CRÔNICA IDENTIFICAÇÃO DE UNIDADES FRASEOLÓGICAS ESPECIALIZADAS EVENTIVAS DE LÍNGUA ESPANHOLA NO ÂMBITO DA RESTAURAÇÃO E CONSERVAÇÃO: ESTUDO PILOTO

Manuela Arcos Machado

Submetido em 28 de maio de 2018.

Aceito para publicação em 01 de novembro de 2018.

Cadernos do IL, Porto Alegre, n.º 56, mês de novembro. p. 160-175

POLÍTICA DE DIREITO AUTORAL

Autores que publicam nesta revista concordam com os seguintes termos:

- (a) Os autores mantêm os direitos autorais e concedem à revista o direito de primeira publicação, com o trabalho simultaneamente licenciado sob a Creative Commons Attribution License, permitindo o compartilhamento do trabalho com reconhecimento da autoria do trabalho e publicação inicial nesta revista.
- (b) Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada nesta revista (ex.: publicar em repositório institucional ou como capítulo de livro), com reconhecimento de autoria e publicação inicial nesta revista.
- (c) Os autores têm permissão e são estimulados a publicar e distribuir seu trabalho online (ex.: em repositórios institucionais ou na sua página pessoal) a qualquer ponto antes ou durante o processo editorial, já que isso pode gerar alterações produtivas, bem como aumentar o impacto e a citação do trabalho publicado.
- (d) Os autores estão conscientes de que a revista não se responsabiliza pela solicitação ou pelo pagamento de direitos autorais referentes às imagens incorporadas ao artigo. A obtenção de autorização para a publicação de imagens, de autoria do próprio autor do artigo ou de terceiros, é de responsabilidade do autor. Por esta razão, para todos os artigos que contenham imagens, o autor deve ter uma autorização do uso da imagem, sem qualquer ônus financeiro para os Cadernos do IL.

POLÍTICA DE ACESSO LIVRE

Esta revista oferece acesso livre imediato ao seu conteúdo, seguindo o princípio de que disponibilizar gratuitamente o conhecimento científico ao público proporciona sua democratização.

<http://seer.ufrgs.br/cadernosdoil/index>

Quinta-feira, 22 de novembro de 2018.

IDENTIFICAÇÃO DE UNIDADES FRASEOLÓGICAS ESPECIALIZADAS EVENTIVAS DE LÍNGUA ESPANHOLA NO ÂMBITO DA RESTAURAÇÃO E CONSERVAÇÃO: ESTUDO PILOTO

IDENTIFICATION OF EVENTIVE SPECIALIZED PHRASEOLOGICAL UNITS IN SPANISH IN THE SCOPE OF CONSERVATION AND RESTORATION: A PILOT STUDY

Manuela Arcos Machado*

RESUMO: Este trabalho trata das Unidades Fraseológicas Especializadas Eventivas (UFE eventivas) e tem como objetivo propor uma metodologia para sua identificação em corpora textuais da linguagem da área de Conservação e Restauração de bens materiais em suporte papel em língua espanhola. As UFE eventivas se caracterizam por transmitir ações e processos especializados de um âmbito do conhecimento. Como metodologia, extraímos os colocados dos cinco termos de maior keyness score no corpus textual que esses termos conformam. Como resultado, foi possível estabelecer uma metodologia de identificação e extração de unidades fraseológicas especializadas, utilizando os recursos da linguística de corpus, seguindo critérios quantitativos e qualitativos. Também foi possível observar as variações morfossintáticas e as correlações semânticas entre as diferentes unidades.

PALAVRAS-CHAVE: fraseologia especializada; terminologia; linguística de corpus.

ABSTRACT: This study deals with Eventive Specialized Phraseological Units (Eventive SPUs) and aims to propose a methodology for their identification in textual corpora of the language of the field of Conservation and Restoration of material assets in paper in Spanish. Eventive SPUs are characterized by conveying knowledge concerning processes and actions in the scope of knowledge. As methodology, we researched the collocates of the five terms with higher keyness score in the textual corpus that these terms conform. As results, it was possible to establish a methodology to identify phraseological units, using corpus linguistics' tools, based on quantitative and qualitative criteria. Furthermore, it was possible to observe the morphosyntactic variation and the semantic correlation between different units.

KEYWORDS: specialized phraseology; terminology; corpus linguistics.

1 Introdução

A fraseologia e seu objeto de estudo, as unidades fraseológicas, têm despertado um interesse crescente nos estudos terminológicos nos últimos anos, de um ponto de vista terminográfico, isto é, para a elaboração de materiais terminológicos, mas também

*Mestranda na área de Lexicografia, Terminologia e Tradução do Programa de Pós-Graduação do Instituto de Letras da Universidade Federal do Rio Grande do Sul (PPGLet UFRGS), bolsista de produtividade CAPES, arcosmanuela@gmail.com.

lexicológico, no que se refere a sua definição, caracterização e critérios para seu reconhecimento. Como consequência desse grande interesse, coexistem, na literatura da área, diversas perspectivas e definições dessas unidades.

O objetivo deste trabalho é apresentar uma proposta metodológica de identificação e extração de Unidades Fraseológicas Especializadas Eventivas (UFE eventivas) em corpora textuais por meio de ferramentas computacionais. Dessa forma, este trabalho se insere nos estudos da fraseologia especializada e, portanto, no âmbito da Terminologia.

Apesar da diversidade de definições e denominações que a literatura oferece para as unidades da fraseologia especializada (ver L'HOMME; BERTRAND, 2000; HAUSMANN, 1990; CABRÉ; LORENTE; ESTOPÀ, 1996, entre outros), centraremos nosso estudo nas UFE eventivas conforme a concepção de Bevilacqua (2004), que as considera como unidades formadas por um núcleo terminológico na forma de um termo e um núcleo eventivo na forma de um verbo, nominalização ou particípio, por exemplo, *consumir energia*, *consumo de energia* e *energia consumida*. Portanto, sua constituição parte de um núcleo eventivo que se refere a ações e processos especializados de um âmbito do conhecimento.

O reconhecimento das UFE eventivas se justifica pelo papel importante que tanto os núcleos terminológicos como os núcleos eventivos desempenham na transmissão de conhecimento dentro da linguagem especializada, uma vez que são responsáveis por comunicar as ações próprias de um âmbito do conhecimento. Essas unidades também mostram a convencionalidade das áreas de conhecimento, sendo, portanto, relevantes para o tratamento da expressão textual por tradutores ou outros mediadores de comunicação.

Nessa direção, o reconhecimento dos diferentes tipos de unidades fraseológicas especializadas é, também, importante para a elaboração de produtos terminográficos, já que, de acordo com a finalidade ou objetivo que orienta sua elaboração, se faz necessário estabelecer limites entre as diferentes unidades para decidir quais estarão representadas e de que maneira (BEVILACQUA, 2004).

Para cumprir com os objetivos, em um primeiro momento, apresentamos a proposta de definição das UFE eventivas (BEVILACQUA, 2004) dentro do marco teórico e metodológico da Teoria Comunicativa da Terminologia (TCT) (CABRÉ, 2002). Em um segundo momento, descrevemos o *corpus* de análise e os passos que seguimos para a identificação e extração das UFE eventivas. Por fim, ilustramos os resultados quantitativos e qualitativos a que chegamos através da metodologia proposta.

Ressaltamos que este trabalho está inserido em um projeto de pesquisa maior, realizado junto ao grupo TERMISUL (Projeto Terminológico Cone Sul)¹, cujo propósito é identificar e reunir, em uma base de dados, a terminologia da área da Conservação e Restauração de Bens Culturais Móveis em suporte papel. Portanto, a pesquisa desenvolvida está direcionada e interessa a tradutores e redatores técnicos, entre outros.

2 Posicionamento teórico-metodológico

¹ Grupo de pesquisa teórica e aplicada em Terminologia e Terminografia da Universidade Federal do Rio Grande do Sul (UFRGS). <http://www6.ufrgs.br/termisul/>.

Este trabalho se insere no quadro teórico da Teoria Comunicativa da Terminologia (TCT) (CABRÉ, 2002), que concebe a linguagem especializada não em oposição à língua geral, mas como uma manifestação da mesma em contextos específicos de comunicação. Nessa perspectiva, o termo é definido como uma unidade léxica poliédrica que, utilizada em certas condições discursivas, adquire um valor especializado e, portanto, é, ao mesmo tempo, igual e diferente das unidades léxicas de uma língua – palavras, na lexicologia. Em vista disso, a TCT contempla a Terminologia como o conjunto de Unidades de Significação Especializada (doravante USE) e propõe seu tratamento a partir do modelo das portas – uma perspectiva tripla que permite acessar essas unidades desde uma teoria da linguagem, uma teoria do conhecimento e uma teoria da comunicação. A entrada da teoria linguística – a que se aplica a este trabalho – compreende as USE como unidades de forma e conteúdo que assumem um valor especializado no uso efetivo dentro do âmbito de uma especialidade. Portanto, é no texto especializado que seu significado e suas condições de uso se materializam.

Entre as USE, segundo Cabré (2002), encontram-se unidades não linguísticas (que pertencem a sistemas simbólicos não naturais) e unidades linguísticas (que pertencem à língua natural). As unidades linguísticas especializadas podem se descrever a partir de regras e condições formais dos diferentes componentes da gramática, como a morfologia (morfemas), o léxico (unidades léxicas), os sintagmas (unidades fraseológicas) – objeto de estudo deste trabalho – e as unidades sintáticas (frases), mas também deverão incluir elementos semânticos (significado, por meio de uma definição) e pragmáticos (condições de uso).

Quanto às USE fraseológicas, apesar da diversidade conceitual a respeito dessas unidades na bibliografia da área, como já mencionado, neste piloto serão identificadas e descritas as Unidades Fraseológicas Especializadas Eventivas (doravante UFE eventivas), conforme a proposta de Bevilacqua (2004). A autora considera as UFE eventivas como o “ambiente” em que ocorre uma unidade terminológica (UT) ou como a combinação de uma UT e seus coocorrentes. As UFE eventivas se diferenciam das UT porque representam as ações e processos específicos de um âmbito de especialidade, enquanto as UT são unidades léxicas simples ou sintagmáticas, de caráter denominativo-conceitual e referencial, que denominam um nó da estrutura conceitual de um âmbito de especialidade. Na área da conservação e da restauração em língua espanhola, temos, como exemplo de cada uma dessas unidades *tinta*, *tinta ferrogálica* e *emplear tinta ferrogálica*, sendo:

UT simples → *tinta*

UT sintagmática → *tintas ferrogálicas*

UFE eventiva → *emplear tinta ferrogálica*.

Assim, as UFE eventivas são unidades sintagmáticas formadas por um ou mais termos – que constituem seu núcleo terminológico (NT) – e por um núcleo eventivo (NE), de caráter terminológico ou não, que pode se manifestar como verbo, nome verbal ou participio:

a) [NE]_v + [NT]_N² *consumir energía*

² O [NE]_v é formado por um verbo (que indica ação ou processo especializado) que seleciona o [NT]_N, cuja função sintática é de objeto direto. Contudo, a estrutura [NE]_V + [NT]_N também se manifesta em UFE eventivas cujo NE é um verbo e cujo NT tem função sintática de sujeito, como em *la luz incide* (LORENTE; BEVILACQUA; ESTOPÀ, 2002).

- b) [NE]_{Ndev} + [NT]_{SP} *consumo de energía*
 c) [NT]_N + [NE]_{PartAdj} *energía consumida*³ (BEVILACQUA, 2004)

O NT é de caráter nominal, conceitual e referencial e denomina um nó ou núcleo de conhecimento no mapa conceitual da área de especialidade, podendo aparecer como termo simples ou termo sintagmático. O NE, por sua vez, é de categoria verbal ou derivada de verbo (nomes deverbais e participípios), é relacional e denota os processos e ações próprias de determinado âmbito de especialidade. Entre os dois núcleos – terminológico e eventivo – se estabelecem relações semânticas e pragmáticas, determinadas pelo seu uso no âmbito temático, que conferem à unidade características como certo grau de estabilidade e fixação e uma frequência relevante nos textos especializados (BEVILACQUA, 2004).

Para a identificação das UFE eventivas em nosso *corpus* de análise, tomamos a proposta teórica e metodológica da Terminologia Textual (BOURIGAULT; SLODZIAN, 2004), que entende a análise terminológica como uma tarefa descritiva das estruturas lexicais existentes num *corpus* textual, e que deve, portanto, partir de uma análise desse mesmo *corpus*. À vista disso, a Terminologia Textual propõe uma abordagem descritiva da linguagem especializada na sua realização efetiva, isto é, no texto especializado. Nesse sentido, a tarefa de estabelecer uma terminologia de certa área de conhecimento deve partir da análise de *corpora* textuais, uma vez que é nos textos produzidos e utilizados pela comunidade de especialistas que estão expressos e disponíveis os conhecimentos compartilhados por essa comunidade.

Antes de apresentar a metodologia de análise do *corpus* textual, apresentamos, nas próximas, seções informações sobre o projeto em que se insere este estudo piloto, o modo como o *corpus* de conservação e restauração do espanhol foi compilado e quais critérios seguimos para a identificação das UFE eventivas.

3 Estudo piloto

Este trabalho piloto está inserido em um projeto de pesquisa maior, vinculado ao Grupo TERMISUL, intitulado *A linguagem do patrimônio cultural brasileiro: conservação dos bens culturais móveis*, cujo foco é a linguagem da área de conservação dos bens culturais móveis em suporte papel, isto é, de acervos documentais, bibliográficos e arquivísticos⁴. O projeto tem como objetivo identificar e organizar a terminologia da área em uma base de dados terminológica *on-line* em uma perspectiva multilíngue.

Para tanto, o projeto tem como língua de partida o português e visa estabelecer equivalentes em espanhol, alemão, francês, inglês, italiano e russo. Seu público-alvo são pesquisadores, professores e estudantes da área, bem como profissionais que atuam na produção de textos, como tradutores, assessores linguísticos e redatores técnicos, entre outros. Nosso foco, neste trabalho, é estabelecer uma metodologia de extração das UFE eventivas a partir do *corpus* textual da área de conservação e restauração em língua espanhola.

³ V = verbo; N = nome; Ndev = Nome deverbal; SP = sintagma preposicionado; PartAdj = Participípio/Adjetivo

⁴ Não abrangem mapas e gráficos, nem material fotográfico.

Nas próximas seções, explicamos como o *corpus* do espanhol foi compilado, quais critérios levamos em consideração para a identificação das UFE eventivas, qual metodologia utilizamos para extrai-las e a que resultados parciais chegamos.

3.1 Corpus de análise

O *corpus* do espanhol foi elaborado por bolsistas colaboradores do TERMISUL, seguindo os seguintes critérios para a seleção de textos: a) conterem as palavras-chave “documento, documentação, conservação, papel, patrimônio, preservação, restauração e restauro”; b) pertencerem a gêneros acadêmicos – livros, manuais, revistas científicas, trabalhos de conclusão de curso, dissertações, teses e boletins informativos de associações da área; c) estarem incluídos em fontes confiáveis – *sites* de universidades e instituições de pesquisa, dentre outros, cuja língua original fosse o espanhol.

Atualmente, o *corpus* conta com 124 textos oriundos de países hispano-falantes, como Argentina (37 textos), Chile (20 textos), Espanha (30 textos), México (16 textos) Paraguai (01 texto) e Uruguai (20 textos). Todos os textos coletados passaram por um processo de limpeza (retirada de todas as informações não significativas para a pesquisa), por um processo de conversão para o formato *.txt*, e, por fim, por catalogação. Todos esses procedimentos fazem parte de uma metodologia para a construção de *corpus*, de modo a garantir a qualidade na posterior extração das unidades terminológicas. Hoje, o *corpus* de língua espanhola contém 805,519 *types* e 928,784 *tokens*, com o objetivo de atingir um milhão de palavras.

3.2 Critérios e método de análise

Para estabelecer nossa metodologia de identificação das UFE eventivas da área da conservação e restauração no *corpus* do espanhol, consideramos seus critérios de constituição: critério linguístico (sintáticos e semânticos), critério pragmático-discursivos e critério quantitativo (BEVILACQUA, 2004).

Com relação ao critério linguístico (sintático e semântico), consideramos as estruturas morfossintáticas superficiais V+N; N+SP (prep+(art+)N) e N+Particípio derivadas das estruturas [NE]_v + [NT]_N, [NE]_{Ndev} + [NT]_{SP} e [NT]_N + [NE]_{PartAdj}, respectivamente. Portanto, as UFE eventivas devem incluir, no mínimo, uma unidade terminológica (NT simples ou sintagmático) referente à área da conservação e restauração, bem como um elemento eventivo (NE), na forma de verbo, particípio ou nominalização, que represente as ações e os processos caracterizadores da área, conforme ilustra o Quadro 1.

Quadro 1 – Estruturas das UFE eventivas

[NE] _V + [NT] _N	<i>sintetizar celulosa</i>
[NE] _{Ndev} + [NT] _{SP}	<i>degradación de celulosa</i>
[NT] _N + [NE] _{PartAdj}	<i>celulosa blanqueada</i>

Fonte: a autora

Quanto ao critério pragmático-discursivo, salientamos duas características do texto especializado: a fixação temática e a fixação do texto pelas propriedades

pragmático-discursivas. Dessa forma, as UFE eventivas adquirem valor especializado não somente pela temática (conservação e restauração), mas também pela função comunicativa que desempenham no contexto de uso.

Antes de explicar a metodologia de extração de UFE eventivas que seguimos, ressaltamos que tanto a pertinência temática quanto a função pragmático-discursiva interferiram, neste trabalho, no terceiro critério: o quantitativo. Embora a alta frequência seja um traço típico das UFE eventivas no texto especializado, em função das características temáticas e pragmáticas, decidimos incluir itens com frequências baixas ou iguais a um no *corpus*. Optamos por considerar as *hápax legomena* por se tratar de um estudo prospectivo, cujo propósito é estabelecer a metodologia de identificação das unidades fraseológicas. Dessa forma, esse recorte quantitativo nos permitiu selecionar:

- unidades que, embora com uma frequência baixa, possuem valor especializado no âmbito;
- unidades que apresentavam variação morfossintática (*eliminación de lignina/eliminación de la lignina, incrustación de lignina/incrustaciones de lignina*);
- unidades que possuíam um mesmo NT e um mesmo NE, mas que apresentavam outros termos inseridos (*reacción de las tintas/reacción adversa de las tintas, celulosa producida/celulosa bacteriana producida*);
- unidades que podem ser consideradas sinônimas⁵ (*migración de tinta/traspaso de tinta*).

Para estabelecer o método de extração das UFE eventivas de nosso *corpus* textual, usamos como metodologia a Linguística de *corpus* (BERBER SARDINHA, 2004), por meio da ferramenta *Sketch Engine* (SE)⁶ e seus recursos, como o concordanciador, isto é, um gerador de contextos de palavra, os filtros de identificação de colocados e os índices estatísticos.

Por se tratar de um trabalho prospectivo, cujo objetivo é a testagem das ferramentas e dos recursos informáticos para propor uma metodologia de identificação das unidades fraseológicas, neste primeiro momento, partimos dos cinco termos de maior *keyness score*⁷ do nosso *corpus* de análise. Para tanto, num primeiro momento, tivemos que identificar os termos simples para, em seguida, partindo desses termos, identificar as UFE eventivas. Todo o processo se deu nas três etapas que descrevemos na continuação.

Na primeira etapa, levantamos os 100 primeiros candidatos a termos de maior *keyness score* através da ferramenta extratora de termos do SE. Em seguida, para confirmar o estatuto de termo, os candidatos a NT foram analisados por dois critérios: primeiro pela classificação conforme a árvore de domínio da área da restauração e conservação utilizada pelo grupo TERMISUL e, em seguida, pela análise manual e individual dos seus contextos definitórios no *corpus*, por meio da ferramenta de concordâncias. Nessa etapa, identificamos 49 termos simples (como *papel, humedad*,

⁵ As sinonímias que observamos neste estudo se projetaram, principalmente, no NE, uma vez que nossa identificação das UFE eventivas partiu do NT para o NE.

⁶ *Sketch Engine* é uma ferramenta on-line (<https://www.sketchengine.co.uk/>) que permite criar e fazer pesquisas em *corpora* textuais.

⁷ O *keyness score* é um índice estatístico usado pelo SE para identificar termos de um *corpus* comparado com um *corpus* de referência. Na nossa pesquisa, usamos, para contrastar com o *corpus* de conservação e restauração do TERMISUL, o *corpus* de referência oferecido pelo próprio SE, que inclui variantes do espanhol da América e Europa (*Spanish Web 2011 (esTenTen11, Eu + Am)*).

celulosa) e 13 nominalizações de processos especializados (como *conservación, restauración, preservación*).

Na segunda etapa, tomamos os cinco termos de maior *score* do *corpus* para iniciar a identificação das UFE eventivas: *celulosa, lignina, adhesivo, tinta* e *gramaje*. Cada termo foi submetido a quatro métodos de pesquisa por meio do filtro *word sketch*⁸:

- Modifiers of "x"* – oferece os adjetivos colocados do termo (*celulosa bacteriana, tinta ferrogálica*);
- Verbs with "x" as object* – oferece os verbos cujo colocado é o termo pesquisado com função de objeto direto (*generar celulosa, fotodegradar la lignina*). Esse filtro também direciona para as estruturas cujo NE está no particípio, como em *celulosa generada* ou *adesivo activado*.
- Verbs with "x" as subject* – oferece os verbos cujo colocado é o termo pesquisado com função de sujeito (*el adhesivo penetra, la lignina agrava*);
- Prepositional phrases "de x"* – oferece os colocados (geralmente substantivos e nomes deverbais) que formam sintagmas preposicionados com o termo pesquisado (*oxidación del adhesivo, migración de tinta*).

A Figura 1 exemplifica, com o termo *celulosa*, os quatro métodos de pesquisa da ferramenta *word sketch* que utilizamos.

Fonte: adaptado de <https://www.sketchengine.co.uk/>

celulosa <small>(noun)</small> <small>TERMSUL freq = 879 (946.39 per million)</small>		
a) modifiers of "celulosa" 41.98		
bacteriano +	291	13.74
de celulosa bacteriana		
vegetal	30	10.94
de la celulosa vegetal		
puro	7	9.07
alfa	3	8.04
adyacente	3	8.00
cristalino	3	7.97
b) verbs with "celulosa" as object 17.06		
generar	29	10.85
de celulosa bacteriana generadas		
sintetizar	5	9.94
crecer	5	9.78
blanquear	4	9.55
proporcionar	6	9.44
obtener	12	9.38
digerir	3	9.31
purificar	3	9.24
degradar	3	9.04
producir	12	8.87
deber	4	8.63
c) verbs with "celulosa" as subject 3.98		
crecer	4	11.54
ir	4	9.97
presentar	5	9.85
tener	5	8.23
poder	3	7.20
d) prepositional phrases		
... de "celulosa"	481	54.72
... de "celulosa"		51.88
capa	65	11.72
las capas de celulosa bacteriana		
cadena	23	10.55
de las cadenas de celulosa		
producción	25	10.49
la producción de celulosa bacteriana		
fibra	24	10.45
las fibras de celulosa		
éter	14	9.91
éteres de celulosa		
nitrito	13	9.82
de nitrito de celulosa		

Figura 1 – *Word sketch*.

Cada método de busca pode direcionar a determinadas unidades terminológicas. O método "a" indica os termos sintagmáticos (*celulosa bacteriana*)⁹. O método "b"

⁸ Ferramenta do SE que oferece os colocados de diferentes classes gramaticais da palavra consultada.

⁹ Identificamos os termos sintagmáticos, uma vez que podem compor o NT da UFE eventiva (por exemplo, uso de *tinta ferrogálica*). Contudo, não formam parte das unidades fraseológicas que são o foco

indica unidades de duas estruturas: [NE]_v + [NT]_N (*generar celulosa*) e [NT]_N + [NE]_{PartAdj} (*celulosa generada*)¹⁰. Contudo, para chegar às estruturas com NE no particípio, era necessário observar as concordâncias, uma vez que o filtro oferece o verbo sempre no infinitivo. A busca “c” também leva às UFE eventivas de estrutura [NE]_v + [NT]_N, porém, direciona a itens onde o NE é o verbo e o NT tem função de sujeito sintático (*el adhesivo penetra*). Por fim, a busca “d” aponta para a estrutura [NE]_{Ndev} + [NT]_{SP}, na qual o NE é uma ação ou processo expresso por um nome deverbal e o NT é um sintagma preposicionado, como em *producción de celulosa*.

Na terceira e última etapa, tomamos os dados obtidos pela ferramenta *word sketch* e confirmamos quais unidades conformavam ações ou processos especializados da área. Para isso, verificamos as concordâncias (contextos) de cada uma. Assim, foi possível descartar combinações como *presentar tinta*, que não se refere a uma ação ou processo especializado, mas a uma propriedade do termo *tinta*.

Depois das três etapas, organizamos todas as UFE eventivas extraídas do *corpus* em uma planilha separada pelos métodos de pesquisa a, b, c e d, e com o *score* e frequência absoluta de cada unidade. Essa organização por métodos de pesquisa nos serviu para observar:

- quais estruturas morfossintáticas de UFE eventivas são mais ou menos produtivas na linguagem da conservação e restauração em língua espanhola;
- quais estruturas morfossintáticas apresentam correspondência (*eliminar la lignina/eliminación de lignina, extraer la lignina/extracción de lignina*).
- a sinonímia entre diferentes NE de diferentes estruturas morfossintáticas (*uso de tinta/emplear tinta*).

Na seção seguinte, apresentamos os resultados quantitativos e qualitativos a que chegamos a partir da metodologia seguida para a extração de UFE eventivas.

3.3 Resultados parciais da análise

A metodologia seguida nos permitiu identificar um total de 121¹¹ UFE eventivas a partir dos cinco NT de maior *keyness score* do *corpus* textual analisado. A Tabela 1 expõe o número de unidades extraídas para cada uma das estruturas de UFE eventivas: [NE]_v + [NT]_N (sendo o NT com função sintática de objeto direto e de sujeito), [NE]_{Ndev} + [NT]_{SP} e [NT]_N + [NE]_{PartAdj}.

da nossa identificação, nem fazem parte de nossos resultados por não indicarem ação ou processo especializado e, portanto, não constituírem UFE eventivas.

¹⁰ Nesse caso, em especial, a análise manual das concordâncias foi fundamental, posto que alguns resultados com verbos no particípio poderiam indicar termos sintagmáticos, isto é, casos em que as terminações verbais *ado/ido* têm função de adjetivo. Assim, consideramos UFE eventivas quando o NE no particípio, bem como as nominalizações, herda a estrutura do verbo do qual derivam, e com o qual é possível gerar as três estruturas: [NE]_V + [NT]_N *generar celulosa*, [NE]_{Ndev} + [NT]_{SP} *generación de celulosa*, e [NT]_N + [NE]_{PartAdj} *celulosa generada*.

¹¹ Todas as UFE eventivas extraídas se encontram no Anexo 1, no final deste trabalho, separadas pela estrutura morfossintática e com o índice de frequência que apresentam no *corpus*.

Tabela 1: resultado numérico das UFE eventivas identificadas

Estruturas	UFE eventivas identificadas
[NE] _v + [NT] _N (objeto direto)	39
[NE] _v + [NT] _N (sujeito)	05
[NE] _{Ndev} + [NT] _{SP}	49
[NT] _N + [NE] _{PartAdj}	28

Fonte: a autora.

Os resultados quantitativos de cada estrutura morfossintática das UFE eventivas extraídas sugerem que as UFE eventivas cujo NE são nominalizações são as estruturas mais produtivas na linguagem da conservação e restauração em espanhol e as mais frequentes no *corpus*, enquanto aquelas cujo NT tem função sintática de sujeito são as menos produtivas, ocorrendo somente com dois NT: *adhesivo* e *lignina*.

Quanto à correspondência entre as estruturas morfossintáticas, observamos que os cinco termos pesquisados apresentaram estruturas com correlação absoluta entre sintagmas verbais e nominalizações, como ilustra o Quadro 2:

Quadro 2 – correlação entre sintagmas verbais e nominalizações

[NE] _v + [NT] _N (objeto direto)	[NE] _{Ndev} + [NT] _{SP}	[NT] _N + [NE] _{PartAdj}
<i>eliminar lignina</i>	<i>eliminación de lignina</i>	<i>lignina eliminada</i>
<i>crecer celulosa</i>	<i>crecimiento de celulosa</i>	<i>celulosa crecida</i>
<i>usar adhesivo</i>	<i>uso de adhesivo</i>	<i>adhesivo usado</i>
<i>generar celulosa</i>	<i>generación de celulosa</i>	<i>celulosa generada</i>
<i>producir celulosa</i>	<i>producción de celulosa</i>	<i>celulosa producida</i>

Fonte: a autora

O mesmo ocorreu com *aplicar adhesivo/aplicación de adhesivo*, *decolorar las tintas/decoloración de las tintas* e *determinar el gramaje/determinación del gramaje*, entre outros. Por outro lado, quando o NT tinha função sintática de sujeito, como em *el adhesivo penetra*, o sintagma verbal não apresentou correlação com nominalização.

Quanto à metodologia de busca que utilizamos, observamos que somente os resultados oferecidos pela ferramenta *word sketch* não são suficientes para a identificação de UFE eventivas, sendo necessária a análise manual e individual das concordâncias. Com isso, pudemos:

- confirmar ou descartar quais NE (verbos, participios ou nominalizações) indicavam ações ou processos especializados;
- identificar UFE eventivas com NT sintagmáticos, uma vez que a ferramenta *word sketch* só permite a pesquisa a partir de um termo simples. Desse modo, ao pesquisar o termo *celulosa*, por exemplo, obtivemos como colocado o NE *producir*, mas só foi possível ver que o NT era sintagmático – *producir celulosa bacteriana* – por meio da leitura das concordâncias;
- observar que os termos sintagmáticos, quando ocorrem, costumam aparecer nas UFE eventivas cujo NT é um objeto direto (OD) ou um sintagma preposicionado (SP), como em *formar tinta metaloácida* (OD) e *aplicación del adhesivo termofundente* (SP), não ocorrendo nas demais estruturas;
- verificar que, no método de pesquisa “d” (sintagmas preposicionados colocados ao termo pesquisado), quando o colocado oferecido pelo *word sketch* era um verbo no infinitivo, este não constituía NE. Por exemplo, a

pesquisa “d” apresentou como colocados para o termo *lignina* os verbos no infinitivo *exentar*, *elear*, *definir*. Contudo, ao analisar individualmente as concordâncias, constatamos que o verbo não constituía um NE, mas um adjetivo que se referia a outro elemento que não o termo: [...] *fibra de celulosa de algodón*, *exenta de lignina* [...] [...] *pueden contener cantidades elevadas de lignina* [...] [...] *describir una estructura definida de la lignina* [...].

Por fim, a análise individual das concordâncias também nos permitiu identificar UFE eventivas sinônimas. Reconhecemos 12 grupos de unidades sinônimas:

- *utilizar adhesivo/usar adhesivo*;
- *reblandecer el adhesivo/ablandar el adhesivo*;
- *adhesivo reblandecido/adhesivo ablandado*;
- *eliminar lignina/extraer lignina*;
- *eliminación de la lignina/extracción de la lignina*;
- *emplear tinta/utilizar tinta/usar tinta*;
- *corrimiento de tinta/traspaso de tinta/migración de tinta*;
- *generar celulosa/crecer celulosa/producir celulosa*;
- *generación de celulosa/producción de celulosa/crecimiento de celulosa*;
- *descomposición de celulosa/degradación de celulosa*;
- *degradar celulosa/sintetizar celulosa*;
- *purificar celulosa/blanquear celulosa*.

Novamente, o trabalho manual com as concordâncias, que consistiu na leitura de cada contexto, foi fundamental, uma vez que algumas unidades descontextualizadas poderiam parecer sinônimas. Por exemplo, em *degradar celulosa*, *digerir celulosa* e *sintetizar celulosa*, observamos, por meio dos contextos, que *degradar* e *digerir* funcionavam como sinônimas por se referirem ao mesmo processo, cujo agente são fungos. Contudo, *sintetizar celulosa* é um processo cujo agente são insetos e, portanto, não se refere ao mesmo processo na linguagem especializada.

4 Considerações finais

O propósito deste estudo piloto foi estabelecer uma metodologia de extração de UFE eventivas em *corpus* textual. Julgamos ter alcançado o objetivo inicial e ter sido possível desenhar uma metodologia de identificação de UFE eventivas. Contudo, embora tenhamos partido de um conjunto de critérios que as caracterizam e utilizado uma ferramenta informática (*Sketch Engine*), a seleção manual através das concordâncias e dos contextos das UFE eventivas foi fundamental para:

- compreender sua constituição complexa e determinar seu valor especializado em contextos específicos;
- observar suas possíveis variações morfossintáticas e semânticas (como as correlações entre verbos e nominalizações e a presença de sinônimos);
- estabelecer que os filtros de extração como o *Word Sketch* servem para uma identificação semiautomática que oferece estruturas candidatas a UFE eventivas, que deverão ser confirmadas manualmente, uma vez que a ferramenta se baseia em cálculos estatísticos que podem gerar ruído nos resultados.

Por outro lado, a análise dos dados que extraímos reforça os critérios de caracterização e reconhecimento das UFE eventivas no que se refere a:

- sua conformação estável por um NE e um NT, apesar da sua variação morfossintática;
- sua caracterização como unidades especializadas determinadas pelas propriedades pragmático-discursivas dos contextos em que são usadas;
- considerar relevante, mas não determinante, o critério de frequência, já que o caráter especializado que a unidade adquire no texto parece ter mais peso.

Apesar de considerarmos que o trabalho permitiu um avanço no estabelecimento de uma metodologia de reconhecimento e extração de UFE eventivas em *corpora* textuais, acreditamos que o trabalho abre, como possibilidades futuras:

- a classificação morfossintática das UFE eventivas pela estrutura argumental dos verbos e das nominalizações, a fim de observar mais profundamente como se comportam as unidades fraseológicas da área da conservação e restauração em língua espanhola;
- o teste de uma segunda metodologia de extração partindo do NE, seja dos verbos ou das nominalizações, gerando seus colocados e concordâncias, com a finalidade de observar se diferentes UFE eventivas são identificadas ou determinar se a extração a partir do NE pode ser mais produtiva do que partindo do NT;
- o estabelecimento da medida em que a metodologia seguida e os dados obtidos na extração de UFE eventivas servem como ponto de partida para identificar os equivalentes no português e outras línguas.

O trabalho propõe, como direcionamento futuro, uma análise partindo de um número maior de termos, a fim de identificar e descrever mais unidades fraseológicas da área da conservação e restauração. Por outro lado, este estudo piloto também abre, como caminho possível, a testagem da metodologia aqui proposta para outras línguas, além da possibilidade de estabelecer UFE eventivas equivalentes entre as diferentes línguas. Esses direcionamentos dialogam com a elaboração do produto final do projeto do grupo TERMISUL, posto que as UFE eventivas deverão constituir uma das partes das fichas catalográficas da base de dados terminológica multilíngue voltada para tradutores, redatores, profissionais e estudantes da área da Conservação e Restauração, que está sendo elaborada pelo grupo.

REFERÊNCIAS

BEVILACQUA, Cleci Regina. *Unidades Fraseológicas Especializadas: estado de la cuestión y perspectivas*. 1999. 208 f. Tese (Doutorado em Linguística Aplicada) – Instituto Universitário de Linguística Aplicada (IULA), Universidade Pompeu de Fabra, Barcelona, 1999.

_____. *Unidades Fraseológicas Especializadas: elementos para seu reconhecimento em corpora textuais*. *Revista Intercâmbio*, Montes Claros, vol. XII, p. 215-223, 2003.

_____. *Unidades Fraseológicas Especializadas Eventivas: descripción y reglas de formación en el ámbito de la energía solar*. 2004. 241 f. Tese (Doutorado em Linguística Aplicada) – Instituto Universitário de Linguística Aplicada (IULA), Universidade Pompeu de Fabra, Barcelona, 2004.

BERBER SARDINHA, T. *Linguística de corpus*. Barueri: Manole, 2004.

- BOURIGAULT, D.; SLODZIAN, M. Por uma terminologia textual. *Cadernos de Tradução*, Porto Alegre, n. 17, p. 101-108, out./dez. 2004.
- CABRÉ, M. T. Terminologie et linguistique: la théorie des portes. *Terminologies nouvelles: Terminologie et diversité culturelle*, Bruxelas, n. 21, p.10-15, 2000.
- CABRÉ, M. T.; LORENTE, M.; ESTOPÀ, R. Terminología y fraseología. In: V SIMPOSIO DE TERMINOLOGÍA IBEROAMERICANA, 1996. Cidade do México. *Anais do V Simposio de Terminología Iberoamericana...* Cidade do México: Colegio de México, 1996. p. 67-81.
- HAUSMANN, F. J. Le dictionnaire de collocations. In: HAUSMANN, F. J. et al. *An International encyclopedia of lexicography*. Vol. 1. Berlim/Nova York: Walter de Gruyter, 1990. p. 1010-1019.
- L'HOMME, M. C.; BERTRAND, C. Specialized Lexical Combinations: should they be described as collocations or in terms of selectional restrictions? In: NINTH EURALEX INTERNATIONAL CONGRESS. *Proceedings of the Ninth Euralex International Congress...* Stuttgart: Universitat Stuttgart. 2000. p. 4974-4506.
- LORENTE, M; BEVILACQUA, C.R.; ESTOPÀ, R. El análisis de la fraseología especializada mediante elementos de la lingüística actual. In: VI SIMPÓSIO DA REDE IBERO-AMERICANA DE TERMINOLOGIA, 2002. Lisboa. *Atas do VI Simpósio da Rede Ibero-americana de Terminologia: Terminologia, desenvolvimento e identidade nacional...* Lisboa: Ed. Colibri, p. 635-653.

ANEXO 1 – UFE eventivas identificadas e extraídas

ADHESIVO	Estruturas identificadas	UFE eventivas identificadas	Frequência no <i>corpus</i>
[NE]V + [NT]N (objeto direto)	8	<i>dejar el adhesivo.</i>	1
		<i>aplicar (el) adhesivo;</i>	26
		<i>reblandecer el adhesivo;</i>	2
		<i>ablandar el adhesivo;</i>	1
		<i>poner el adhesivo;</i>	4
		<i>utilizar adhesivo;</i>	1
		<i>usar adhesivos;</i>	3
		<i>introducir adhesivo;</i>	2
[NE]Ndev + [NT]N (sujeito)	1	<i>el adhesivo penetrar</i>	2
[NE]Ndev + [NT]SP	11	<i>uso de adhesivos</i>	7
		<i>uso de adhesivos acuosos</i>	1
		<i>uso de adhesivos termofusibles</i>	1
		<i>uso de adesivos termoplásticos</i>	1
		<i>uso de adhesivos acrílicos;</i>	1
		<i>resquebrajamiento de los adesivos;</i>	2
		<i>aplicación del/ de los adhesivo(s)</i>	4
		<i>aplicación del adhesivo termofundente;</i>	1
		<i>preparación del/ de los adhesivo(s);</i>	2
		<i>oxidación del adhesivo;</i>	2
		<i>encuadernación de adhesivo.</i>	2

[NT]N + [NE]PartAdj	7	<i>adhesivo reblandecido;</i>	1
		<i>adhesivo(s) ensayado(s);</i>	3
		<i>adhesivo(s) seleccionado(s);</i>	4
		<i>adhesivo ablandado;</i>	1
		<i>adhesivo activado;</i>	2
		<i>adhesivo utilizado;</i>	13
		<i>adhesivos usados;</i>	1

LIGNINA	Estructuras identificadas	UFE eventivas identificadas	Frecuência no corpus
[NE]V + [NT]N (objeto direto)	5	<i>ablandar la lignina.</i>	1
		<i>disolver la lignina;</i>	3
		<i>figurar la lignina;</i>	2
		<i>fotodegradar la lignina;</i>	1
		<i>eliminar la lignina;</i>	6
[NE]Ndev + [NT]N (sujeito)	4	<i>la lignina contribuye.</i>	1
		<i>la lignina agrava;</i>	1
		<i>la lignina protege;</i>	1
		<i>la lignina funciona;</i>	1
[NE]Ndev + [NT]SP	7	<i>detección de lignina.</i>	1
		<i>eliminación de (la) lignina;</i>	5
		<i>incrustación(es) de lignina;</i>	2
		<i>oxidación de (la) lignina;</i>	3
		<i>extracción de la lignina;</i>	2
		<i>recoloración de la lignina blanqueada;</i>	1
		<i>concentración de lignina;</i>	1
[NT]N + [NE]PartAdj	2	<i>lignina eliminada;</i>	1
		<i>lignina extraída.</i>	1

TINTA	Estructuras identificadas	UFE eventivas	Frecuência no corpus
[NE]V + [NT]N (objeto direto)	9	<i>obtener (la/s) tinta(s).</i>	3
		<i>reavivar (la) tinta(s);</i>	4
		<i>decolorar las tintas;</i>	4
		<i>recibir la tinta;</i>	4
		<i>emplear tinta;</i>	1
		<i>emplear tinta de carbón;</i>	1
		<i>utilizar tinta;</i>	5
		<i>usar tinta;</i>	4
		<i>formar tinta metaloácida;</i>	1
[NE]Ndev + [NT]N (sujeito)	0	-	
[NE]Ndev + [NT]SP	17	<i>corrosión de tinta(s);</i>	30
		<i>solubilidad de (la/s) tinta(s);</i>	22
		<i>migración de (la/s) tinta(s);</i>	17
		<i>composición de (la/s) tinta(s);</i>	18
		<i>corrimiento de (la/s) tinta(s);</i>	8
		<i>traspado de (la) tinta(s);</i>	7
		<i>oxidación de la(s) tinta(s);</i>	8
		<i>uso de la tinta ferrogálica;</i>	9
		<i>uso de la tinta de carbón;</i>	1
		<i>degradación de (la/s) tinta(s);</i>	6
		<i>fijación de (la/s) tinta(s);</i>	5
		<i>reacción de (la/s) tinta(s);</i>	5
		<i>reacción adversa de las tintas;</i>	1

		<i>solubilización de (la/s) tinta(s);</i>	4
		<i>sangrado de (la) tinta</i>	3
		<i>decoloración de (la/s) tinta(s);</i>	4
		<i>disolución de tintas.</i>	3
[NT]N + [NE]PartAdj	4	<i>tinta confeccionada;</i>	3
		<i>tinta ferrogáfica empleada;</i>	1
		<i>tinta utilizada;</i>	4
		<i>tintas formadas.</i>	2

GRAMAJE	Estructuras identificadas	UFE eventivas	Frequência no corpus
[NE]V + [NT]N (objeto direto)	3	<i>alcanzar gramajes;</i>	1
		<i>determinar el gramaje;</i>	4
		<i>obtener el gramaje.</i>	1
[NE]Ndev + [NT]N (sujeito)	0	-	
[NE]Ndev + [NT]SP	4	<i>determinación del gramaje;</i>	5
		<i>multiplicación del gramaje;</i>	1
		<i>disminución del gramaje;</i>	1
		<i>variación(es) del gramaje.</i>	2
[NT]N + [NE]PartAdj		-	

CELULOSA	Estructuras identificadas	UFE eventivas	Frequência no corpus
[NE]V + [NT]N (objeto direto)	14	<i>generar (la) celulosa;</i>	2
		<i>sintetizar la celulosa;</i>	5
		<i>crecer (la) celulosa;</i>	4
		<i>proporcionar (la/una) celulosa;</i>	6
		<i>obtener (la/una) celulosa;</i>	5
		<i>digerir la celulosa;</i>	3
		<i>purificar la celulosa;</i>	1
		<i>degradar la celulosa;</i>	3
		<i>producir celulosa;</i>	1
		<i>producir celulosa bacteriana;</i>	1
		<i>tratar la celulosa;</i>	1
		<i>contener celulosa;</i>	3
		<i>usar celulosa;</i>	2
		<i>utilizar celulosa.</i>	2
[NE]Ndev + [NT]N (sujeito)	0	-	
[NE]Ndev + [NT]SP	10	<i>producción de celulosa;</i>	11
		<i>producción de celulosa bacteriana;</i>	14
		<i>oxidación de la celulosa;</i>	12
		<i>degradación de la celulosa;</i>	12
		<i>generación de celulosa bacteriana;</i>	11
		<i>aplicación de la celulosa bacteriana;</i>	7
		<i>descomposición de la celulosa;</i>	5
		<i>descomposición oxidativa de la celulosa;</i>	2
		<i>crecimiento de celulosa;</i>	5

		<i>crecimiento de la celulosa bacteriana.</i>	4
[NT]N + [NE]PartAdj	15	<i>celulosa generada;</i>	7
		<i>celulosa bacteriana generada;</i>	6
		<i>la celulosa crecida;</i>	1
		<i>celulosa blanqueada;</i>	4
		<i>celulosa obtenida;</i>	5
		<i>celulosa bacteriana obtenida;</i>	2
		<i>purificada la celulosa;</i>	1
		<i>celulosa bacteriana purificada;</i>	1
		<i>celulosa degradada;</i>	1
		<i>celulosa producida;</i>	4
		<i>celulosa bacteriana producida;</i>	1
		<i>celulosa tratada;</i>	2
		<i>celulosa contenida;</i>	1
		<i>celulosa usada;</i>	1
		<i>nitrato de celulosa utilizado;</i>	1