

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 Loguercio

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 Odilio 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

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

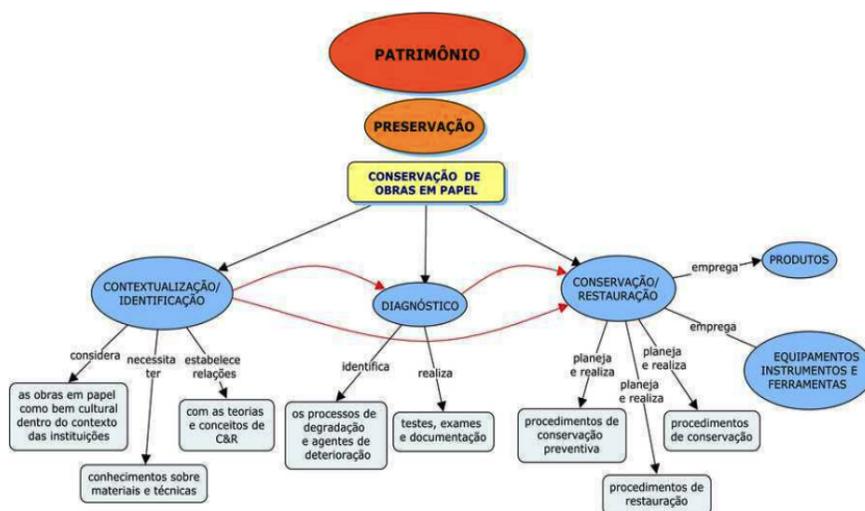
Essa etapa de aproximação com a terminologia da área pode ser seguida ainda da elaboração de uma árvore de domínio ou de um *mapa conceitual* (já tratados no Capítulo 2), que ajuda a compreender efetivamente a arquitetura dessa área: suas subáreas, intersecções, campos temáticos, além de suas relações 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 tâbuas, 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 contains 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 **chavicidade**, 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 chavicidade, 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 artefat
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, conseqüentemente, 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 shows the AntConc software interface. At the top, there is a menu bar with options: Concordance, Concordance Plot, File View, Clusters/N-Grams, Collocates, Word List, and Keyword List. Below the menu bar, it says 'Concordance Hits 202'. The main window displays a list of text hits with line numbers (1-19) and file names (brMA011.txt to brMA030.txt). The search term 'de acordo com' is highlighted in pink in the text. Below the main window, there is a search window with the search term 'de acordo com' and options for 'Words', 'Case', and 'Regex'. The search window also has a 'Search Window Size' dropdown set to 60, and a 'Show Every Nth Row' dropdown set to 1. At the bottom, there are 'Kwic Sort' options for Level 1, 2, and 3, and a 'Clone Results' button.

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

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*. Montevidéo:

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-associacao-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.

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 [...]”.

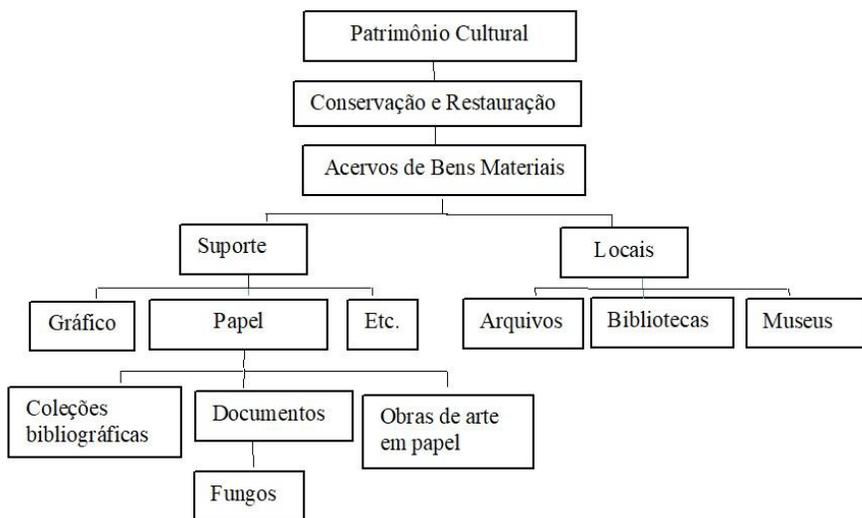
Exercício 2: As UFEs formadas a partir do termo *arquivo* são do tipo colocação (nesse caso, UFE eventivas), pois estão formadas por [verbo + termo]

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

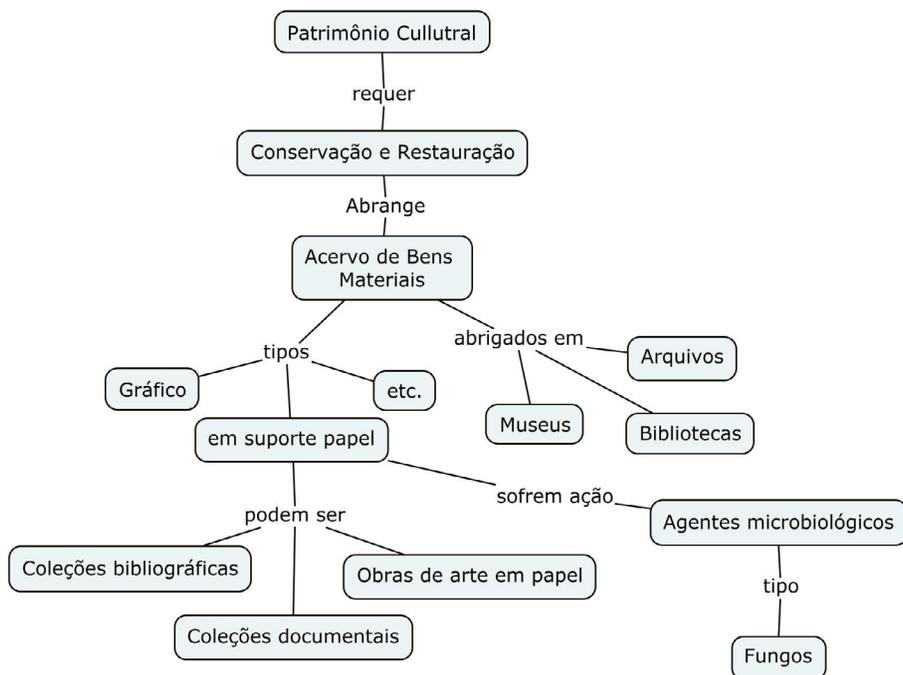
Exercício 3: A área de conhecimento pode ser identificada, mais amplamente, como sendo do **Patrimônio Cultural** (cf. linhas 1, 2 e 5), e mais especificamente, como a de **Conservação e Restauração**, vista na referência ao *corpus* de onde foi extraído o texto. Já o assunto abordado é **fungos em acervo 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.