação de termos e de UFEs. Retomando várias noções já abordadas ao longo do manual e levando à risca a abordagem linguístico-textual, trataremos de palavras, textos e contextos das diferentes amostras discursivas (organizadas em *corpora*), coletadas em meio à produção de sujeitos de saber de uma área de conhecimento ou ofício. Essas amostras servirão como uma janela ao tradutor, normalmente não especialista, de modo que possa espiar o que dizem e como dizem os especialistas – que não deixam de informar sobre a própria natureza de seu saber –, e assim transitar pelos diferentes discursos que ajuda a criar, através dos encontros entre culturas que promove, com mais segurança e mesmo com conhecimento de causa.

Como identificar termos? Ou o que, afinal, é próprio ao que dizem os especialistas?

Relembrando o que já se disse no Capítulo 1, podemos compreender **termo** como uma unidade lexical que é, simultaneamente, uma unidade linguística e uma unidade de conteúdo especializado. Se traçarmos um paralelo com a língua geral, os termos são as palavras cujo conteúdo é da ordem dos **conceitos***, enquanto o conteúdo das palavras comuns é da ordem dos **significados*** (Krieger; Finatto, 2004, p. 77). A diferença, então, entre um termo e uma palavra comum reside no **sentido** e no **uso especializado** do primeiro.

Por serem, portanto, oriundos de um discurso, que vai se moldando entre pares que compartilham saberes sobre o mundo, os termos são, predominantemente, de natureza designativo-denominativa. Daí corresponderem, normalmente, a uma forma nominal e receberem qualificativos conforme o conhecimento vai sendo aprofundado em determinada área. Por exemplo, passamos a separar *lixo* em *lixo orgânico* e *lixo seco*, a partir do que ambientalistas nos ensinaram. Essa característica sintagmática, em que o nome é composto de mais de uma palavra, predomina nos discursos especializados. Mas o termo pode adquirir outras formas, todas resultantes dos processos de comunicação e de representação do conhecimento, como formas truncadas (*micro* para dizer microcomputador, *diagnóstico por imagem* ou ainda *exame de imagem* que correspondem à radiologia e ao diagnóstico por imagem), siglas e acrônimos (AVC, Enem, Iphan, Termisul) e mesmo unidades não verbais, como as fórmulas (H_2O e CO_2).

Não podemos esquecer ainda que, em razão de sua natureza linguística, por mais restrito que seja o círculo de especialistas ou de iniciados em determinado saber e por mais esforço que venham a empenhar para normalizar a linguagem, ou seja, “falar a mesma língua”, as unidades terminológicas refletem fenômenos comuns às palavras de maneira geral, como a variação, tanto denominativa quanto conceitual, como visto no Capítulo 1. Essas questões vão aparecer na extração e seleção dos termos e

* Conceito é compreendido como significado estabilizado, fruto de uma elaboração teórica, e compartilhado entre especialistas, colegas de profissão ou pessoas que realizam uma dada atividade, ou seja, forjado por uma comunidade de saber.

* Diferentemente do conceito, significado, aqui, é mais instável, dependente do contexto de uso e da interpretação dos sujeitos.

não podem ser negligenciadas quando optamos por uma perspectiva descritiva da terminografia.

Comentamos a seguir algumas etapas essenciais desse estudo descritivo.

1. Aproximação com a área de estudo

A seleção de termos de uma dada área está intimamente associada a um estudo semântico, sobretudo, que busca identificar o valor especializado que as unidades lexicais ativam em dado contexto de comunicação. Com base nos princípios da TLT, conforme descritos no Capítulo 1, a abordagem que propomos leva em conta elementos que dizem respeito às condições de produção e recepção dos textos, que se refletem nas características mais gerais dos gêneros textuais (levadas em conta já na composição do *corpus*, como visto no Capítulo 3), mas igualmente na configuração e nas relações entre saberes dentro de determinada área – sua arquitetura –, assim como nos elementos textuais (internos a cada manifestação textual) e discursivos (que se repetem ou se completam quando apreendidos em um conjunto de textos).

Esse estudo dos aspectos textuais antecede a extração das unidades lexicais especializadas, fornecendo as primeiras pistas para sua identificação. Se estamos trabalhando com textos da legislação ambiental, sabemos de imediato que haverá, no mínimo, duas categorias de unidades terminológicas: aquelas relativas mais especificamente ao ato de legislar, ao Direito, e aquelas relativas a subáreas ambientais em questão, por *exemplo*, *contaminação e preservação da água, do solo, do ar, sonora* etc. Do mesmo modo, as características do gênero textual, sobretudo no que diz respeito à maneira como o texto se estrutura e à postura retórica predominante (tecnopragmática, epistêmica, deontica, estética etc.), vão apontar para fórmulas de rotina – frases feitas ou outros tipos de fraseologias que marcam passagens dos textos – e um tipo de vocabulário. Assim, por exemplo, em toda e qualquer legislação, evidencia-se a importância dos verbos *promulgar, decretar, sancionar*, entre outros; em uma procuração, *nomear, instituir*;

em uma ação penal, *julgar, punir*. Afinal, faz parte da postura deôntica do Direito “ordenar, tanto no sentido de dar ordens, como no de colocar as coisas em ordem” (Maciel, 2001, p. 92). Em artigos ou relatórios de áreas técnicas, como a de Conservação e Restauração de bens em papel, por exemplo, será comum o uso de locuções que denotem finalidade, como *a fim de* e *com o objetivo de*, e o emprego de formas instrucionais, como *é preciso*, *é necessário*, *deve-se* etc. Isso se verifica porque a descrição de uma técnica, resultante de experiências empíricas, busca justificar um modo de fazer.

2. Construção da árvore de domínio e de mapa conceitual

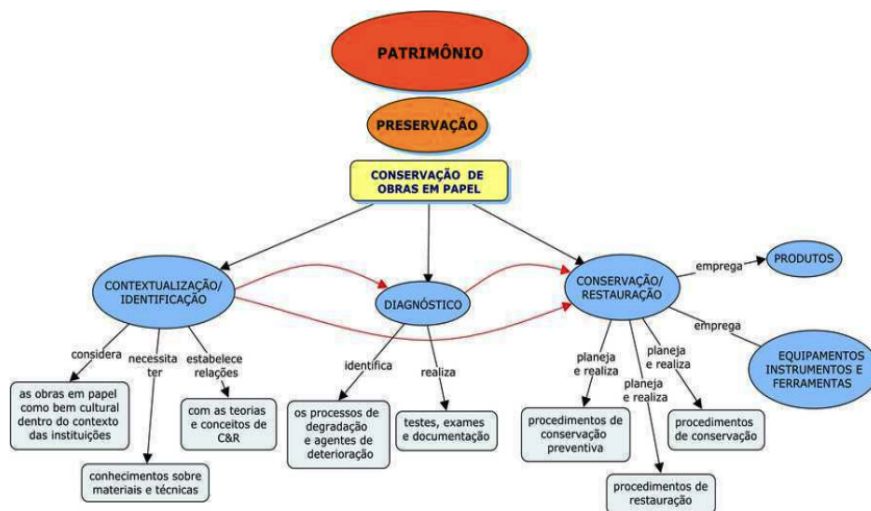
Essa etapa de aproximação com a terminologia da área pode ser seguida ainda da elaboração de uma árvore de domínio ou de um *mapa conceitual* (já tratados no Capítulo 2), que ajuda a compreender efetivamente a arquitetura dessa área: suas subáreas, intersecções, campos temáticos, além de suas relações com outras áreas. A árvore ou o mapa pode ter distintas configurações, sendo fruto de uma leitura interpretativa acerca da hierarquia entre os termos ou categorias de termos (campos nocionais). Pode partir de termos mais genéricos e chegar aos mais específicos – como normalmente se vê na árvore de domínio –, ou adquirir outros desenhos menos hierárquicos, mais dinâmicos – como se vê nos mapas –, especificar ou não a natureza de suas relações. Tanto a relação entre os termos ou campos quanto seu detalhamento podem variar segundo a perspectiva do sujeito e seus propósitos. A seguir, vemos uma árvore de domínio e um mapa conceitual.

Figura 4.1 – Estrutura de conhecimento da área da Conservação e Restauração de documentos gráficos



Fonte: Bojanoski; Michelon; Bevilacqua (2017, p. 43).

Figura 4.2 – Mapa conceitual da organização da área da Conservação e Restauração



Fonte: Bojanoski (2018, p. 166).

Como explicam as autoras da árvore de domínio, essa “estrutura de conhecimento” foi elaborada “a partir dos princípios e teorias já estabelecidos na área de Conservação e Restauração sobre as abordagens dos bens culturais” (Bojanoski; Michelon; Bevilacqua, 2017, p. 43). No mapa

conceitual, de estrutura mais complexa e dinâmica, a autora explicita as relações entre os conceitos. Tanto a árvore quanto o mapa são constituídos de palavras-chave (hiperônimos) para a área. É comum, porém, que o próprio levantamento terminológico, conforme vai sendo aprofundado, produza alterações na árvore e/ou no mapa, na medida em que pode haver novas compreensões das relações estabelecidas ou se descobrir novos campos. Por outro lado, a seleção das unidades terminológicas e, mais adiante, sua definição, são facilitadas por essa espécie de esquema inicial, que delimita os contornos do campo de conhecimento.

3. *Extração de candidatos a termos*

Essa primeira etapa de estruturação da área, de seus limites e de exploração dos textos é seguida por uma etapa de extração propriamente dita de **candidatos a termos***. Para a extração, que leva em conta os critérios comentados adiante, são usados programas de processamento automático da linguagem, como aqueles que foram mencionados no Capítulo 3, entre outros, o *AntConc* e o *Sketch Engine*. As ferramentas e as estratégias que elencamos e sobre as quais discorreremos não são interdependentes, nem configuram passos a serem seguidos necessariamente na ordem proposta; são antes pistas para o reconhecimento e a seleção de candidatos a termos, que, por sua vez, também estarão atrelados, como visto no Capítulo 2, a decisões prévias, como, entre outras, o público-alvo (o consulente a ser privilegiado) e os propósitos ou a função do trabalho terminográfico. Para fins de ilustração, apoiamo-nos principalmente no levantamento da terminologia da área da Conservação e Restauração de obras em papel, abordando três critérios: índice de frequência, chavicidade e presença de contextos definitórios.

Um dos critérios utilizados para a extração é o índice de frequência das unidades lexicais no *corpus* textual. Nesse caso, recorre-se à lista de palavras (*wordlist*), como vemos na figura a seguir.

* As etapas de extração e seleção de terminologia devem ser seguidas por uma etapa de validação dos termos, o que normalmente conta com a colaboração de especialistas da área de conhecimento abordada.

Figura 4.3 – Wordlist do Corpus Papel

	Type	Rank	Freq	Range		Type	Rank	Freq	Range
1	de	1	60439	161	22	ou	22	5294	143
2	e	2	34533	161	23	das	23	4563	153
3	a	3	32986	161	24	não	24	4562	152
4	o	4	21773	161	25	ao	25	4472	159
5	que	5	17677	161	26	papel	26	4132	148
6	do	6	16914	161	27	ser	27	4013	144
7	da	7	14903	160	28	conservação	28	3903	150
8	em	8	14004	161	29	à	29	3820	151
9	para	9	11275	160	30	são	30	3206	142
10	se	10	9883	160	31	documentos	31	3205	116
11	os	11	9782	158	32	mais	32	3108	148
12	com	12	8820	159	33	sua	33	3028	147
13	as	13	7525	159	34	preservação	34	3013	121
14	no	14	7407	157	35	foi	35	2784	149
15	um	15	7292	157	36	sobre	36	2460	143
16	é	16	6864	150	37	acervo	37	2422	116
17	dos	17	6802	154	38	p	38	2363	62
18	uma	18	6699	157	39	restauração	39	2134	115
19	por	19	6440	159	40	nos	40	2030	133
20	como	20	6337	157	41	também	41	1893	139
21	na	21	5824	156	42	pela	42	1835	141

Fonte: Corpus Papel e AntConc.

Imediatamente, percebemos que as primeiras unidades lexicais de conteúdo – ou seja, que não correspondem a preposições, conjunções, pronomes etc. – remetem à temática do *corpus* e mesmo aos hiperônimos que compõem a árvore de domínio ou mapa conceitual comentados acima. A partir dessa primeira lista, que tornará ainda mais evidente a relevância dessas unidades se utilizarmos o recurso de exclusão de palavras (*stoplist*), “limpando” a listagem daquilo que não interessa à pesquisa, podemos consultar a lista de concordâncias (*concordance*), que ajuda a perceber rapidamente se se trata de uma unidade lexical simples ou complexa ou mesmo outros aspectos da linguagem (colocações, estruturas gramaticais, entre outros). Além disso, a concordância nos leva aos contextos imediatos em que se encontra a unidade lexical observada, auxiliando na análise semântica propriamente dita.

Figura 4.4 – Concordâncias do termo *papel*

File	Left Context	Hit	Right Context
670 ptPP006.txt	cadernos, os pares de folhas que deveriam ser unidos com	papel	japonês fino bege e metil celulose, mantendo as fibras
671 ptPP124.txt	e mata-borrão e tábua, sob pesinhos. O excesso de	papel	japonês foi retirado com o bisturi. Essa técnica foi
672 ptPP012.txt	°C, observamos que mesmo com pressão menor, 0,5 T ATM, o	papel	japonês funde-se satisfatoriamente ao suporte. Todavia, obtivemos r
673 ptPP030.txt	e na zona dos encaixes e pequenas perfurações, enxertos com	papel	japonês grosso. A etiqueta de couro foi reconstituída com
674 ptPP006.txt	japonês fino bege e metil celulose, mantendo as fibras de	papel	japonês no mesmo sentido de orientação do papel original. –
675 ptPP152.txt	s documentos, procedendo pequenos reparos (remendos), utilizando	papel	japonês ou outro alcalino e cola metilcelulose para impedir
676 ptPP124.txt	estudo deste trabalho. Então, foi necessário fazer o tingimento do	papel	japonês para realizar os reparos. *Tingir papeis é o
677 ptPP022.txt	Optamos por efetuar a velatura dos mapas com tecido e	papel	japonês pelas seguintes razões: 1) grande dimensão do material, 2) p
678 ptPP148.txt	deve ser completada com a fixação de um reforço de	papel	japonês pelo verso da área tratada. Enfim, para a
679 ptPP012.txt	as vezes, o documento laminado torna-se ilegível. Embora o	papel	japonês possa ser encontrado na cor natural, o que
680 ptPP037.txt	seda. (fig nº 1 Processo de karibari). A última camada de	papel	japonês que cobre o karibari é impermeabilizada com suco
681 ptPP002.txt	alcalina. (pH final do material: 7.5) – Reconstituição: . Velaturas com	papel	japonês resistente: todas as pranchas foram fixadas em novo
682 ptPP075.txt	rs e eucalipto refinadas em holandesa, fornecidas pelo laboratório de	papel	do Arquivo Nacional, com colagem alcalina e coloração formulada
683 ptPP144.txt	ria coletiva e documentada dos povos. Ao analisarmos criticamente o	papel	do Arquivo Nacional, frente a essa política nacional de
684 ptPP031.txt	ao apoio da equipe e às instalações da Oficina de	papel	do Arquivo Nacional, pode implantar imediatamente esta nova metod
685 ptPP116.txt	e longa DSC02207 (Eucaliptus e Pinus SP) no laboratório de	papel	do Arquivo Nacional no Rio de Janeiro. Os corpos
686 ptPP140.txt	rel e detalhamento de sua constituição físico-química (Laboratório de	Papel	do Arquivo Nacional–RJ e Laboratório de Celulose e
687 ptPP127.txt	ina. op. cit., s.p. 121 Jacqueline Assis, conservadora-restauradora de	papel	do Museu Nacional de Belas Artes, ressaltou a importância

Fonte: *Corpus Papel e programa AntConc.*

Ao observar a lista de concordâncias da palavra *papel*, percebemos imediatamente que ela compõe unidades de sentido maiores, como *papel japonês* ou *Laboratório de Papel do Arquivo Nacional (RJ)*; assim como podemos identificar elementos que caracterizam esse objeto, apontando para sua especificidade conceitual: sabemos, por exemplo, que o *papel japonês* pode ser *fino* ou *grosso*, tem um pH *alcalino*, pode ser encontrado na *cor natural* ou pode ser *tingido*, entre outras informações, que poderão ser usadas ou não na ficha terminológica e na redação final do verbete.

Para encontrar mais rapidamente unidades lexicais complexas, que tendem a ser recorrentes na linguagem especializada, ou mesmo certificar-se do que foi observado na lista de concordância, pode-se recorrer às ferramentas n-gramas e/ou Colocados (*Collocate*). A primeira, depois de definir o número de itens lexicais mínimo e máximo, por exemplo, de 2 a 3 palavras gráficas – o que pode ser definido por meio de testagens –, bem como, se pertinente para o projeto, o mínimo de ocorrências no *corpus*, fornece uma lista de “blocos de palavras”. A segunda, a partir de um item de busca, possibilita refinar a observação, permitindo que se decida olhar somente para o que vem à direita da palavra ou à esquerda, ou ainda para

ambos os lados, e também a frequência mínima com que os itens aparecem lado a lado, como ilustramos a seguir.

Figura 4.5 – Colocados do termo *papel*

The screenshot shows the AntConc Collocate window. At the top, there are tabs for KWIC, Plot, File, Cluster, N-Gram, Collocate (selected), Word, and Keyword. Below the tabs, it displays 'Collocate Types 46', 'Collocate Tokens 5440', and 'Page Size 100 hits'. The main area is a table with the following columns: Collocate, Rank, FreqLR, FreqL, FreqR, Range, Likelihood, and Effect. The search query is 'papel', and the results are sorted by Likelihood. The search parameters are: Search Query: Words, Case, Regex; Window Span: From 1L to 1R; Min. Freq: 10; Min. Range: 1.

	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	do	1	1070	1004	66	100	2570.961	2.89
2	o	2	797	713	84	104	1154.647	2.1
3	japonês	3	112	2	110	33	874.122	6.68
4	de	4	1105	897	208	113	554.109	1.099
5	em	5	446	345	101	93	538.683	1.899
6	no	6	268	176	92	58	372.594	2.084
7	translúcido	7	38	0	38	1	270.395	6.287
8	suporte	8	91	85	6	42	266.557	3.384
9	alcalino	9	42	1	41	21	262.095	5.741
10	mata	10	36	0	36	18	222.477	5.702
11	vegetal	11	33	0	33	11	191.615	5.457
12	jornal	12	28	1	27	9	159.921	5.394
13	ácido	13	36	1	35	18	156.670	4.45
14	neutro	14	28	1	27	17	140.650	4.925
15	ingres	15	16	1	15	7	121.005	6.55
16	artesanal	16	20	0	20	9	112.481	5.336
17	a	17	132	15	117	51	102.292	-1.093
18	ao	18	112	100	12	45	94.727	1.553
19	sobre	19	77	73	4	28	89.117	1.874
20	da	20	42	1	41	20	78.483	-1.599
21	produzido	21	18	0	18	12	77.014	4.398

Fonte: *Corpus Papel* e AntConc.

A combinação dessas ferramentas, para observar o vocabulário do *corpus*, vai ampliando e, ao mesmo tempo, filtrando a lista de candidatos a termos. Assim, além de *papel japonês*, encontramos outros termos

frequentes, como *papel translúcido*, *papel alcalino*, *papel vegetal*, *papel neutro*, *papel artesanal*, *papel jornal* etc. Essa filtragem, que se dá principalmente pela frequência lexical, pode ser acompanhada de um segundo critério, o de **chavidade**, isto é, aquele que identifica quais são os “termos-chave” verdadeiramente típicos da área. Isso é feito por meio da comparação do léxico do *corpus* de trabalho (especializado) com o léxico de um *corpus* de referência, mais geral e maior (jornalístico, por exemplo). Quando o programa cruza esses dois *corpora*, ele vai destacar, pelo índice de chavidade, as unidades lexicais representativas no *corpus* de trabalho, porque são raras (menos frequentes) ou inexistentes no *corpus* mais geral, mostrando finalmente o que parece ser próprio ao vocabulário especializado. O índice de frequência, nesse caso, é relativo ao número total de palavras de cada *corpus*, apontando para unidades que são fortemente candidatas a termo. Ilustramos, a título de exemplo, as primeiras dez unidades lexicais frequentes mais prototípicas do *Corpus Papel* (Figura 4.6), por sua pouca frequência ou mesmo inexistência no *corpus* de referência.

Figura 4.6 – Primeiras dez palavras-chave do *Corpus Papel*

SINGLE-WORDS ✓				
reference corpus: Portuguese Web 2011 (ptTenTen11)				
	Word	Frequency?		
		Focus	Reference	
1	encadernação	920	4,455	...
2	arquivístico	491	4,128	...
3	conservação-restauração	223	57	...
4	desacidificação	193	57	...
5	lignina	202	1,374	...
6	acervo	3,867	117,318	...
7	condicionamento	412	8,621	...
8	restaurador	290	6,588	...
9	pergaminho	276	6,446	...
10	restauração	1,921	72,420	...

Fonte: *Corpus Papel* e Sketch Engine.

Nessa lista, “*Focus*” representa a frequência dessas palavras no *corpus* de estudo, enquanto “*Reference*” representa a sua frequência no *corpus* de referência. Para compreender como funciona o conceito de chavicidade, é importante levar em conta o tamanho dos *corpora* que estão sendo contrastados. O *corpus* de estudo em questão, *Corpus Papel*, tem aproximadamente 1 milhão de palavras. O *corpus* de referência, por sua vez, tem aproximadamente 20 bilhões de palavras.

Muito embora algumas palavras como *encadernação*, *arquivístico* e até mesmo *acervo* tenham muito mais ocorrências no *corpus* de referência, o cálculo da chavicidade se faz pela representatividade dessas palavras em cada *corpus*. Ou seja, no caso de *encadernação*, 920 ocorrências em um *corpus* de 1 milhão de palavras significa que a palavra é, proporcionalmente, muito representativa da temática do *corpus*, uma vez que, quando

contrastada à sua frequência no *corpus* de referência, ela aparece registrada 4.555 vezes, porém, em um universo de mais de 4 bilhões de palavras.

Para além dessas estratégias de extração lexical semiautomática, comparação e observação dos contextos, com base sobretudo na frequência de uso, cabe chamar a atenção para um último critério de seleção de termos: a presença de **contextos definitórios**. Definições de palavras e expressões encontradas nos textos que compõem o *corpus* são um indício patente de seu sentido especializado, apontando para o conceito elaborado na área ou por especialista(s), portanto, para seu uso terminológico. A partir de definições, podemos identificar, inclusive, palavras de ocorrência única no *corpus*, conhecidas como *hápax**. Deparamo-nos com essas definições quando da leitura dos textos certamente, mas também é possível identificá-las através dos programas de extração, o que é uma grande vantagem quando trabalhamos com *corpora* textuais. Isso pode ser feito de modo não sistemático, quando observamos, por exemplo, uma lista de concordância de determinada unidade lexical; ou de modo mais sistemático, quando criamos, a partir de observações prévias do gênero textual, um método de busca de contextos definitórios, através, por exemplo, de expressões que introduzem explicações ou reformulações (paráfrases), tais como *isto é, ou seja, em outras palavras, a saber*, ou de um léxico que explicita a ação de definir, como *ser, significar, querer dizer, definir/definição, compreender/compreensão* etc.

A título de exemplo, vemos a seguir duas situações distintas: na primeira, com conteúdo extraído do *corpus* de Direito Ambiental Internacional, definições compõem uma seção do próprio gênero textual, portanto, aparecem de maneira explícita nos textos (figura 4.7); na segunda, com conteúdo extraído do *Corpus* Papel, ilustramos uma busca sistematizada de contextos definitórios (figuras 4.8 e 4.9).

* Casos de *hápax*, que contrariam o critério de frequência, são comuns em situações de neologismo, quando termos são cunhados pelos especialistas para dar conta de um novo conceito. Algumas áreas do conhecimento são notórias pela presença de *hápax*, como é o caso da psicanálise, cujos neologismos foram estudados, da perspectiva da tradução, por Reuillard (2007).

Figura 4.7 – Definições em textos legislativos

PARTE I INTRODUÇÃO

ARTIGO 1 Termos utilizados e âmbito de aplicação

1. Para efeitos da presente Convenção:

- 1) 'Área' significa o leito do mar, os fundos marinhos, e o seu subsolo além dos limites da jurisdição nacional;
- 2) 'Autoridade' significa a Autoridade Internacional dos Fundos Marinhos;
- 3) 'atividades na Área' significa todas as atividades de exploração e aproveitamento dos recursos na Área;
- 4) 'poluição do meio marinho' significa a introdução pelo homem, direta ou indiretamente, de substâncias ou de energia no meio marinho, incluindo os estuários, sempre que a mesma provoque ou possa vir provocar efeitos nocivos, tais como danos aos recursos vivos e à vida marinha, riscos à saúde do homem, entrave às atividades marítimas, incluindo a pesca e as outras utilizações legítimas do mar, alteração da qualidade da água do mar, no que se refere à sua utilização, e deterioração dos locais de recreio;
- 5) a) 'alijamento' significa:
 - i) qualquer lançamento deliberado no mar de detritos e outras matérias, a partir de embarcações, aeronaves, plataformas ou outras construções;
 - ii) qualquer afundamento deliberado no mar de embarcações, aeronaves, plataformas ou outras construções;

Fonte: Convenção das Nações Unidas sobre o Direito do Mar. Decreto nº 1.530, de 22 de junho de 1995.

Figura 4.8 – Busca por contextos definitórios do candidato a termo *papel*

29	ptPP140.txt	papel é	uma folha formada, seca e acabada, de uma
30	ptPP148.txt	papel é	uma lâmina que se obtém a partir da
31	ptPP085.txt	papel é	uma propriedade especialmente importante para restauradores porque e
32	ptPP006.txt	papel é	constituído de fibras de trapo de linho que
33	ptPP154.txt	papel é	constituído de fibras vegetais de diferentes origens, compostas
34	ptPP135.txt	papel é	constituído do mesmo suporte do material gráfico. O
35	ptPP091.txt	papel é	sensível a agentes deteriorantes como umidade, pragas, calor,
36	ptPP099.txt	papel é	sensível a agentes deteriorantes como umidade, pragas, calor,
37	ptPP115.txt	papel é	sensível à oxidação, portanto, será oxidado em pouco

Fonte: *Corpus* Papel e AntConc.

Além de auxiliarem a evidenciar a especificidade do sentido da palavra de busca, ou seja, o conceito, as informações contidas nos contextos definitórios podem ser usadas tais quais na ficha do glossário ou servir de base para a redação de definições terminológicas. Por exemplo, percebemos que, para restauradores, *papel* “é uma lâmina obtida a partir de [...]” (linha 30 da Figura 4.8), pode ser “constituído de fibras vegetais de diferentes origens” (linha 33), além de ser “sensível a agentes deteriorantes como umidade, pragas, calor [...]” (linhas 35 e 36). Essas informações são preciosas para tradutores e intérpretes, na medida em que alimentam sua bagagem de conhecimento sobre determinada área ou assunto, preenchendo, muitas vezes, os não ditos dos textos especializados, e serão essenciais para o estabelecimento da equivalência terminológica, tratada no Capítulo 6.

Figura 4.9 – Busca de termos e definições pela estrutura “*definid** como”

	File	Left Context	Hit	Right Context
1	ptPP159.txt	o, cujo prazo de guarda é	definido como	permanente, são recolhidos para o DAG. Este acervo está armazenad
2	ptPP159.txt	o, cujo prazo de guarda é	definido como	permanente, são recolhidos para o acervo do DAG. 108 Este acervo es
3	ptPP158.txt	ca; * Consolidação, que é	definida como	a aplicação ou regeneração de materiais consolidantes em um artefatu
4	ptPP154.txt	nde a conservação, que é	definida como	aquelas ações, que envolvendo o mínimo de intervenção técnica, são
5	ptPP159.txt	lovo, onde o Patrimônio é	definido como	o conjunto de bens móveis e imóveis existentes no país cuja conserv
6	ptPP159.txt	. (2008) a rinite alérgica é	definida como	uma inflamação da mucosa nasal, induzida pela exposição a alérgeno
7	ptPP142.txt	ica brasileira, o arquivo é	definido como:	um conjunto de documentos produzidos e recebidos por órgãos públ
8	ptPP100.txt	s; assim, a conservação é	definido como	um “conjunto de procedimentos e medidas contra agentes de deterior
9	ptPP106.txt	ologia de bem cultural, é	definido como	patrimônio a ser preservado, sob a égide da noção de excepcionalida
10	ptPP142.txt	vística, o termo original é	definido como “	a forma pela qual um documento, feito por vontade expressa de seu
11	ptPP115.txt	.NBR 14348:1999) O pH é	definido como	a concentração de íons hidrogênio (H+) em uma solução. Para que un
12	ptPP154.txt	UNESCO, a preservação é	definida como,	a soma das medidas necessarias para garantir a acessibilidade perma

Fonte: *Corpus* Papel e AntConc.

Já a Figura 4.9 ilustra contextos definitórios de termos diferentes, tais como *consolidação* (linha 3), *Patrimônio* (linha 5), *conservação* (linha 8), *preservação* (linha 12), para ficar nos termos exclusivamente da área abordada. A forma *definid** usada para a busca, com o uso de asterisco, ou seja, com a palavra truncada, serve para identificar variações morfológicas.

Nesse caso, são incluídas as variações de gênero e número, tal como *definido/a/os/as*. O capítulo 3 apresentou essas possibilidades.

A identificação e a seleção dos termos (ou candidatos a termos) podem também constituir o ponto de partida para a identificação de outras unidades lexicais especializadas, como as UFEs, que passamos a abordar a seguir. Essas são tão ou mais importantes para o trabalho de pesquisa e documentação do tradutor e/ou do intérprete quanto os termos, sobretudo por apontarem para um jeito de dizer específico das comunidades de saber.

Depois dos termos, como identificar unidades fraseológicas especializadas?

A partir da noção de termo, é possível identificar construções maiores, que transcendem o espaço discursivo do termo. Ao redor dele orbitam outras palavras que, juntas, formarão estruturas que podem expressar **ações e processos especializados** de um campo de saber ou, ainda, modos de dizer característicos de uma área.

Portanto, além dos termos, outras unidades lexicais dos textos de especialidade também são responsáveis pela representação e transmissão de conhecimento. Mas por que é tão importante identificar e registrar essas unidades em produtos terminográficos?

Conhecer as unidades fraseológicas de uma determinada área é de grande importância nas práticas de produção de textos especializados – especialmente na sua tradução –, uma vez que nem sempre se encontram registrados em glossários e bases terminológicas os equivalentes das palavras que coocorrem com um determinado termo. Bevilacqua (2002, p. 140) explica que, muitas vezes, as dúvidas dos redatores e tradutores se referem não necessariamente aos termos em si, mas às unidades que aparecem junto deles. Por exemplo, a estrutura *fazer/contratar uma apólice* tem como equivalente em espanhol *suscribir una póliza*. Ou seja, não é propriamente chegar à relação de equivalência dos termos *apólice* (português)/*póliza* (espanhol) que constitui uma dificuldade para o tradutor, mas encontrar as palavras que se combinam com esses termos em uma e outra língua.

A equivalência entre *fazer/contratar* (português) e *suscribir* (espanhol), no contexto mencionado, se dá, justamente, pela combinação estável entre esses verbos e os termos *apólize/póliza*, formando, respectivamente, fraseologias especializadas de cada língua, como explicado no Capítulo 1.

Cada área de conhecimento estará mais ou menos marcada pela presença de UFEs de diferentes tipos, tendo em vista suas especificidades discursivas e os tipos de conceitos e de processos que se pretendem expressar. Por isso, para a tarefa de identificação de UFEs de uma determinada área, é importante, em primeiro lugar, delimitar estrutural e conceitualmente as unidades que se pretende extrair e selecionar, sempre levando em conta as necessidades que deverão ser supridas pelo produto terminográfico em questão, conforme visto no Capítulo 2.

Tomando como referência produtos já compilados pelo grupo Termisul, podemos afirmar que uma base terminológica multilíngue que busca representar, por exemplo, a linguagem jurídica e oferecer equivalentes em outras línguas deverá registrar UFEs que são prototípicas dessa linguagem, como é o caso das fórmulas. Já bases de outras áreas, como da Conservação e Restauração, cujos textos versam sobre métodos e práticas de salvaguarda e recuperação do patrimônio cultural, deverão oferecer aos consulentes UFEs semelhantes às colocações.

Como vimos na tarefa de reconhecimento de termos, para identificar e extrair UFEs de um *corpus* textual, também é necessária uma etapa semiautomática de extração de estruturas candidatas a UFEs. Conforme foi explicado no Capítulo 3, no Termisul, utilizamos tanto o programa *AntConc* quanto o *Sketch Engine* para identificar UFEs¹.

O primeiro passo é identificar estruturas candidatas a UFEs eventivas. Para isso, é necessário conhecer o termo (NT) a partir do qual a UFE se formará. Esse NT será a palavra-chave da busca por verbos e nominalizações que apareçam junto desse termo. A figura 4.10 ilustra a busca por

¹ O *Sketch Engine* oferece recursos que otimizam a tarefa de recuperação de dados (ver capítulo 3), portanto pautamos nossa explicação a partir desse programa e da sua ferramenta *Word Sketch*.

candidatos a UFEs a partir do NT *acervo*, organizados de acordo com seu índice de frequência de ocorrência no *corpus*².

Figura 4.10 – Candidatos a UFE do termo *acervo*

verbo + acervo	verbo + acervo
compor 28 ... que compõem o acervo	constituir 9 ... de itens que constituem o acervo
abrigar 19 ... que abrigam acervos	proteger 8 ... de proteger o acervo
preservar 16 ... preservar o acervo	cuidar 7 ... cuidar do acervo
afetar 14 ... afetam acervos	integrar 7 ... e livros que integram o acervo desse museu
possuir 14 ... possuem acervos	ambientar 7 ... em ambientes de acervos
trabalhar 11 ... trabalham diretamente com o acervo	atacar 7 ... atacam o acervo
manter 11 ... manter os acervos	danificar 6 ... danificam os acervos

Fonte: *Corpus* Papel e Sketch Engine.

Muito embora todos esses candidatos sejam estruturas formadas por **verbo + termo*** (substantivo), isso não significa, necessariamente, que conformam UFE. Isso se deve ao fato de que o verbo deve constituir um NE, expressando, assim, ações e processos especializados realizados com o NT. Além disso, deve existir entre esses elementos uma restrição de seleção combinatória que está determinada pelas especificidades do âmbito temático em que são utilizados.

² O grupo Termisul estabelece como corte de frequência para a identificação de UFEs o mesmo adotado para a identificação de termos. Dependendo do tamanho do *corpus* de estudo, esse corte pode variar entre, no mínimo, 5 e 10 ocorrências.

* É importante levar em conta que, apesar da estrutura da UFE e de seu grau de fixação, essas unidades podem aparecer intercaladas por outras palavras, por exemplo, “abrigar o seu próprio acervo”. Esse intervalo de palavras (“o seu próprio”) chama-se *span* e pode ser definido como o espaço ocupado por outras palavras entre a base e o coocorrente de uma colocação, neste caso, entre o verbo e o termo nas UFEs. Algumas ferramentas, como o *Word Sketch*, recuperam estruturas que aparecem separadas por essa janela de palavras. Outras, como o *Cluster* do *AntConc*, exigem que se façam buscas utilizando formas truncadas (para saber mais sobre formas truncadas, ver capítulo 3).

Para confirmar se esses candidatos são UFE, é necessário, em primeiro lugar, analisar as concordâncias – etapa manual do trabalho. Por exemplo, para constatar que a estrutura *preservar acervo* constitui, de fato, uma UFE, diferentemente de *possuir acervo*, é necessário olhar para os contextos de uso:

ptPP095 (...) Para *preservar os acervos* arquivísticos é necessário organizá-los e conservá-los para servir de referência, de informação, prova, testemunho e fonte de pesquisa. (...)

ptPP136 (...) As bibliotecas *possuem acervos* constituídos por material altamente combustível armazenado em prateleiras, o que cria áreas densas e, consequentemente, um ambiente vulnerável a incêndios. (...)

A partir da análise dos contextos, observa-se que o candidato *possuir acervo* não expressa uma ação especializada da área da Conservação e Restauração, como é o caso de *preservar acervo*, processo que envolve fazeres específicos do âmbito temático, como a organização e conservação dos materiais que fazem parte dos acervos.

No caso da identificação das UFEs do tipo fórmulas, pode-se utilizar a ferramenta *N-grams* (tanto do *Sketch Engine*, quanto do *AntConc*). Uma vez que o *corpus* de estudo é constituído por textos jurídicos, os quais estão marcados discursivamente por essas estruturas complexas, próximas ao nível da frase, e que apresentam certo grau de fixação, a busca por *n-gramas* permite chegar a construções candidatas a UFEs desse tipo.

Por exemplo, fazendo uma busca por *n-gramas* de uma extensão de 3 a 6 palavras, com uma frequência mínima de 2 ocorrências e um **range** mínimo de 2 textos, identificamos estruturas complexas como *de acordo com* (figura 4.10).

* *Range*, ou distribuição, é um filtro que indica a distribuição de uma palavra ou expressão em textos diferentes. Por exemplo, a indicação de *range 2* significa que a palavra ou expressão deve ocorrer, no mínimo, em dois textos diferentes. Evita-se, assim, que se recupere uma construção que tem alta frequência em um único texto e autor, o que seria pouco representativo da área.

Figura 4.11 – n-gramas candidatos a formadores de UFEs do tipo fórmulas

Concordance		Concordance Plot		File View		Clusters/N-Grams		Collocates		Word List		Keyword List			
Total No. of N-Gram Types				142128				Total No. of N-Gram Tokens				584930			
Rank	Freq	Range	N-gram												
1	840	207	nç de de												
2	771	180	do meio ambiente												
3	638	163	lei nç de												
4	517	160	lei nç de de												
5	458	123	de que trata												
6	366	124	de acordo com												
7	329	201	presidente da república												
8	310	63	do ministério da												
9	306	110	a que se												
10	302	114	meio ambiente e												
11	288	114	que se refere												
12	284	92	da lei nç												
13	281	249	data de sua												
14	281	82	no caso de												
15	280	280	brma name title												
16	280	280	name brma name												
17	280	280	...												

Search Term		<input checked="" type="checkbox"/> Words	<input type="checkbox"/> Case	<input type="checkbox"/> Regex	<input checked="" type="checkbox"/> N-Grams	N-Gram Size	
		Advanced			Min. 3	Max. 6	
Start	Stop	Sort		Min. Freq.		Min. Range	
Sort by		<input type="checkbox"/> Invert Order	Search Term Position		2	2	
Sort by Freq		<input checked="" type="checkbox"/> On Left	<input type="checkbox"/> On Right				

Fonte: Corpus Legis e AntConc.

Em seguida, deve-se buscar as concordâncias geradas a partir dessa expressão, a fim de identificar possíveis fórmulas da área.

Figura 4.12 – Concordâncias formadas a partir do candidato *de acordo com*

The screenshot displays the AntConc software interface. At the top, the menu bar includes 'Concordance', 'Concordance Plot', 'File View', 'Clusters/N-Grams', 'Collocates', 'Word List', and 'Keyword List'. Below the menu, the search term 'de acordo com' is entered in the search bar. The search window size is set to 60. The results are displayed in a list with columns for 'Hit', 'KWIC', and 'File'. The KWIC column shows the search term highlighted in red in the original image. The files listed in the 'File' column are various .txt files, such as 'brMA011.txt', 'brMA012.txt', etc.

Fonte: Corpus Legis e AntConc.

Nesse caso, o concordanciador recupera expressões como *de acordo com a lei* ou *de acordo com esta lei*, ou, ainda, variações dessa fórmula, como *de acordo com a legislação* ou *de acordo com o inciso X do artigo Y da lei*. Essas expressões são um indicativo de uma fórmula prototípica mais ou menos fixa e frequente em gêneros discursivos do Direito.

Fórmulas como essas costumam apontar para um outro aspecto que, ao lado de termos e fraseologias especializadas, caracteriza a expressão das comunidades de saber: a *fraseologia de gênero* (Tutin, 2007). Trata-se de combinações lexicais (normalmente colocações), padrões léxico-gramaticais ou fórmulas comuns a um gênero de discurso, mas transversais a diferentes áreas de conhecimento ou temáticas. Assim, por exemplo, no *artigo científico*, gênero privilegiado para a divulgação das ciências de maneira geral, fraseologias como *levantar a hipótese*, *descrever os procedimentos*, *o objetivo deste trabalho [artigo]* etc. remetem ao discurso sobre os objetos e ao fazer científico (léxico metacientífico), bem como para a relação entre interlocutores (Kilian; Loguercio, 2015; Loguercio; Kilian, 2017; Loguercio,

2019; Loguercio, 2020). Já no *tratado internacional* (e seus subgêneros, como *acordo*, *convênio*, *protocolo de cooperação*, entre outros), texto legislativo oriundo da celebração de um acordo entre dois ou mais países com efeitos jurídicos no plano internacional, é comum encontrar fraseologias como: *para os efeitos deste [tratado, acordo, convênio, protocolo], este [tratado, acordo, convênio, decreto] entra em vigor, as Partes acordam que, os Estados signatários [comprometer-se a, cooperar, reconhecer...]* (Bevilacqua; Maciel, 2018). Nesses casos, algumas das estratégias de busca semiautomática que apresentamos aqui, como o uso da ferramenta *N-grams* e de concordanciador, também facilitam a compilação.

Vale ressaltar, por fim, que cada trabalho de descrição da linguagem de uma comunidade de saber, com seus propósitos específicos e seu contexto de realização e aplicação, é único em relação às unidades lexicais que pretende descrever, bem como à definição de seus procedimentos metodológicos. Tomadas as decisões prévias, há uma etapa de estudo exploratória, normalmente de familiarização com a área temática e os gêneros discursivos, seguida de testagem de estratégias e ajustes das buscas e da análise lexical, que antecede a extração e seleção das unidades lexicais. Neste capítulo, apresentamos um breve panorama dos tipos de unidades que podem ser descritos para auxiliar o tradutor ou intérprete em suas pesquisas, por meio de procedimentos e segundo princípios que costumam ser adotados pelo Grupo Termisul em seus trabalhos de abordagem linguístico-textual. Trata-se mais de pistas e ideias do que de receitas metodológicas.

Atividades: Seleção de unidades terminológicas

1. Leia o trecho abaixo do *Corpus Papel*, da área de Conservação e Restauração, e identifique o termo que está sendo definido. Leve em consideração as estruturas de contextos definitórios e o número de vezes (frequência de ocorrência) que esse termo aparece no trecho.

Na legislação arquivística brasileira, o arquivo é definido como: um conjunto de documentos produzidos e recebidos por órgãos públicos, instituições de caráter público e entidades privadas, em decorrência de exercício de atividades específicas, bem como por uma pessoa física, qualquer que seja a natureza do suporte da informação ou a natureza dos documentos (Brasil, 1991). Portanto, o arquivo não se define pela forma dos documentos ou por sua origem, mas pela razão para que foram criados e por sua forma de acumulação orgânica. Para Schellenberg (1974, p. 15-17) os elementos que definem os arquivos podem ser resumidos em três fatores que são abstratos: o primeiro é a finalidade pela qual os materiais foram produzidos e acumulados; o segundo é pelos valores pelos quais os arquivos são preservados; e o terceiro tem relação com o que Jenkinson (1922) denominou como “custódia ininterrupta”. Sobre este ponto, o autor esclarece que teriam qualidade para serem denominados de “arquivo” os conjuntos de documentos que não podem ser questionados na sua autenticidade, não apenas legal, mas evidencial, quando estes foram constituídos por meio de uma trajetória ininterrupta de custódia responsável (Jenkinson, 1922, p. 10 e 11).

(*Corpus* Papel PT – Termisul)

2. A partir do termo identificado no exercício 1, assinale, no trecho abaixo, quais UFEs são formadas a partir dele e classifique de que tipo elas são (colocações ou fórmulas).

Verificamos que a maioria das escolas que abrigam seu próprio arquivo de documentos o faz em ambientes que não têm as mínimas condições de acondicionamento. Na maior parte das vezes, esses arquivos ficam em locais sem ventilação, tomados por umidade e fungos. [...] Há um item relativo à Seção de Assistência Técnica a quem cabe “elaborar normas para a organização, conservação, catalogação e microfilmagem de arquivos”.

(Adaptado de *Corpus* Papel PT – Termisul)

3. Leia o trecho textual abaixo e identifique: a área de conhecimento em que se insere, o assunto tratado e o gênero textual.

Nas últimas décadas, uma ampliação nas discussões em torno do patrimônio permitiu que outras categorias de bens culturais começassem a ganhar uma maior visibilidade no cenário nacional, crescendo o entendimento da necessidade de se contemplar e de se reconhecer a diversidade da história e da cultura brasileira além das múltiplas memórias formadoras de nossa nacionalidade. Profissionais de diferentes áreas e formações têm se envolvido na discussão, pois sendo o Patrimônio Cultural matéria de conhecimento interdisciplinar, [...], necessita dos vários olhares das diversas áreas do saber no estudo de suas singularidades.

A ideia de democratização do patrimônio implica, qualquer que seja a perspectiva, no fato de que o Estado não deve ser o único ator social a se envolver com a preservação do patrimônio cultural de uma sociedade. [...]. Em relação aos acervos em suporte de papel (compreendidos pelas coleções bibliográficas, documentais e obras de arte em suporte de papel), é interessante observar que, em termos quantitativos, eles representam um dos maiores estoques informacionais e culturais da nação, ocupando salas de arquivos, bibliotecas e museus públicos. [...]

No presente trabalho, optou-se por aprofundar as pesquisas em fungos já que, além de serem mais comuns em ambientes de arquivos, eles apareceram em um número bem mais representativo quando da análise microbiológica realizada na época em que ocorreu a infiltração d'água no acervo do Arquivo Histórico de Joinville.

Os fungos são microrganismos que podem acometer todo tipo de acervo, independentemente da natureza material que o constitui. A presença ou suspeita de contaminação por fungos em um acervo de papel requer muita atenção, uma vez que podem provocar manchas e destruição da informação em textos e gravuras que podem ser irreversíveis, além dos possíveis quadros alérgicos que podem acometer pessoas expostas a esse tipo de organismo. [...].

A eliminação de agentes microbiológicos em bens materiais em suporte de papel apresenta dois problemas: ou são agressivos ao documento ou são tóxicos ao ser humano. A noção das dificuldades de tratamento desse tipo de material, a escassez de trabalhos específicos nessa área e a necessidade de ações que evitem a perda de acervos afetados por esse tipo de contaminação orientaram esta pesquisa que pretende, por meio de uma investigação mais detalhada acerca de microrganismos em ambientes de arquivos, tendo como estudo de caso o Arquivo Histórico de Joinville (AHJ), estudar métodos de tratamento para eliminação de agentes fúngicos em documentos em suporte de papel. A pesquisa busca viabilizar uma proposta de intervenção que elimine o agente agressor sem danificar o suporte do documento.

(Adaptado de *Corpus Papel PT* – Termisul)

4. Com base nas informações anteriores e em outras presentes no excerto, relativas à área e ao assunto específico do trabalho, desenhe uma pequena árvore de domínio ou um mapa conceitual.

5. Quais unidades lexicais e fraseológicas remetem, mais precisamente, ao gênero textual? Selecione algumas.

6. Que tipo de informação as unidades elencadas no exercício anterior introduzem no texto? Associe as colunas.

a. no presente trabalho optou-se por	() material de análise
b. esta pesquisa pretende	() objetivo aplicado, prático, propositivo
c. por meio de uma investigação	() tipo de método
d. a pesquisa busca	() tema geral do artigo
e. estudo de caso	() objetivo específico, de testagem

Capítulo 5 – A ficha terminológica

Cleci Regina Bevilacqua

Denise Regina de Sales

Márcia Moura da Silva

O projeto terminográfico inclui o uso de fichas para registro dos dados das unidades selecionadas – termos e UFEs, conforme referido no capítulo 1. No desenho da ficha, levamos em conta as decisões prévias (ver capítulo 2) específicas de cada projeto. Por isso, neste capítulo, em vez de apresentar um modelo padrão, fornecemos noções gerais sobre o tema e exemplos concretos de fichas utilizadas em projetos do Termisul.

O que é a ficha terminográfica e para que serve?

Na literatura da terminografia, podemos encontrar os termos “ficha terminográfica”, “ficha terminológica” e “ficha de trabalho” para designar o modelo de registro de dados das unidades selecionadas. Se antigamente as fichas eram preenchidas em cartões ou folhas avulsas, hoje em dia os modelos são criados e armazenados no computador. Como apontam Krieger, Maciel e Finatto (2001), a transição entre esses suportes demandou estudos em conjunto com profissionais da informática, visto que o registro e o tratamento de dados passaram a ser feitos com o auxílio do computador. Conforme as autoras:

Na operacionalização da pesquisa terminológica e no seu posterior aproveitamento na elaboração do produto terminográfico, a equipe cuidou de lançar mão dos recursos oferecidos pela informática. Desde os primeiros momentos das atividades, o registro dos dados e seu tratamento foi feito com o auxílio do computador. [...] Buscávamos uma ferramenta que nos permitisse desenhar as fichas terminológicas sob medida, isto é, com campos de número

ilimitado e de extensão livre, para o registro dos dados relativos ao termo (Krieger; Maciel; Finatto, 2001, p. 331).

A ficha terminográfica visa fornecer informações relativas à entrada (termo ou UFE) e aponta para a estrutura interna dos verbetes, ou seja, o que se conhece como microestrutura do dicionário. Por isso, as fichas são um dos itens essenciais para a elaboração de um glossário, dicionário ou base de dados. De forma geral, elas incluem:

- Entrada: registro do termo e/ou UFE selecionado, a respeito do quais são dadas as demais informações constantes na ficha; nos produtos do Termisul, a entrada principal é o termo em português;
- Categoria gramatical, gênero e número;
- Fonte da entrada: referência do texto de onde foi coletado o termo ou UFE;
- Definição: explicação do significado do termo; em geral, para as UFEs, não são apresentadas definições, posto que se referem a ações e processos ou a fórmulas que possuem funções específicas nos textos;
- Fonte da definição: referência do texto de onde foi retirada a definição ou se foi criada pelo grupo de trabalho;
- Contexto: exemplo de uso do termo ou UFE;
- Fonte do contexto: referência do texto de onde foi retirado o contexto;
- Remissivas: indicação de entradas para formas sinônimas ou entradas que fornecem informações complementares; podem estar indicadas por “Ver também”, “Outra denominação”;
- Equivalentes: termos nas línguas estrangeiras correspondentes ao termo ou UFE em português;
- Notas: observações que complementam as informações dadas para especificar a que se refere a entrada, para esclarecer diferenças entre as línguas, entre outras possibilidades.

Como se observa, essas informações podem ser gramaticais (categoria, gênero e número), semânticas (definição, remissivas, pois podem remeter a formas sinônimas ou variantes denominativas) e pragmáticas (contextos de uso, notas). É para isso que serve a ficha, para definir

e organizar as informações pertinentes a cada projeto terminográfico. Assim, conforme o objetivo, isto é, a função do dicionário, além de seus usuários e temática com a qual se trabalha, se definem os campos da ficha.

Entretanto, com a prática de trabalho terminológico no Termisul, ao longo do tempo alguns padrões foram se consagrando como os melhores ou mais adequados para os nossos projetos. De modo geral, usamos a forma gráfica básica ou canônica das entradas, isto é, infinitivo para os verbos e singular masculino para os substantivos e adjetivos. Contudo, nas obras terminográficas, caso se identifique no *corpus* de estudo que termos e UFEs ocorrem apenas no plural, é possível registrá-las dessa forma. É o caso de águas subterrâneas no *Dicionário de Direito Ambiental* (Krieger *et al.*, 1998). A variação denominativa (por exemplo, *Controle Integrado de Pragas e CIP*) e a forma sinônima (por exemplo, *resíduos* e *lixo*) são apontadas com remissivas na microestrutura e aparecem como nova entrada na macroestrutura. E quaisquer informações adicionais consideradas relevantes para descrever particularidades do termo ou da UFE são fornecidas no campo “Notas”.

Além disso, as entradas podem estar registradas em determinado padrão gráfico, sendo mais frequente o uso de fonte maiúscula e em negrito. Do mesmo modo, os nomes dos demais campos podem estar grafados com fonte em negrito para chamar a atenção do consulente. Você pode observar esses aspectos nos exemplos de fichas apresentados adiante.

Como se estrutura a ficha terminográfica

Para ilustrar as diferentes possibilidades de fichas, trazemos três modelos utilizados em projetos do Termisul. O primeiro é do *Dicionário de Direito Ambiental: terminologia das leis do meio ambiente*, de 2008; o segundo é da base *Combinatórias Léxicas Especializadas da Linguagem Legal, Normativa e Científica* (PROJECOM), de 2016 e o terceiro é do projeto *A linguagem do patrimônio cultural brasileiro: conservação dos bens culturais móveis em suporte papel* (Projeto Papel), em fase de revisão para publicação.

* Nome do ambiente de programação que agrega uma IDE (Integrated Development Environment – Ambiente de Desenvolvimento Integrado) com o compilador da linguagem de programação Object Pascal.

Talvez não se note, ao olharmos esses modelos, que o Termisul foi consolidando o uso de recursos informáticos em seus projetos com o passar do tempo. No entanto, os avanços tecnológicos não só facilitaram a coleta de um número crescente de dados, sobretudo com a criação da internet, como também tornaram mais rápido e preciso seu processamento com ferramentas criadas especificamente para esse fim (conforme explicado no capítulo 3).

No caso do Termisul, incluímos nos projetos bolsistas com conhecimento específico de linguagens de programação, como PHP e MySQL. Contudo, caso não seja possível dispor de um informático, a base pode ser construída em Excel, em ferramentas disponíveis *on-line* ou nos próprios recursos de construção de glossários das memórias de tradução.

Ficha do projeto do *Dicionário de Direito Ambiental*

Em relação ao primeiro modelo de ficha, usando a ferramenta **Delphi***, da Borland, a equipe gerou tanto o dicionário impresso quanto o eletrônico, a partir de 2000 fichas terminológicas, correspondentes às entradas. O desenho das fichas terminológicas do *Dicionário de Direito Ambiental* é apresentado nas figuras 5.1 e 5.2.

Figura 5.1 – Campos da ficha em formato eletrônico do *Dicionário de Direito Ambiental (TermDic)*

Campo	Nome do Campo	Conteúdo do Campo
1	Entrada	Termo em português
2	Assunto	Categorização de subáreas temáticas na legislação ambiental
3	Referência da fonte	Identificação do documento fonte
4	Ementa ou texto legal	Breve resumo do conteúdo do diploma legal
5	Equivalente em inglês	Termo em inglês com código e página da referência
6	Equivalente em espanhol	Termo em espanhol com código e página da referência
7	Definição	Descrição do conceito referenciado pelo termo na área temática da legislação ambiental.
8	Referência da definição	Os códigos LgBR , LgRS , LgPOA e LgP indicam que a definição foi encontrada na Legislação Federal Brasileira, na Legislação do Rio Grande do Sul, de Porto Alegre ou de Portugal. Outros códigos se referem às fontes de consulta. A ausência de informação indica que a definição é de autoria do Projeto Termisul.
9	Observações	Notas técnicas, linguísticas ou científicas relevantes para a compreensão do conceito ou o uso do termo.
10	Remissiva	Termo relacionado ao termo do campo Entrada
11	Remissiva	Termo relacionado ao termo do campo Entrada
12	Remissiva	Termo relacionado ao termo do campo Entrada
13	Remissiva	Termo relacionado ao termo do campo Entrada
14	Remissiva	Termo relacionado ao termo do campo Entrada

Fonte: Krieger; Maciel, 2001, p. 332.

Figura 5.2 – Exemplo de entradas do *Dicionário de Direito Ambiental* em formato papel

<p style="text-align: center;"><i>Dicionário de direito ambiental</i></p> <p>MINERAL NUCLEAR LgBR LEI 4118 de 27/08/62, art. 2º. nuclear mineral WTO mineral nuclear WMED Mineral que contenha em sua composição um ou mais elementos nucleares. LgBR ⇒ ELEMENTO NUCLEAR; MINÉRIO NUCLEAR; SUBPRODUTO NUCLEAR.</p> <p>MINÉRIO NUCLEAR LgBR LEI 4118 de 27/08/62, art. 2º. nuclear ore WUSHR mineral nuclear WSECYT Concentração natural de mineral nuclear na qual o elemento ou elementos nucleares ocorrem em proporção e condição tais que permitam sua exploração econômica. LgBR ⇒ ELEMENTO NUCLEAR; MINERAL NUCLEAR; SUBPRODUTO NUCLEAR.</p> <p>MINIFÚNDIO LgBR LEI 4504 de 30/11/64, art. 4º, IV. small landholding WFAO minifundio WINTA Imóvel rural de área e possibilidades inferiores às da propriedade familiar. LgBR ⇒ EMPRESA RURAL; LATIFÚNDIO; MÓDULO RURAL; PROPRIEDADE FAMILIAR.</p> <p>MINISTÉRIO PÚBLICO LgBR CF, art. 127. Ministry of Justice WUSAID Ministerio Público PNU92:449 Instituição permanente, essencial à função jurisdicional do Estado, incumbendo-lhe a defesa da ordem jurídica, do regime democrático e dos interesses sociais e individuais indisponíveis. LgBR ◇ <i>Entre suas atribuições compete-lhe promover o inquérito civil e a ação civil pública, para a proteção do patrimônio público e social, do meio ambiente e de outros interesses difusos e coletivos.</i> ⇒ AÇÃO CIVIL PÚBLICA; AÇÃO POPULAR; COMPETÊNCIA DO MUNICÍPIO; MINISTÉRIO PÚBLICO FEDERAL; PATRIMÔNIO PÚBLICO.</p> <p>MINISTÉRIO PÚBLICO FEDERAL Federal Department of Justice MEL 94:145</p>	<p>Ministerio Público Federal WLANETA Ramo do Ministério Público da União que atua judicialmente junto ao Supremo Tribunal Federal, ao Superior Tribunal de Justiça, à Justiça Federal Comum, ao Tribunal Superior Eleitoral, aos Tribunais Regionais Federais, e à Justiça Federal de 1ª Instância, ao qual são cometidas as mesmas atribuições institucionais definidas na Constituição Federal para os Ministérios Públicos dos Estados e para os demais ramos do Ministério Público da União, respeitadas as áreas de competência em que cada Instituição atua. ⇒ MINISTÉRIO PÚBLICO.</p> <p>MISTICETÁCEOS LgBR DEC 73497 de 17/01/74. baleen whale WEPA misticetos WUANTO Baleias que possuem barbatanas ou cerdas bucais ao invés de dentes, e constituem a subordem Mysticeti. ⇒ BALEIA AZUL; BALEIA CINZENTA; MEGAPTERO JUBARTE; BALEIA FRANÇA; BALEIA SEL.</p> <p>MISTURA EM TANQUE LgBR DEC 4074 de 04/01/02, art. 1º, XXV. tank mixture WEPA mezcla en tanque WGRIF Associação de agrotóxicos e afins no tanque do equipamento aplicador, imediatamente antes da aplicação. LgBR</p> <p>MISTURA OLEOSA LgBR DEC 2508 de 04/03/98, Anexo I, 2 oily mixture KIS83:327 mezcla oleosa TRE83c:226 Mistura com qualquer conteúdo de óleo. LgBR ⇒ ÓLEO COMBUSTÍVEL.</p> <p>MISTURA OLEOSA *1 LgBR LEI 9966 de 28/04/00, art. 2º, IX oily mixture KIS83:327 mezcla oleosa TRE83c:226 Mistura de água e óleo, em qualquer proporção. LgBR ⇒ ÓLEO *2</p>
--	---

Fonte: Krieger et al., 2008, p. 202.

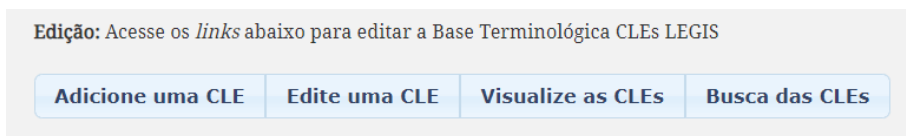
É possível observar que estão disponíveis ao usuário a entrada (termo) em ordem alfabética; a referência do texto legislativo de onde foi coletado o termo; o(s) equivalente(s) nas línguas inglesa e espanhola com as referências de onde foram coletados; a definição retirada da lei com sua fonte, ou, quando for o caso, a definição formulada pelo legislador ou especialista; informações complementares (notas) e termos que indicam relação de proximidade ou de complementariedade com o termo de entrada (indicados por \Rightarrow). Esse conjunto de informações constitui o que se denomina em Lexicografia e Terminografia de microestrutura, conforme indicamos acima.

Observe que a entrada está em fonte maiúscula e em negrito; os equivalentes estão em fonte minúscula; e os termos que indicam relação de proximidade ou de complementariedade estão em caixa-alta e iniciam por \Rightarrow ; já as remissivas para os sinônimos estão em itálico e precedidas por um losango (◊) e *Ver.* Desse modo, além de definir quais as informações fazem parte da ficha, também é preciso pensar na formatação gráfica de cada elemento para diferenciá-los na entrada e indicar para o consulente que se referem a distintas informações.

Por sua vez, a base de **Combinatórias Léxicas Especializadas da Legislação Ambiental (BDT CLEs Legis)** teve como tema, como o próprio nome diz, o estudo e a representação de Combinatórias Léxicas Especializadas (CLEs) – outra denominação para as UFEs definidas no capítulo 1 – da linguagem legal e envolveu as línguas portuguesa, alemã, espanhola, francesa, inglesa e italiana. A ideia foi enfocar as CLEs, seus contextos de uso e seus equivalentes nessas línguas, razão pela qual não incluímos, na ficha, definições, pensando nas informações que podem ser úteis principalmente para os tradutores e redatores de textos especializados. A base pode ser acessada de forma gratuita no *link* <http://www.ufrgs.br/termisul/cles/>.

A figura 5.3 ilustra os campos da interface de trabalho dos pesquisadores para incluir uma nova combinatória, editá-la, visualizá-la ou buscar uma CLE.

Figura 5.3 – Interface de trabalho



Fonte: Termisul.

Ao clicar em adicionar uma CLE, abre-se a ficha terminológica que contém os campos seguintes:

CLE: apresenta a combinatória completa.

Contexto: contexto de uso da CLE e a fonte de onde foi retirada, indicado entre parênteses;

Outras formas e seus contextos: inclui CLEs que apresentam a inserção de algum elemento linguístico ou alguma alteração morfosintática na CLE principal; também inclui seus contextos e fontes;

Ver também: *hiperlink* que remete à(s) ficha(s) de combinatória(s) sinônima(s) à CLE entrada;

Notas: informações complementares para o uso da CLE;

Equivalentes: *hiperlinks* que remetem à(s) ficha(s) da CLE correspondente nas línguas estrangeiras, com todos os campos anteriores. Para os contextos dos equivalentes recolhidos na internet, são indicados os endereços dos *sites*, os quais os consulentes podem acessar.

Na ficha de trabalho dos pesquisadores, há ainda o campo **Comentários** que inclui observações para os próprios pesquisadores com vistas à revisão das informações. Esse campo não aparecerá no módulo visível aos consulentes.

A seguir, trazemos o exemplo da ficha de *acondicionar resíduos*.

CLE: ACONDICIONAR RESÍDUOS

Língua: Português

Contexto:

Sempre que estabelecido sistema de coleta seletiva pelo plano municipal de gestão integrada de resíduos sólidos e na aplicação do art. 33, os consumidores são obrigados a: I – **acionar** adequadamente e de forma diferenciada os **resíduos** sólidos gerados. (Lei n. 12.305, de 2 de agosto de 2010; br)

Outras formas:

acionamento de resíduos

Contexto:

Nesta seção deverão estar relacionados os equipamentos e materiais de resposta a incidentes de poluição por óleo, tais como aqueles destinados à contenção, remoção e isolamento das áreas vulneráveis, limpeza de áreas atingidas, produtos absorventes e adsorventes, **acionamento de resíduos** oleosos, veículos (leves e pesados), cuja utilização está prevista pela instalação. (Resolução n. 293, de 12 de dezembro de 2001; br)

Equivalente(s) em Alemão:

[Lagerung von Abfällen](#)

Equivalente(s) em Inglês:

[package waste](#)

Equivalente(s) em Espanhol:

[acionar residuos](#)

Equivalente(s) em Francês:

[stocker des déchets](#)

Equivalente(s) em Italiano:

[stoccare residui](#)

Notas: Os resíduos podem ser oleosos, perigosos, sólidos, entre outros

Como é possível observar, os nomes dos campos estão indicados em negrito e itálico. No interior dos contextos, as CLEs são destacadas em negrito. Os equivalentes estão em forma de *hiperlinks* que remetem às fichas completas das entradas nas línguas estrangeiras.

Em 2016, iniciamos o projeto Papel. Seu objetivo foi coletar e sistematizar os termos da área que está em expansão em nosso país e precisa de obras de referência que compilem seus termos e que sirvam de subsídios tanto aos profissionais e aprendizes da área quanto aos tradutores e redatores de textos e demais interessados no tema. Foi criada então a Base

Papel. Diferentemente dos projetos anteriores do grupo, essa base incluiu termos e UFEs em uma única obra.

A ficha proposta seguiu a mesma estrutura e contém praticamente os mesmos campos que a ficha proposta para o projeto BDT CLEs Legis. A figura 5.4 ilustra a ficha de *tinta ferrogálica*.

Figura 5.4 – Ficha da entrada *tinta ferrogálica*

TINTA FERROGÁLICA
Língua: Português
Contexto: A composição da tinta ferrogálica é, basicamente, uma mistura de sulfato de ferro (II), um extrato rico em taninos de certas vesículas de origem vegetal e goma arábica em um líquido que pode ser água, vinho ou vinagre. A adição de corantes é também comum em algumas formulações. (ptPPP004)
Ver também: tinta
Equivalente(s) em Inglês: iron gall ink
Equivalente(s) em Espanhol: tinta ferrogálica
Equivalente(s) em Francês: encre ferrogallique
Equivalente(s) em Italiano: inchiostro ferrogallico
Equivalente(s) em Russo: железогалловые чернила [jelezogallovye tchernila]

Fonte: Base Papel, Termisul.

Colocamos também o exemplo de ficha de uma UFE, *degradação do papel* (figura 5.5):

Figura 5.5 –Ficha da entrada *degradação do papel*

DEGRADAÇÃO DE PAPEL
Língua: Português
Contexto: Por degradação do papel, entende-se a cisão da ligação entre moléculas de celulose, ou seja, a cisão da ligação entre monômeros de glicose. A degradação produz fibras com menor grau de polimerização, reduzindo o tamanho das moléculas formadoras das fibras de celulose, afetando, portanto, as propriedades do papel que dependem do comprimento da cadeia molecular da celulose, como a resistência mecânica. (ptPPP115)
Equivalente(s) em Inglês: paper degradation
Equivalente(s) em Espanhol: degradación de papel
Equivalente(s) em Francês: dégradation du papier
Equivalente(s) em Italiano: degrado della carta
Equivalente(s) em Russo: деградация бумаги [degradatsia bumagui] разрушение бумаги [razruchenie bumagui]

Fonte: Base Papel, Termisul.

Como comentamos anteriormente, a entrada pode incluir termos (*tinta ferrogálica*) ou UFEs (*degradação do papel*). Outra diferença em relação ao projeto anterior é que o campo *ver também* se refere não apenas ao sinônimo do termo ou UFE constante na entrada (por exemplo, *papel japonês* que tem como sinônimo *papel washi*), mas também aos termos sintagmáticos que especificam o termo de entrada, ou seja, há remissão do termo mais amplo (hiperônimo) para os termos específicos (hipônimos). É o caso de *tinta* (hiperônimo), que remete para *tinta ácida* e *tinta ferrogálica* (seus hipônimos), como vemos na figura 5.6:

Figura 5.6 – Entrada de *tinta* e seus hipônimos – *Ver também*

TINTA
Língua: Português
Contexto: A tinta é um dos compostos mais importantes na documentação. Foi e é usada para escrever em papéis, pergaminhos e materiais similares, desde que o homem sentiu necessidade de registrar seu avanço técnico e cultural, e é ainda indispensável para a criação de registros e para atividades relacionadas aos interesses de vida diária. (ptPP130)
Ver também: tinta ácida tinta ferrogálica

Fonte: Base Papel, Termisul.

Por isso, na figura 5.7 da entrada de *tinta ferrogálica*, no campo *ver também* há remissão para *tinta* (seu hiperônimo) e para *tinta ácida* (sinônimo de *tinta ferrogálica*), ou seja, um caso de variação denominativa.

Figura 5.7 – Entrada de *tinta ferrogálica*

TINTA FERROGÁLICA
Língua: Português
Contexto: A composição da tinta ferrogálica é, basicamente, uma mistura de sulfato de ferro (II), um extrato rico em taninos de certas vesículas de origem vegetal e goma arábica em um líquido que pode ser água, vinho ou vinagre. A adição de corantes é também comum em algumas formulações. (ptPP004)
Ver também: tinta tinta ácida

Fonte: Base Papel, Termisul.

Os exemplos de fichas deste capítulo confirmam que a especificidade de cada projeto é que define os campos incluídos na ficha; por isso, a Terminografia não trabalha com um modelo fixo único de ficha. Como meio usado pelo terminógrafo para registrar as informações sobre termos e UFEs de modo facilmente recuperável e compreensível, a ficha é imprescindível. Na face visível ao consulente, indicam-se a fonte textual onde se observou o termo; os contextos de uso; as variantes denominativas etc. Na face visível, apenas aos integrantes do grupo de trabalho, podem ser inseridas informações de gerenciamento, como o nome de quem abriu a ficha, o registro de alterações nas fichas com indicação da data em que foram feitas, os comentários surgidos no processo de revisão da base ou dicionário etc.

A apresentação de exemplos serve ainda para indicar as possibilidades de mudança e inovação. Informações mais heterogêneas e menos comuns podem se tornar regra no futuro. Podemos prever, por exemplo, o uso cada vez mais frequente de ilustrações, áudios, *links* com fontes externas e imagens 3D, além de várias interfaces que permitam aos usuários selecionar os campos que lhe interessam, gerando resultados de buscas específicos às suas necessidades.

ATIVIDADES: Elaborando fichas terminográficas

1. Um vírus danificou a base do projeto Papel, e as informações dos campos das fichas se embaralharam. Corrija os erros da ficha abaixo, recolocando as informações nos campos corretos.
2. Retome os termos encontrados no *corpus* que você construiu na atividade 2 do capítulo 3. Crie uma ficha terminográfica para uso em um projeto de dicionário eletrônico desses termos destinados a tradutores.
3. Você é contratado por uma empresa brasileira do setor metalúrgico para solucionar problemas de comunicação em um projeto de transferência tecnológica para uma empresa estrangeira. Numa conversa com os técnicos

brasileiros e estrangeiros, você conclui que os problemas são gerados pela falta de padronização terminológica e então propõe a criação de uma base de dados bilíngue destinada a todos os funcionários envolvidos no projeto. Crie uma ficha terminográfica para essa base e justifique a inclusão de cada campo.

PARA SABER MAIS

COSTA, Maria Izabel Plath da Costa. *Terminologia jurídico-policia: proposta de elaboração de um glossário eletrônico*. Orientação: Cleci Regina Bevilacqua. Disponível em: <https://lume.ufrgs.br/handle/10183/102211>. Acesso em: 11 set. 2021.

BEVILACQUA, Cleci Regina; MACIEL, Anna Maria Becker; REUILLARD, Patrícia Chittoni Ramos. *Terminologia da Conservação dos bens móveis em papel: considerações iniciais*. 2018. Apresentação. ANPOLL. Disponível em: <https://www.ufrgs.br/termisul/files/file771959.pdf>. Acesso em: 11 set. 2021.

Capítulo 6 – Busca e identificação de equivalentes em línguas estrangeiras

Alexia Gonçalves Pokorski
Ana Letícia Prado de Campos
Iago Marques Barragan
Marina Canofre dos Santos
Tainara Cecília Silveira Balt

Neste capítulo, explicaremos o que entendemos por equivalência e como, a partir disso, buscamos os equivalentes para os termos em português nas diversas línguas estrangeiras que compõem a pesquisa do grupo Termisul. Os exemplos deste capítulo foram retirados das pesquisas relacionadas aos projetos *Conservação e Restauração de Bens Culturais Móveis em Papel (Projeto Papel)* e *Conservação dos bens culturais móveis em papel: Identificação e representação de unidades fraseológicas especializadas* (UFEs).

O que são equivalentes?

Em um trabalho terminológico bi ou multilíngue que tem o português como língua de partida, após a identificação e extração dos termos e UFEs em língua portuguesa, parte-se para a busca de seus equivalentes nas línguas estrangeiras. Em primeiro lugar, é importante retomar a definição de Termo Equivalente (TE) apresentada no capítulo 1. Os TEs são os correspondentes conceituais entre um termo ou uma UFE em língua portuguesa e um termo ou uma UFE em uma língua estrangeira que cumprem a mesma função nas duas línguas e culturas. Além disso, devem pertencer à mesma área, temática e campo semântico, ocorrer no mesmo gênero textual e ser empregados em situações comunicativas similares. Por exemplo, para o termo *papel* (português) foram identificados os seguintes equivalentes nos textos que compõem os *corpora* especializados em cada

língua: *paper* (inglês), *papel* (espanhol), *papier* (francês), *carta* (italiano) e *бумага* [*bumaga*] (russo).

Dessa forma, os equivalentes correspondem às denominações utilizadas em gêneros textuais semelhantes ao gênero textual no qual o termo em português foi encontrado e que possuem a mesma função comunicativa (Nord, 2016). A busca por equivalentes é um trabalho que exige do terminólogo uma pesquisa específica para cada campo conceitual ou recorte temático a fim de fornecer informações seguras e confiáveis sobre a linguagem e as terminologias utilizadas em cada área de especialidade (Bevilacqua; Finatto; Reuillard, 2009). Isso porque é preciso entender o conceito e o uso de determinado termo ou UFE na área especializada para buscar o equivalente em língua estrangeira, já que os conceitos podem variar entre as línguas de acordo com as áreas do conhecimento.

No projeto mencionado no início do capítulo, por exemplo, para buscar qual é o equivalente de *dobradeira* em outra língua, primeiro foi preciso entender seu conceito a partir dos contextos em que aparecia em textos especializados da área de conservação e restauração de papel em português. Segundo os contextos do *corpus* especializado dessa área, a *dobradeira* é um instrumento similar a uma régua, que pode ser feito de madeira ou de outros materiais, e que serve para dobrar o papel, como o próprio nome sugere. A partir disso, pode-se pensar que os equivalentes nas línguas estrangeiras podem derivar do verbo *dobrar*, como parece ser o caso do termo em português. Em espanhol, por exemplo, poder-se-ia pensar na possibilidade de existência do equivalente **dobladera*, que seria derivado do verbo *doblar*. No entanto, ao fazer a busca no *corpus* especializado do Termisul, percebeu-se que não havia nenhuma ocorrência de **dobladera*. Nesses casos, é preciso pensar em outro possível equivalente ou fazer buscas em outras obras de referência. Como um sinônimo de *doblar* é *plegar*, fizemos uma busca com a forma truncada *plega** para verificar se existia algum termo derivado do verbo *plegar*. A partir dessa forma, encontraram-se muitas ocorrências do termo *plegadera* no *corpus* em espanhol. O próximo passo foi a verificação de que realmente se tratava de um equivalente, nesse caso, se *dobradeira* e *plegadera* são o mesmo instrumento. Com

a ferramenta *Concordance* do programa *AntConc*, foi possível analisar os contextos em que esse termo aparecia, ou seja, seu uso em frases, que muitas vezes explicam os conceitos que os termos carregam. Com base nessas análises, foi possível verificar que *plegadera* é o equivalente em espanhol do termo *dobradeira* em português.

No entanto, às vezes, a busca de equivalentes pode ser mais complexa, pois os conceitos e suas denominações nem sempre estão alinhados da mesma forma em línguas e culturas distintas, pois sofrem a influência de diversos fatores. Assim, os conceitos de uma área de especialidade são organizados de acordo com a visão de mundo de cada comunidade científica ou conforme a área do conhecimento ou ciência em que se desenvolve a comunicação especializada, o que pode acarretar em graus variados de semelhança ou distanciamento em relação aos conceitos.

Por essas razões, muitas vezes, nos deparamos com algumas dificuldades durante a busca de equivalentes. No caso da língua russa, por exemplo, ao procurar pelos termos referentes aos agentes de deterioração do papel, mais precisamente os nomes de insetos, constatamos que, enquanto alguns deles possuíam equivalentes no *corpus* em russo, como *traça* [МОЛЬ (*mol*)] e *besouro* [жук (*juk*)], outros termos como *broca*, *barata*, *cupim* e *rato* deixavam lacunas, não sendo encontrados no referido *corpus*. Isso se deve a fatores como a temperatura e o clima de cada país, que interferem na fauna e na flora e, por consequência, nos termos relacionados a esses campos. Pelo mesmo motivo, observamos que insetos diferentes apareciam nos textos especializados em russo e em português, como пёстрый точильщик [*piostryi totchilshik*] e *Xestobium rufovillosum*, que constavam apenas no *corpus* em língua russa. Isso revela que os desafios da tradução especializada vão além da diversidade dos sistemas linguísticos: também as diferenças socioculturais, o desenvolvimento e a consolidação da área de especialidade em cada lugar são fatores responsáveis pela variação terminológica encontrada ao longo da busca dos equivalentes.

A variação terminológica, segundo Kilian (2007) e conforme visto no capítulo 1, se divide em dois tipos: denominativa e conceitual. A variação denominativa causa alterações na forma, podendo haver diferentes

nomes ou termos para nomear o mesmo conceito. Por exemplo, em português há duas maneiras de referir-se ao mesmo tipo de papel: *papel japonês* e *washi*. Por sua vez, uma das formas de variação conceitual se dá quando o mesmo termo expressa dois conceitos distintos, alterando seu significado ou algum traço semântico. Por exemplo, na base de dados terminológicos do grupo Termisul, é possível encontrar duas entradas para o termo água em português, já que uma delas se refere à água como agente de deterioração dos bens em papel, e outra, à água como um material utilizado nas técnicas de conservação e restauração.

Por conta desses fatores de variação explicitados acima, nem sempre o número de termos será exatamente o mesmo entre duas línguas, ou seja, pode haver mais ou menos equivalentes para cada termo. Por exemplo: o termo *cola* em português possui três equivalentes em espanhol (*adhesivo*, *cola* e *pegamento*); o termo *cartão alcalino* possui dois equivalentes em italiano (*carta barriera* e *carta con riserva alcalina*); a UFE *encadernação de livro* possui dois equivalentes em francês (*reliure d'un volume* e *reliure de livre*); e o termo *barbante* possui dois equivalentes em inglês (*string* e *thread*). Da mesma forma, o termo *termita*, em espanhol, é, ao mesmo tempo, o equivalente de *cupim* e *térmita* em português. Em português, os termos *velatura* e *laminação* se referem ao mesmo processo, portanto, há variação terminológica. Tal variação não é encontrada em inglês, espanhol e francês, que possuem somente um equivalente para ambos os termos (*lamination*, *laminación* e *lamination*, respectivamente).

Como identificar os equivalentes?

De maneira geral, a metodologia de busca de equivalentes adotada pelos membros do Termisul é a seguinte: após a identificação dos termos e das UFEs em português, é realizada a busca dos equivalentes desses termos e UFEs em cada língua estrangeira em um *corpus* textual especializado. Tais *corpora* são compilados em uma etapa anterior e estão compostos por textos de mesmo gênero (manuais, relatórios, teses, dissertações, anais de eventos, artigos científicos etc.) e temática que os textos do *corpus* em

português e coletados em *sites* confiáveis (universidades, centros de pesquisa, instituições públicas etc.).

Para o estabelecimento dos equivalentes, alguns critérios devem ser seguidos, entre eles: a) confiabilidade e ocorrência em textos especializados ou *corpora* especializados; b) frequência de uso da forma buscada; c) abrangência; e d) precisão conceitual (Bevilacqua; Finatto; Reuillard, 2009).

A **confiabilidade** e a **ocorrência em textos especializados** são necessárias para que sejam encontradas as terminologias inseridas na área do conhecimento em questão e não na linguagem geral. A **frequência de uso**, ou quantas vezes o termo ou a UFE aparece nos textos do *corpus* consultado, é importante para que não sejam identificadas como equivalentes formas pouco usuais ou até mesmo idiossincráticas.

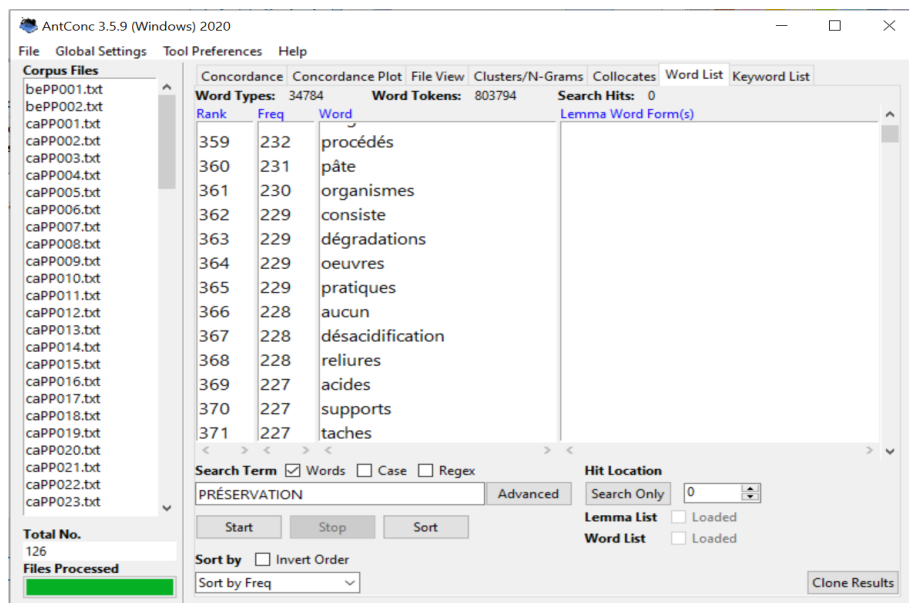
Segundo Bevilacqua, Finatto e Reuillard (2009, p. 67), a **abrangência** é “entendida como a possibilidade de maior ou menor aproximação conceitual entre os pares de línguas”, ou seja, a identificação de equivalência conceitual total, parcial ou ainda a sua inexistência. Ainda segundo as autoras, para esses casos podem ser elaboradas notas complementares para esclarecer o usuário da base de dados, de glossários etc. sobre as especificidades de cada termo ou UFE em cada língua. Em nossas buscas pelos equivalentes do projeto relacionado às UFEs, encontramos que, em português, o termo *deterioração* referia-se mais a processos físicos, e *degradação*, mais a processos químicos. No entanto, em espanhol, foi possível identificar variação denominativa e conceitual entre os termos *deterioro* e *degradación*, pois ambos são usados para referir-se tanto a processos químicos quanto físicos. Isso levou à existência de termos aparentemente contraditórios como *deterioro químico*, posto que em português, em princípio, a *deterioração* seria somente física. Para alertar o usuário da base de dados sobre essa variação no que se refere à **abrangência**, incluiu-se uma nota na ficha terminológica desse termo em espanhol.

A **precisão conceitual**, por sua vez, refere-se à análise de diferentes contextos de uso de um termo ou UFE nas línguas estrangeiras para identificar seus traços semânticos, compará-los com os dos termos e UFEs em

português e, assim, verificar sua correspondência conceitual, ou seja, se são de fato equivalentes. Com base nesses quatro parâmetros explicitados anteriormente, pode-se definir se um termo é o equivalente de outro em língua estrangeira.

Para a busca dos equivalentes, podem ser utilizados programas e *softwares* como o *AntConc* e o *Sketch Engine*, que compilam o *corpus* de cada língua e oferecem análises de frequência detalhadas, podendo também gerar listas de palavras que são candidatas a termo, conforme visto nos capítulos 3 e 4 e na figura 6.1 abaixo. O uso dos concordanciadores desses programas também auxilia muito a busca terminológica, pois permite que se visualizem os termos em seus contextos reais de uso, ou seja, em frases e parágrafos de textos escritos por especialistas da área.

Figura 6.1 – Lista de palavras do *corpus* em francês

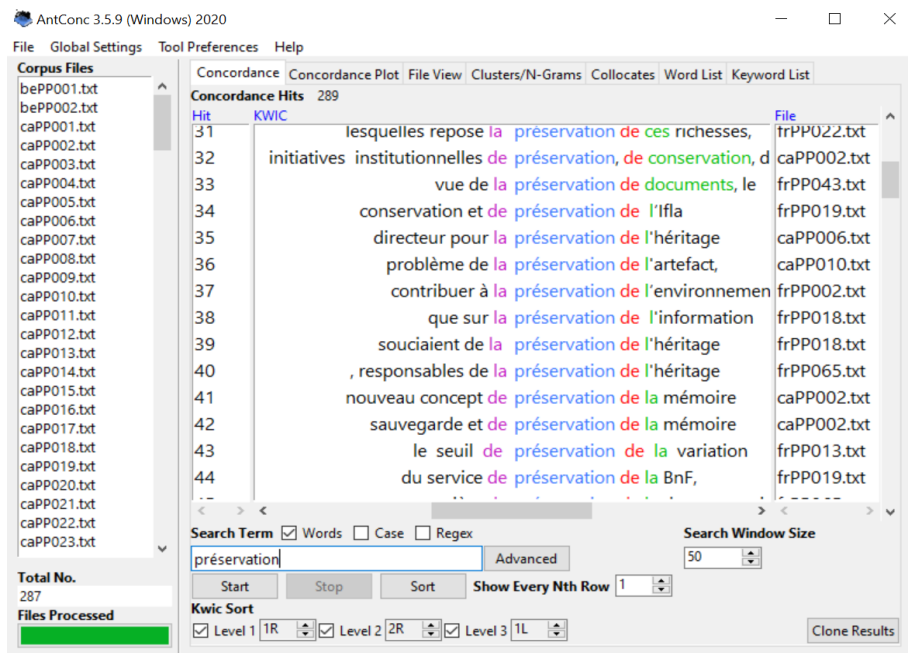


Fonte: AntConc (2020).

A seguir, ilustramos uma busca de termos através da ferramenta *Concordance*, do programa *AntConc*. A partir do termo *preservação*, em português, utilizou-se a ferramenta para buscar, no *corpus* especializado,

se seu equivalente em francês seria *préservation* ou outra forma. Após a análise dos resultados encontrados a partir dos parâmetros explicitados acima (confiabilidade, frequência, abrangência e precisão), identificou-se que *préservation* era realmente o equivalente correto para o termo.

Figura 6.2 – Concordâncias do termo *préservation* em francês



Fonte: AntConc (2020).

Também é possível realizar essas buscas por meio do *Sketch Engine*. Na figura 6.3, vemos a lista de palavras-chave do *corpus* do projeto sobre conservação e restauração de bens em papel em espanhol, que resultou da comparação com as palavras do *corpus* de referência em espanhol disponível no programa, conforme explicado no capítulo 4.

Figura 6.3 – Lista de palavras-chave do corpus em espanhol – *Corpus Papel*

SINGLE-WORDS ✓			
reference corpus: Spanish Web 2018 (esTenTen18)			
Word	Frequency	Score [?]	
1 archivístico	725	434.2	...
2 celulosa	915	249.9	...
3 encuadernación	572	226.1	...
4 lignina	190	143.6	...
5 archivos	421	102.4	...
6 conservación	3,460	91.5	...
7 adhesivo	571	85.8	...
8 tinta	848	75.6	...
9 cuadernillo	154	74.5	...
10 laminación	114	73.4	...
11 preservación	610	72.9	...
12 archivística	89	72.2	...
13 ciclododecano	70	68.8	...
14 celulósico	83	68.6	...

Fonte: Sketch Engine (2022).

Após a identificação dos possíveis equivalentes, a equipe de trabalho responsável por cada língua analisa e revisa os termos encontrados. Se for confirmada a equivalência, os termos são inseridos na base de dados terminológicos do Termisul.

Caso o tradutor não tenha acesso a um *corpus* especializado, é possível buscar equivalentes em dicionários ou glossários especializados ou em textos escritos por especialistas da área de conhecimento em questão. Para tal fim, também é preciso estabelecer e seguir critérios como confiabilidade da fonte (*sites* institucionais e de pesquisa etc.), nível de especialização do autor, temática e gêneros textuais semelhantes na língua fonte e na língua alvo, conforme referimos anteriormente. Por exemplo, durante as

buscas do projeto sobre papel, não foi possível encontrar no *corpus* em inglês o equivalente para o termo *documento digital*. Ao buscar, fora do *corpus*, em uma fonte confiável na internet (grupo ligado à National Digital Library Federation), foi encontrado o equivalente *digital document* e, portanto, ele foi inserido na base de dados como equivalente para o termo em português.

Variação terminológica

Como supracitado e visto no capítulo 1, durante o processo de busca de equivalentes, é constante a necessidade de resolver casos de variação terminológica, como, por exemplo, a utilização de formas lexicais distintas para referir-se ao mesmo conceito ou expressar a mesma função no contexto discursivo especializado dos termos e UFEs. Temos como exemplos desse tipo de variação retirados da base de dados do Termisul as seguintes UFEs:

<i>Causar mancha</i>	<i>Provocar mancha</i>
As sujidades que se acumulam no acervo podem vir do ar ou da falta de cuidado dos usuários, que podem manipular alimentos perto dos documentos. Essas sujidades podem favorecer a propagação de insetos, roedores e microrganismos, além de causar manchas e oxidação no documento. (ptPP124)	Pragas: Insetos (cupins, brocas, baratas e traças), roedores e pássaros. Perfuram o papel causando perda de informação, defecam nos depósitos, provocam manchas e enfraquecimento do suporte. (ptPP156)

Nesses casos, atentamo-nos ao tipo de variação, que pode ser lexical, no caso de unidades sinonímicas (*proteção/segurança do acervo*), ou morfológica, no caso de alternância da forma verbal e deverbal (*preservar/preservação de acervo*); alternância do modo verbal (*abriguem/abrigam o acervo*); alternância da forma gramatical (singular/plural, masculino/feminino, voz ativa/passiva, tempo verbal); inserção de um elemento léxico na combinatória (*análise detalhada da obra*) ou siglas (*máquina obturadora de papel/*

MOP) (Maciel e Reuillard, 2015; Kilian, 2007). No caso das unidades fraseológicas, a variação pode ocorrer nos verbos, como no exemplo acima, *causar/provocar mancha*, ou apresentar a mesma forma verbal ocorrendo com termos sinonímicos, como *remoción de adhesivo* e *remoción de celo*, em espanhol.

Conforme visto no capítulo 5, para a inserção dessas UFEs, dispomos de dois campos na base de dados: *Outras formas*, para as UFEs que apresentam alguma variação morfológica em relação à UFE original, e *Ver também*, para as UFEs que apresentam sinonímia. Esse trabalho é realizado no intuito de que o nosso usuário conheça não somente a forma mais utilizada da UFEs, mas também as suas variações. Assim, podemos oferecer sugestões de equivalentes variantes adequados para determinado contexto. Por exemplo, são apresentados dois equivalentes em italiano para o termo *cartão alcalino* em português: *carta barriera* e *carta con riserva alcalina*, que, por sua vez, estão conectados pelo campo *Ver também*.

Dessa forma, a atividade de busca de equivalentes para a criação de uma base de dados terminológicos estabelece equivalentes que possuem sinônimos e variações morfológicas nas línguas que fazem parte do projeto. Para isso, se verifica a existência desses fenômenos em cada língua de trabalho e se analisam os diferentes casos encontrados. Na língua russa, por exemplo, para a UFE *restauração de documento* foram encontrados dois equivalentes: Восстановление документа [*vosstanovlenie dokumienta*] e Реставрация документа [*riestavratsia dokumienta*]. Nesse caso, a palavra *реставрация* [*riestavratsia*] é um estrangeirismo, um empréstimo do latim, *restaurāre*, enquanto *восстановление* [*vosstanovlenie*] é uma palavra de origem russa que possui o mesmo significado.

No russo, além de sinônimos, também foram encontradas variações morfológicas, assim como no português. A UFE *fabricação de papel*, por exemplo, possui *производство бумаги* [*proizvodstvo bumagui*] como equivalente, porém, há ainda a variação morfológica *бумажное производство* [*bumajnoe proizvodstvo*]. A variação morfológica ocorre no termo *бумага* [*bumaga*] (*papel*, em português). No primeiro caso, esse termo aparece declinado no caso genitivo singular, que é um caso que indica uma relação de

posse entre a palavra no genitivo e uma outra palavra, neste caso, “de/do papel”. Já no segundo caso, *бумажное* [*bumajnoe*] está cumprindo a função de adjetivo. Não temos uma forma verbal nesse caso, mas a palavra “papel” em russo está cumprindo a mesma função de maneiras diferentes: em um caso como adjetivo e, no outro, como um substantivo no genitivo. Como ambas as UFEs constavam no *corpus* em russo e são utilizadas na linguagem da área, as duas foram acrescentadas na ficha terminológica da base de dados como equivalentes de *fabricação de papel*.

Além disso, podemos encontrar casos em que há a ocorrência de sinonímia em uma língua e não na outra entre termos ou UFEs que apresentam uma relação de equivalência. É o caso de *remoção de adesivo*, que possui dois equivalentes em espanhol: *remoción de adhesivo* e *remoción de celo*. Ou seja, enquanto nosso *corpus* em língua portuguesa possui apenas uma UFE para esse conceito, no *corpus* em língua espanhola é possível identificar duas formas para representá-lo.

ATIVIDADES: identificação de equivalentes

1. Nos contextos apresentados abaixo, identifique os equivalentes das línguas que conhece ou com as quais trabalha. Considere os princípios citados anteriormente (entendimento do conceito do termo, análise dos contextos, adequação etc.) e, se for preciso, consulte dicionários ou glossários especializados.

Espanhol

Contexto em português:	Contexto em espanhol:
Após planificação, fez-se o acondicionamento individualizado dos mapas em pastas de cartão alcalino , especialmente confeccionadas nas suas dimensões. (ptPP075)	Siempre que existan en el mercado cajas, carpetas y otros elementos de guarda, de cartón libre de ácido, se deberá priorizar su uso sobre los de cartón normal. (arPP027)
Termo: cartão alcalino	Equivalente:

Francês

Contexto em português:	Contexto em francês:
A qualidade da fibra, em particular a quantidade de celulose e de hemi-celulose, e a presença ou ausência de lignina são fundamentais no envelhecimento do papel . (ptPP151)	On utilisait de l'eau désionisée pour le lavage, car on pensait que les ions métalliques dissous, comme le fer, le cuivre et le manganèse, qui sont parfois présents dans l'eau du robinet, pouvaient accélérer le vieillissement du papier et des textiles cellulosiques. (caPP032.txt)
Termo: envelhecimento do papel	Equivalente:

Inglês

Contexto em português:	Contexto em inglês:
A atmosfera anóxia é obtida pela conjugação de diferentes recursos: bolsas de plástico de alta barreira; sachês absorvedores de oxigênio; indicadores de oxigênio residual; sachês de sílica gel [...]. (ptPP081)	Inert Cases in the Control of Museum Insect Pests is a compendium of information on the biological mechanisms by which nontoxic gases kill insects; the methods and materials needed to create and maintain an anoxic atmosphere; treatments; the construction and use of chambers and bubbles; and the procedures for treating objects. (Fonte – 23/04/2019; us)
Termo: atmosfera anóxia	Equivalente:

Italiano

Contexto em português:	Contexto em italiano:
Após a remoção do excesso de álcali, o tratamento é finalizado com um banho aquoso em solução de Hidróxido de Cálcio, $\text{Ca}(\text{OH})_2$, a 0,02% por 20 minutos. (ptPP085)	Queste esperienze indicarono ai restauratori la validità del lavaggio acquoso come metodo, utile e non drastico, di pulitura dei supporti cartacei spesso sostituito dagli sbiancamenti. (itPP019)
Termo: banho aquoso	Equivalente:

Russo

Contexto em português:	Contexto em russo:
[Agentes] Ambientais – a falta de controle de temperatura e umidade relativa e o contato com poluentes podem provocar danos irreversíveis como manchas, menor resistência do papel, solubilidade de tintas e carimbos, atrair microrganismos como fungos e bactérias. Têm forte influência na degradação química também. (ptPP120)	Определение водостойкости чернил, карандаша, методом копирования выявило устойчивость к водному растворению чернил, графитного карандаша. (ruPP043).
Termo: solubilidade de tintas	Equivalente:

2. Identifique os equivalentes em sua(s) língua(s) de trabalho do termo e das UFEs encontradas nas atividades 2 e 3 do capítulo 4.

3. A partir da criação de fichas terminológicas (atividade 2 do capítulo 5) para os termos identificados na atividade 5 do capítulo 3, identifique seus equivalentes em fontes confiáveis em sua(s) língua(s) de trabalho. Você pode incluir também as UFEs identificadas.

PARA SABER MAIS

BEVILACQUA, Cleci R.; CERESER, M. T. I.; LOGUERCIO, Sandra D. Uma proposta de objeto de aprendizagem para futuros tradutores: a modalização em resumos científicos em português, espanhol e francês. *Texto Livre: Linguagem e Tecnologia*, Belo Horizonte, v. 11, n. 1, p. 43-59, jan.-abr/2018.

BEVILACQUA, Cleci R.; COIMBRA, S. A. C. Equivalência em língua espanhola para termos de gestão ambiental em língua portuguesa: questões de sinonímia e variação. *ABECAN*, 2005.

BEVILACQUA, Cleci R.; MACIEL, A. M. B. A fraseologia da legislação do direito ambiental em línguas e sistemas jurídicos distintos. In: ZAVAGLIA, Claudia; SIMÃO, A. K. G. *Reflexões, tendências e novos rumos dos Estudos Fraseoparemiológicos*. São José do Rio Preto: Unesp/Ibilce, 2017. p. 45-56.

BEVILACQUA, Cleci R.; REUILLARD, P. C. R. *Glossário de Gestão Ambiental: Questões de Equivalência*. In: CONGRESSO INTERNACIONAL DA ABRALIN, 5, 2007. Belo Horizonte: UFMG, 2007.

SALES, Denise Regina de; CANOFRE, Marina. Conservação dos bens culturais móveis em papel: termos-chave em português e seus equivalentes em russo. *TradTerm*, São Paulo, v. 36, dez. 2020, p. 102-127.

Posfácio

Por uma Terminografia que faça a diferença

Maria José Bocorny Finatto

Este livro pretende introduzir o leitor à área da Terminografia, um território de práticas e de estudos. Um espaço de conhecimento que tem a produção de dicionários e obras afins, tais como glossários, bancos e bases de dados, como um ponto focal. Este manual tem, antes de tudo, o objetivo maior de dar a conhecer os pontos básicos do trabalho com textos e discursos especializados, de temática técnico-científica. Mostra-se, enfim, como é feito todo o trabalho que dá origem a essas obras e repertórios dicionarísticos. Quem vê, busca ou usa um produto desse tipo, muitas vezes, não tem conhecimento do que ele demandou para, concretamente, existir.

Além de indicar, didaticamente, um passo a passo, indispensável para a formação de novos terminógrafos e terminólogos, este é um manual, um guia, uma fonte que visa a motivar seus leitores à exploração dessa área prática e dos estudos e pesquisas a ela conectados. Assim, cada capítulo sintetiza pontos importantes, mas também estimula o aprofundamento dos assuntos, dos conhecimentos sobre os instrumentos. São orientações para quem inicia sua caminhada. Em meio à história do nosso grupo de pesquisas, o grupo Termisul, chama-se a atenção para o *modus operandi* de projetos de produção dicionarística. Apresentam-se seus diferentes formatos e etapas, o que se precisa fazer até a organização de bases de dados e de arquivos terminológicos, com especial destaque para os materiais que servem de apoio e fonte para as tarefas de tradução especializada.

Ao receber o convite para este posfácio, foi inevitável não recordar minha participação na construção de uma base de dados multilíngue do Termisul, a que deu origem ao Glossário de Gestão Ambiental (Krieger *et al.*, 2006). Foi a base denominada Gestamb, na qual reunimos, junto com toda uma equipe, as terminologias das normas técnicas ambientais,

dos manuais técnicos e das leis. Além do português, toda a terminologia foi apresentada em quatro línguas estrangeiras: alemão, espanhol, francês e inglês, e cada língua demandou a reunião de conjuntos de fontes documentais diferentes. Muitas encontramos disponíveis apenas em formato impresso, as quais digitalizamos e processamos como *corpora*.

Naquela ocasião, já empregávamos a maioria dos procedimentos destacados nesta obra, que hoje se concretiza em um manual que é também um testemunho de experiências. Experiências de várias pessoas e espelhamento das nossas diferentes perspectivas sobre o que o/um trabalho terminológico representa. Afinal, sempre tentamos equacionar – e não necessariamente resolver – questões diversas, as quais permeiam os títulos de alguns capítulos deste tão necessário Manual de Terminografia: teorias e práticas; tomada de decisões de projetos com temas mais ou menos específicos; identificação do destinatário do trabalho e de suas necessidades; escolha de métodos, equipe de trabalho e de instrumentos; desenho do repertório das fontes documentais de estudo; seleção das fontes de apoio; desenho de sistemas de conceitos e suas conexões; identificação de padrões da linguagem especializada em foco; seleção das terminologias e de respectivas informações conceituais e linguísticas, definição do dossiê dos termos – estabelecimento da ficha com os dados de cada termo/expressão em foco; o estudo contrastivo em diferentes línguas e identificação de equivalências ou mesmo de “vazios” entre termos em línguas diferentes; síntese de seus usos e valores culturais; e, finalmente, o desenho e a concretização do material de saída para consulta do destinatário.

Todas essas etapas de trabalho são, hoje, em grande parte, mediadas e estabelecidas por recursos e técnicas computacionais, advindos da Linguística de *Corpus* e do Processamento de Linguagem Natural. Essas duas áreas redesenharam os processos de Recuperação da Informação e os modos de representação e de conexão de conteúdos dispersos em fontes textuais, mesmo as que estejam em diferentes idiomas. Mas, a despeito desse progresso tecnológico ímpar para a Terminografia e a Terminologia, os elementos-chave que dão mais valor ao percurso e ao produto do trabalho

serão sempre o olhar e o discernimento subjetivos, as atenções das pessoas nele envolvidas.

Nesse caminho, ajudamos a estabelecer e a difundir a ideia de que é importante favorecer a acessibilidade textual e terminológica (ATT). Para quem busca democratizar o acesso à informação e favorecer os bons diálogos entre especialistas e leigos, a Terminografia didático-pedagógica mostra-se essencial. Com ela, destacaram-se as perspectivas teóricas e metodológicas das dicionarizações em geral mais afeitas aos diálogos com um cenário de saberes e conhecimentos em movimento: pessoas que sabem e pessoas que querem saber. A *acessibilidade* envolvida, sempre frisamos, não diz respeito apenas à compreensão das terminologias, “dos termos técnicos”, de vocabulários “peculiares” e respectivos conceitos, estejam eles em isolado, em “redes” ou em nuvens de palavras que, hoje, servem para resumir documentos. Lidamos, sim, com textos e terminologias em situações comunicativas diferenciadas e, sobretudo, com pessoas e conhecimentos em diálogo. Estão em jogo elementos comunicativos, cognitivos, discursivos, sócio-históricos e linguístico-enunciativos, o que demanda uma perspectiva que já referimos como *textualista* (Finatto, 2004, p. 349). Mas, como vejo hoje, é algo que vai bem mais além.

Finalizando, ao saudar a existência desta obra e cumprimentar as pessoas que a concretizaram e reconhecer seus esforços, vale retomar o título deste posfácio: a Terminografia que faz a diferença. Essa, para mim, é aquela que é capaz de conectar as pessoas em torno de um manancial – ou mesmo de uma lacuna – de conhecimentos e de saberes que, para existirem, precisam de movimento e de transformação contínuos. Fazer diferença, via Terminografia, é instrumentalizar os diálogos, sistematizar e mostrar os valores e as culturas que permeiam as práticas de comunicação técnico-científica – em seus mais diferentes cenários e circunstâncias. Assim, a partir dos trabalhos realizados e mesmo já ao longo deles, a Terminografia que faz a diferença chama os envolvidos para a reflexão e, sobretudo, para as ações praticadas em meio aos domínios de especialidade, que são reconhecidos e repertoriados em diferentes projetos: da Culinária à Física Quântica. É um trabalho que, ao nos mostrar um

panorama de conhecimentos expressos e organizados via usos de linguagem, também aponta para aquelas ações que ainda precisam ser construídas. Por isso, desejo que a leitura deste manual nos seja inspiradora!

Porto Alegre – RS, Inverno de 2022.

Referências

KRIEGER, Maria da G.; MACIEL, Anna Maria B.; BEVILACQUA, Cleci R.; FINATTO, Maria José B.; REUILLARD, Patrícia C. R. *Glossário de Gestão Ambiental*. São Paulo: DISAL, 2006.

FINATTO, Maria José B. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, Aparecida N.; KRIEGER, Maria da G. (Orgs.) *As Ciências do Léxico*. Volume II. Campo Grande – MS: Editora UFMS, 2004.

Referências

ALUÍSIO, Sandra M.; ALMEIDA, Glades M. B. O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. *Calidoscópio*, v. 4, n. 3, p. 156-178, 2006.

ALVES, Ieda M. A unidade lexical neológica: do histórico-social ao morfológico. In: ISQUERDO, Aparecida N.; KRIEGER, Maria da Graça. *As Ciências do Léxico*. Lexicologia, Lexicografia, Terminologia. Campo Grande: Editora UFMS, 2004. p. 77-100. v. II.

ARCOS, Manuela; BEVILACQUA, Cleci R. UFE eventivas na área da conservação e restauração de bens culturais móveis em suporte papel: identificação e análise. *Debate Terminológico*, [Porto Alegre], n. 18, p. 4-18, 2020. Disponível em: <https://seer.ufrgs.br/riterm/article/view/98700>. Acesso em: 17 set. 2021.

ARCOS, Manuela; BEVILACQUA, Cleci R. Metodologias para a extração e identificação de unidades fraseológicas especializadas eventivas em *corpora* textuais. *Guavira Letras*, v. 27, p. 75-95, 2018. Disponível em: <http://websensors.net.br/seer/index.php/guavira/article/view/714> . Acesso em: 17 set. 2021.

BAKER, Mona. Corpora in Translation Studies. An overview and suggestions for future research. *Target*, v. 7, n. 2, p. 223-243, 1995.

BAKER, Mona. Corpus Linguistics and Translation Studies: implications and applications. In: BAKER, Mona; FRANCIS, Gill; TOGNINI-BONELLI, Elena (org.). *Text and Technology: in honour of John Sinclair*. Amsterdam: John Benjamins, 1993. p. 233-250.

BERBER SARDINHA, Tony. Histórico e problemática. *D.E.L.T.A.*, v. 16, n. 2, p. 323-367, 2000.

BERBER SARDINHA, Tony. *Linguística de Corpus*. Barueri, SP: Manole, 2004.

BEVILACQUA, Cleci R. Fraseologia Especializada: panorama das pesquisas realizadas no Brasil. In: SILVA, Suzete (org.). *Fraseologia & Cia: entabulando diálogos reflexivos*. Campinas, SP: Pontes, 2020. p. 41-66. v. 2.

BEVILACQUA, C. R. Investigación Sistemática en Terminología. In: ÁLVAREZ CATALÁ, Sara; BARITÉ, Mario (org.). *Teoría y praxis en Terminología*. Montevideo:

Ediciones Universitarias, Unidad de Comunicación de la Universidad de la República, 2017. p. 69-90. v. 1.

BEVILACQUA, Cleci R. *Unidades Fraseológicas Especializadas Eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Tese (Doutorado). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 2004.

BEVILACQUA, Cleci R. Terminologia mono/bi/multilíngue: algumas propostas e reflexões referentes às unidades fraseológicas especializadas. *TradTerm*, n. 8, p. 135-147, 2002.

BEVILACQUA, Cleci R. *A fraseología jurídico-ambiental*. 1996. Dissertação (Mestrado em Estudos da Linguagem). Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Letras, Porto Alegre, 1996.

BEVILACQUA, Cleci R.; FINATTO, Maria José B.; REUILLARD, Patrícia C. R. Glossário de gestão ambiental: estabelecimento de equivalentes em alemão, espanhol e francês. *Tradução & comunicação: Revista Brasileira de Tradutores*, São Paulo, n. 19, p. 61-72, 2009.

BEVILACQUA, Cleci R.; MACIEL, Anna Maria B. A variação terminológica em uma base de dados de combinatórias léxicas especializadas: descrição e tratamento. In: ISQUERDO, Aparecida N.; DAL CORNO, Giselle O. M. (org.). *As Ciências do Léxico*, Campo Grande: Ed. UFSM, 2018. p. 273-290. v. VIII.

BEVILACQUA, Cleci R. *et al.* Combinatórias léxicas especializadas da linguagem legislativa: uma abordagem orientada pelo *corpus*. In: MURAKAWA, Clotilde; NADIN, Odair Luiz (ed.). *Terminologia: uma ciência interdisciplinar*. São Paulo: Cultura Acadêmica, 2013. p. 227-243.

BEVILACQUA, Cleci R. *et al.* Acervo Termisul: implantação das bases textuais. In: CONGRESSO INTERNACIONAL DA ASSOCIAÇÃO BRASILEIRA DE LINGÜÍSTICA (ABRALIN), 7, 2009, João Pessoa. *Anais...* João Pessoa: Ideia, 2009. v. 1. p. 815-824. 2009.

BIBER, Douglas. Representativeness in corpus design. *Literary and Linguistic Computing*, v. 5, n. 4, p. 243-257, 1993.

BOJANOSKI, Silvana F. *Terminologia em Conservação de bens culturais em papel: produção de um glossário para profissionais em formação*. Tese (Doutorado) –Programa de Pós-Graduação em Memória Social e Patrimônio Cultural, UFPEL, 2018.

BOJANOSKI, Silvana F.; MICHELON, Francisca; BEVILACQUA, Cleci Regina. Criação do *corpus* para um estudo terminológico da área da conservação e restauração de bens culturais. *Debate Terminológico*, n. 17, p. 33-45, 2017.

BOURIGAULT, Didier; SLODZIAN, Monique. Pour une terminologie textuelle. *Terminologies Nouvelles*, n. 19, déc. 1998-juin. 1999.

CABRÉ, María Teresa. *Terminología: representación y comunicación*. Una teoría de base comunicativa y otros artículos. Barcelona: Universitat Pompeu Fabra; Instituto Universitario de Lingüística Aplicada, 1999.

CABRÉ, María Teresa. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Editorial Empúries, 1993.

CHICHORRO, Caroline L. C. M. *Terminologia do Licenciamento Ambiental em português e inglês*. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras, Porto Alegre, 2016.

CHURCH, Kenneth W.; HANKS, Patrick. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, n. 16, p. 22-29, 1990.

DIKI-KIDIRI, Marcel. Eléments de terminologie culturelle. *Cahiers du Rifal*, v. 26, 2007.

FABER, Pamela; MÁRQUEZ, Carlos; VEGA, Miguel. Framing Terminology: A Process-Oriented Approach. *Meta: journal des traducteurs / Meta: Translators' Journal*, v. 50, n. 4, 2005. Disponível em: <https://www.erudit.org/en/journals/meta/2005-v50-n4-meta1024/019916ar.pdf>. Acesso em: 15 mar. 2022.

FAULSTICH, E. A socioterminologia na comunicação científica e técnica. *Ciência e Cultura*, v. 58, n. 2, 2006. Disponível em: <http://cienciaecultura.bvs.br/pdf/cic/v58n2/a12v58n2>. Acesso em: 8 abr. 2022.

FINATTO, Maria José B. A definição de termos técnico-científicos no âmbito dos estudos de terminologia. *Rev. Est. Ling.*, Belo Horizonte, v. 11, n. 1, p. 197-222, jan./jun. 2003. Disponível em: <http://periodicos.letras.ufmg.br/index.php/relin/article/view/2351> Acesso em: 18 abr. 2022.

FINATTO, Maria José B. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, Aparecida

N.; KRIEGER, Maria da G. (org.). *As Ciências do Léxico*. Campo Grande, MS: Editora UFMS, 2004. v. II.

FINATTO, Maria José. B. *Definição terminológica: fundamentos teórico-metodológicos para sua descrição e explicação*. 2001. Tese (Doutorado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Letras, Porto Alegre, 2001.

FISH, Stanley E. *Is There a Text in This Class?: The Authority of Interpretive Communities*. Cambridge, Mass: Harvard University Press, 1980.

FROMM, Guilherme *et al.* Wordsmith Tools e Sketch Engine: um estudo analítico-comparativo para pesquisas científicas com uso de corpora. *Revista de Estudos Linguísticos*, Belo Horizonte, v. 28, n. 3, p. 1.101-1.248, 2020.

GAUDIN, François. *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen, 1993. Disponível em: <http://websensors.net.br/seer/index.php/guavira/issue/view/34/showToc>. Acesso em: 7 jun. 2022.

KILIAN, Cristiane K. *A retomada de unidades de significação especializada em textos em língua alemã e portuguesa sobre gestão de resíduos: uma contribuição para a tradução técnico-científica*. Tese (Doutorado) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Instituto de Letras, Porto Alegre, 2007.

KILIAN, Cristiane K.; LOGUERCIO, Sandra D. Fraseologias de gênero em resumos científicos de Linguística, Engenharia de Materiais e Ciências Econômicas. *Tradterm*, n. 26, p. 241-267, 2015.

KRIEGER, Maria da Graça. Terminografia: entre teoria e aplicações. In: ISQUERDO, Aparecida N.; DAL CORNO, Giselle O. M. *As Ciências do Léxico*. Lexicologia, Lexicografia, Terminologia, Campo Grande, MS: Editora UFMS, 2018. p. 329-346. v. VIII.

KRIEGER, Maria da Graça. Porque Lexicografia e Terminologia: relações textuais. In: ENCONTRO DO CELSUL (Círculo de Estudos Linguísticos do Sul), 8., 2008. *Anais...* Pelotas: Educat, 2008.

KRIEGER, Maria da Graça. *Terminologias em construção: procedimentos metodológicos*. In: CONGRESSO INTERNACIONAL DA ABECAN (Associação Brasileira

de Estudos Canadenses), 8., 2005. *Anais...* Gramado, 2005. Disponível em: <http://www.ufrgs.br/termisul/files/file112160.pdf>. Acesso em: 10 nov. 2021.

KRIEGER, Maria da Graça. Do reconhecimento de terminologias: entre o linguístico e o textual. In: ISQUERDO, Aparecida N.; KRIEGER, Maria da Graça. *As Ciências do Léxico*. Lexicologia, Lexicografia, Terminologia, Campo Grande, MS: Editora UFMS, 2004, p. 327-339. v. II.

KRIEGER, Maria da Graça. Sobre Terminologia e seus objetos. In: LIMA, Marília; RAMOS, Patrícia C. (org.). *Terminologia e ensino de segunda língua: Canadá e Brasil*. Porto, Alegre: NEC, IL, UFRGS/Abecan, 2001. p. 45-53.

KRIEGER, Maria da Graça. Terminologia revisitada. *DELTA*, v. 16, n. 2, p. 209-228, 2000.

KRIEGER, Maria da Graça. Terminologia em contextos integradores: funcionalidade e fundamentos. *Organon*, v. 12, n. 26, p. 19-31, 1998.

KRIEGER, Maria da Graça; FINATTO, Maria José B. *Introdução à Terminologia: Teoria & Prática*. São Paulo: Contexto, 2004.

KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker; FINATTO, Maria José Bocorny. Terminografia das leis do meio ambiente: princípios teórico-metodológicos. In: KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker (org.). *Temas de terminologia*. Porto Alegre/São Paulo: Ed. Universidade/UFRGS/Humanitas/USP, 2001. p. 317-335.

KRIEGER, Maria da Graça et al. *Dicionário de Direito Ambiental: terminologia das leis do meio ambiente*. 2. ed. rev. e atualizada. Rio de Janeiro: Lexicon, 2008.

KRIEGER, Maria da Graça et al. *Glossário de gestão ambiental*. Barueri, SP: Disal, 2006.

KRIEGER, Maria da Graça et al. *Glossário Multilíngue de Direito Ambiental Internacional*. Rio de Janeiro: Ed. Forense, 2004.

KRIEGER, Maria da Graça et al. *Dicionário de direito ambiental: terminologia das leis do meio ambiente*. Porto Alegre: Editora da Universidade/UFRGS, 1998.

LAZZARIN, Renan. *AGROTÓXICO E PFLANZENSCHUTZMITTEL: estudo exploratório da variação terminológica e proposição de equivalentes tradutórios no par de línguas português-alemão*. Trabalho de conclusão de curso (Bacharel em Letras – Tradutor Português e Alemão) – Universidade Federal do Rio Grande do Sul.

Instituto de Letras, Porto Alegre, 2007. Disponível em: <https://lume.ufrgs.br/handle/10183/178858>. Acesso em: 22 ago. 2023.

LEECH, Geoffrey Corpora. In: MALMKJAER, Kirsten (ed.). *The Linguistics Encyclopedia*. London: Routledge, 1991. p. 73-80.

LOGUERCIO, Sandra D. A linguagem comum do artigo científico em português brasileiro: um estudo baseado em corpus. *ANTARES*, v. 12, n. 25, p. 140-164, jan./abr. 2020.

LOGUERCIO, Sandra D. Entre buscar contribuir e la contribution: a modalização em resumos científicos em português/francês. *Linguagem & Ensino*, v. 22, n. 3, p. 881-995, jul./set. 2019.

LOGUERCIO, Sandra D.; KILIAN, Cristiane K. Fraseologias de gênero de resumos de artigos científicos (português, alemão e francês). In: Claudia Zavaglia; Angélica Karim Garcia Simão. (Org.). *Reflexões, tendências e novos rumos dos estudos fraseoparemiológicos*. 1ed. São José do Rio Preto (SP): UNESP/IBILCE, 2017, v., p. 94-108.

MACIEL, Anna Maria B. Reflexão sobre a pesquisa terminológica em corpus. In: ENCONTRO NACIONAL DA ANPOLL, 21, São Paulo. *Domínios do Saber: História, Instituições, Práticas*, 2006. Disponível em: https://silo.tips/queue/xxi-encontro-nacional-da-anpoll-associao-nacional-de-pos-graduacao-e-pesquisa-em?&queue_id=-1&v=1654593928&u=MmEwMTo0YjAwOjg0NGQ6YWlWMDo5YzM3OmVlZjplNzMxOmE3ZmM=. Acesso em: 7 jun. 2022.

MACIEL, Anna Maria B. *Para o reconhecimento da especificidade do termo jurídico*. Tese (Doutorado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Letras, Porto Alegre, 2001.

MACIEL, Anna M.; BEVILACQUA, Cleci R. A fraseologia da legislação do Direito Ambiental em línguas e sistemas jurídicos distintos. In: ZAVAGLIA, Claudia; SIMÃO, Angélica (org.). *Reflexões, tendências e novos rumos dos Estudos Fraseoparemiológicos*. São José do Rio Preto: Unesp, 2017. p. 46-56.

MACIEL, Anna Maria B.; REUILLARD, Patrícia C. R. Abordagem da variação terminológica em uma base de dados de combinatórias léxicas. *Tradterm*, São Paulo, v. 26, p. 223-240, 2015. Disponível em: <https://doi.org/10.11606/issn.2317-9511.v26i0p223-240>. Acesso em: 12 out. 2021.

MARCUSCHI, Luiz Antônio. *Produção textual, análise de gêneros e compreensão*. São Paulo: Parábola Editorial, 2008.

MARCUSCHI, Luiz Antônio. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, Ângela P.; MACHADO, Anna Raquel; BEZERRA, Maria Auxiliadora (org.). *Gêneros textuais e ensino*. 4. ed. Rio de Janeiro: Lucerna, 2005.

MCENERY, Tony; HARDIE, Andrew. *Corpus Linguistics: method, theory and practice*. Edinburgh: Cambridge University Press, 2012.

NORD, Christiane. *Traducir, una actividad con propósito*. Introducción a los enfoques funcionalistas. Berlim: Frank & Timme GmbH, 2018.

NORD, Christiane. Lealdade em vez de fidelidade: proposta de uma tipologia funcional da tradução. *Cadernos de Tradução*, Porto Alegre, Número Especial, p. 9-24, 2016.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). *Convenção das Nações Unidas sobre Direito do Mar*. 1990. Disponível em: http://www.unbciencia.unb.br/images/Noticias/2019/12-Dez/Convencao_das_Nacoes_Unidas_sobre_Direito_do_Mar_Montego_Bay.pdf. Acesso em: 8 jul. 2022.

REUILLARD, Patrícia C. R. Neologismos lacanianos e equivalência tradutória. Tese (Doutorado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Letras, Porto Alegre, 2007. Disponível em: <https://lume.ufrgs.br/handle/10183/12506>. Acesso em: 23 ago. 2023.

SWALES, John M. *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP, 2002[1990].

TAGNIN, Stella E. O. Glossário de linguística de corpus. In: *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011. p. 357-361.

TEMMERMAN, R. *Towards new ways of Terminology description*. Amsterdam: John Benjamins, 2000.

TUTIN, Agnès. Autour du lexique et de la phraséologie des écrits scientifiques. *Revue Française de Linguistique Appliquée*, n. 2, v. XII, p. 5-14, 2007.

Sobre os autores

Alexia Gonçalves Pokorski

Bacharela em Letras – Tradutor Português-Espanhol da UFRGS. Foi bolsista Português-Espanhol do Grupo de Pesquisa Termisul entre 2020 e 2022. Atualmente é revisora da Fundação Universidade Empresa de Tecnologia e Ciências (Fundatec).

Ana Letícia Prado de Campos

Bacharela em Letras – Tradutor Português e Espanhol, da UFRGS. Foi bolsista de Português-Russo no Grupo de Pesquisa Termisul entre 2020 e 2022. Atualmente, é mestranda na área de Estudos do Léxico e Tradução do PPG-Letras, UFRGS.

Cleci Regina Bevilacqua

Professora titular do Departamento de Línguas Moderna e professora convidada do programa de Pós-Graduação em Letras da UFRGS, na linha de pesquisa Estudos do Léxico e da Tradução. Doutora em Linguística Aplicada pelo Instituto Universitário de Linguística Aplicada, Universidade Pompeu Fabra, Barcelona. Tem como áreas de pesquisa a Fraseologia Especializada, a Terminologia e a Tradução e suas interfaces com a Linguística de *Corpus*. É autora de livros e artigos publicados nessas áreas. É pesquisadora do Grupo de Pesquisa Termisul e líder do Grupo de Pesquisa Termisul no CNPq e pesquisadora do CNPq (PQ2).

Cristiane Krause Kilian

Doutora em Letras (Teorias do Texto e do Discurso) pela Universidade Federal do Rio Grande do Sul e mestre em Filologia Germânica e Filologia Românica pela Universidade Georg-August Göttingen, Alemanha. É professora de alemão e português na Faculdade Instituto Ivoti e tradutora pública e intérprete comercial pelo Tribunal Regional de Göttingen, Alemanha, e pela Junta Comercial do Estado do Rio Grande do Sul para

as línguas portuguesa e alemã. Como pesquisadora, faz parte do Grupo de Pesquisa Termisul.

Denise Regina de Sales

Professora de Língua, Cultura e Literatura Russa do Instituto de Letras da UFRGS. Integrante do PPG-Letras UFRGS nos Estudos da Linguagem (Lexicografia, Terminologia e Tradução: Relações Textuais) e nos Estudos de Literatura (Teoria, Crítica e Comparatismo). Mestre e doutora em Literatura e Cultura Russa pela Universidade de São Paulo. Tradutora de literatura russa, tendo publicado, entre outros, *Minha vida e Três anos*, de Anton Tchêkhov, e *Contos de Kolimá*, de Varlam Chalámov. Trabalha atualmente nas áreas de Estudo de Tradução; Língua, Literatura e Cultura Russa; Tradução literária; Tradução e Terminologia; Tradução e Estudos Culturais.

Iago Marques Barragan

Graduando em Bacharelado em Letras – Tradutor Português-Espanhol da UFRGS. Bolsista Português-Espanhol do Grupo de Pesquisa Termisul de 2018 a 2022.

Manuela Arcos Machado

Mestre e doutoranda em Estudos do Léxico e Tradução. Bolsista CNPq pelo Programa de Pós-Graduação em Letras da UFRGS e pesquisadora voluntária no Grupo de Pesquisa Termisul desde 2017.

Márcia Moura da Silva

Mestre e doutora em Estudos da Tradução pela Universidade Federal de Santa Catarina (UFSC). Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação do Instituto de Letras da Universidade Federal do Rio Grande do Sul (UFRGS), nas linhas de pesquisa Estudos do Léxico e da Tradução: Relações Textuais. Interesses de pesquisa: formação do tradutor, tradução literária, tradução infantojuvenil e tradução técnico-científica. Pesquisadora do Grupo de Pesquisa

Termisul – Projeto Terminológico Cone Sul e coordenadora do Grupo de Pesquisa ABREVITRAD – Abreviatura em Tradução. Tradutora de língua inglesa.

Marina Canofre dos Santos

Graduanda em Bacharelado em Letras – Tradutor Português-Alemão da UFRGS. Bolsista do Português-Russo do Grupo de Pesquisa Termisul desde 2017 e bolsista de língua russa do NELE (Núcleo de Ensino de Línguas em Extensão).

Patrícia Chittoni Ramos Reuillard

Doutora em Letras pela Universidade Federal do Rio Grande do Sul. Professora associada do Departamento de Línguas Modernas e do Programa de Pós-Graduação do Instituto de Letras da UFRGS, na linha de pesquisa Estudos do Léxico e da Tradução. Áreas de atuação: Estudos de Tradução, Sociologia da Tradução, Terminologia, Neologia e Interpretação Comunitária. Membro do Núcleo de Estudos de Tradução Olga Fedossejeva (NET) do Instituto de Letras. Pesquisadora do Grupo de Pesquisa Termisul, vice-líder do Grupo de Pesquisa do CNPq Termisul. Coordenadora do Grupo de Trabalho em Estudos da Tradução da Anpoll (2021-2023). Autora de livros e artigos publicados nessas áreas de atuação. Tradutora de língua francesa.

Sandra Dias Loguercio

Doutora em Ciências da Linguagem pela Universidade Sorbonne Nouvelle (Paris III), mestre em Teorias do Texto e do Discurso e bacharela em Letras Português-Francês pela Universidade Federal do Rio Grande do Sul (UFRGS). É professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras da UFRGS, nas linhas de pesquisa Estudos do Léxico e da Tradução: Relações Textuais e Sociedade, (Inter) Textos Literários e Tradução nas Literaturas Estrangeiras Modernas. É membro do Projeto Termisul, do Núcleo de Estudos de Tradução Olga Fedossejeva (NET), do Núcleo de Estudos Avançados de Linguagens

(NAIL) e do Laboratório para a Formação Continuada de Professores de Francês (LaboFPLC). Como pesquisadora, desenvolve trabalhos em estudos do léxico em interfaces com letramentos e tradução, bem como voltados para abordagens pedagógicas na formação leitora e tradutória.

Tainara Cecília Silveira Balt

Graduanda em Bacharelado em Letras – Tradutor Português-Francês na UFRGS. Bolsista de Iniciação Científica Português-Francês do Grupo de Pesquisa Termisul de agosto de 2019 a janeiro de 2022.

Chave de respostas das atividades propostas

Capítulo 1 – Quando a teoria e a prática se encontram

O capítulo 1 não possui atividades por ser um capítulo teórico e que embasa os demais capítulos do livro.

Capítulo 2 – As decisões prévias

As respostas para as atividades propostas no capítulo 2 dependem das obras selecionadas para a realização das atividades, razão pela qual não apresentamos um gabarito.

Capítulo 3 – Constituição de *corpora*: critérios de coleta, limpeza e organização

As respostas para as atividades propostas no capítulo 3 dependem da área a ser selecionada para a construção de *corpus*, razão pela qual não apresentamos um gabarito.

Capítulo 4 – Seleção de unidades terminológicas: estratégias de extração e princípios de identificação

Exercício 1: O termo definido no trecho do *Corpus* Papel é *arquivo*. Nesse fragmento, o termo apresenta uma frequência de cinco ocorrências. Além disso, o termo *arquivo* aparece acompanhado pelo verbo *definir* em três contextos definitórios, sendo eles: 1) “[...] o arquivo é definido como: um conjunto de documentos produzidos e recebidos por órgãos públicos (...)”, 2) “[...] o arquivo não se define pela forma dos documentos ou por sua origem, mas pela razão para que foram criados e por sua forma de acumulação orgânica” e 3) “[...] os elementos que definem os arquivos podem ser resumidos em três fatores que são abstratos [...]”.

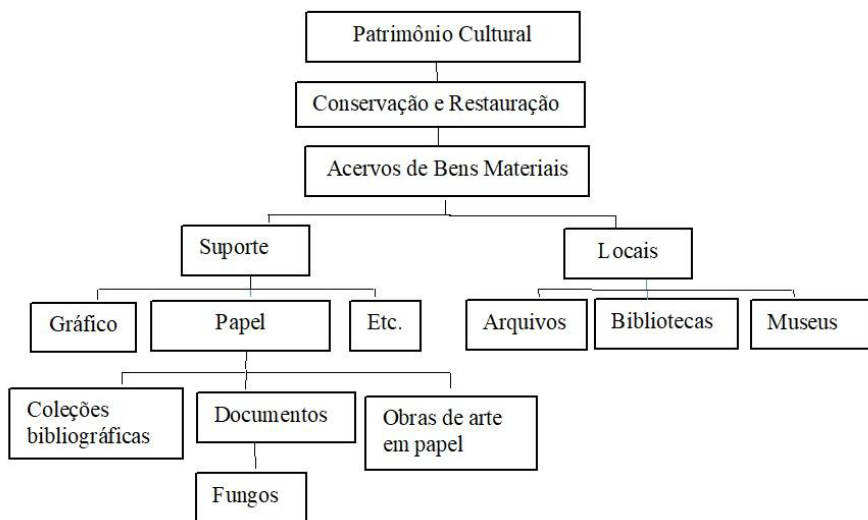
Exercício 2: As UFEs formadas a partir do termo *arquivo* são do tipo colocação (nesse caso, UFE eventivas), pois estão formadas por [verbo + termo]

ou [nominalização + de + termo]. São elas: *abrigar arquivo, organização de arquivo, conservação de arquivo, catalogação de arquivo e microfilmagem de arquivo*.

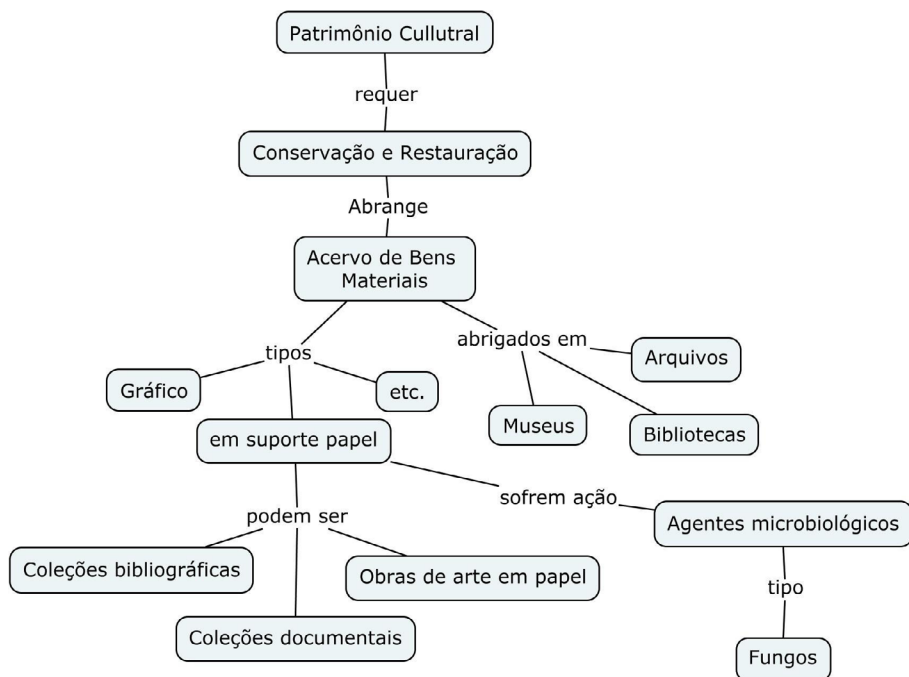
Exercício 3: A área de conhecimento pode ser identificada, mais amplamente, como sendo do **Patrimônio Cultural** (cf. linhas 1, 2 e 5), e mais especificamente, como a de **Conservação e Restauração**, vista na referência ao *corpus* de onde foi extraído o texto. Já o assunto abordado é **fungos em acervo de papel** (introduzido nas linhas 14 a 16 e especificado nas linhas 20, 23, 29 e 30). Isso é feito em um **artigo científico**, gênero identificado pelo registro escrito, pela estrutura textual-discursiva (texto segmentado em parágrafos que trazem contextualização da área e do tema, justificativa da pesquisa, indicação do objeto de estudo e dos objetivos etc.) e por unidades lexicais e fraseológicas que remetem mais especificamente ao relato científico.

Exercício 4:

Sugestão de árvore de domínio



Sugestão de mapa conceitual



Exercício 5: O léxico relativo ao gênero artigo científico (também chamado de léxico metacientífico) torna-se saliente no excerto a partir da linha 14, com *No presente trabalho optou-se por*, em que **trabalho** faz referência ao próprio artigo e a fórmula introduz o tema geral do estudo. Também podem ser identificadas as seguintes unidades: **orientar esta pesquisa**, **esta pesquisa pretende**, **por meio de uma investigação** (l. 26), **estudo de caso** (l. 28), [estudar] **métodos de tratamento para** (l. 29), **a pesquisa busca** (l. 30). Também podemos pensar em palavras como: **trabalho**, **pesquisa**, **investigação**, **estudo de caso**, **estudar**, **método(s)**.

Exercício 6: c / d / e / a / b

Capítulo 5 – A ficha terminológica

Exercício 1:

TERMO: água

Língua: português

Contexto: No tanque superior se dá o processo da reenfibragem, que é a passagem de uma solução de água + polpa de papel através de uma tela semipermeável onde está o documento a ser restaurado. Como resultado esperado temos o depósito da polpa nas áreas do documento onde houve perdas de material. No tanque inferior armazena-se a água após o processo de reenfibragem que, por ser deionizada e trafilada, é de custo elevado, portanto não deve ser desperdiçada. (ptPP023)

Ver também:

água quente

água deionizada

água destilada

água desmineralizada

Equivalente(s) em Inglês:

water 2

Equivalente(s) em Espanhol:

agua 2

Equivalente(s) em Francês:

eau 2

Equivalente(s) em Italiano:

acqua 2

Equivalente(s) em Russo:

вода 2 [voda]

Exercício 2: Como explicado no capítulo, a ficha vai variar de acordo com os diversos fatores envolvidos. Lembre-se de que ela costuma ter Entrada; Categoria gramatical, Gênero e Número; Fonte da entrada; Definição; Fonte da definição; Contexto; Fonte do contexto; Remissivas; Equivalentes; e Notas.

Exercício 3: ver respostas do exercício 1.

Capítulo 6 – Busca e identificação de equivalentes em línguas estrangeiras

Exercício 1:

Língua	Termo	Equivalente
Espanhol	cartão alcalino	cartón libre de ácido
Francês	envelhecimento do papel	vieillessement du papier
Inglês	atmosfera anóxica	anoxic atmosphere
Italiano	banho aquoso	lavaggio acquoso
Russo	solubilidade de tintas	водное растворение чернил [vodnoe rastvorienie tchernil]

Para identificar os equivalentes das atividades 2 e 3, você pode consultar as bases do grupo Termisul disponíveis em www.ufrgs.br ou outras fontes confiáveis de consulta, como *sites* de universidades, de outros grupos de pesquisa e o portal de periódicos da Capes, por exemplo.