

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

UNIDADES FRASEOLÓGICAS ESPECIALIZADAS EVENTIVAS:  
NOVAS QUESTÕES SOBRE SEU RECONHECIMENTO EM CORPORA  
TEXTUAIS

Cleci Regina BEVILACQUA (Universidade Federal do Rio Grande do Sul)  
Ana Rachel SALGADO (Universidade Federal do Rio Grande do Sul)  
Diônifer Alan da SILVEIRA (Universidade Federal do Rio Grande do Sul)

*ABSTRACT: This work deals with difficulties met when recognizing Eventive Phraseological Specialized Units in a corpus of Brazilian environmental legislation. These units are characterized by an Eventive Nucleus (verb or a noun derived from a verb) and a Terminological Nucleus and by conveying knowledge concerning processes and actions of Environmental Law.*

*KEYWORDS: Phraseological Specialized Units; Phraseology; Terminology; Specialized Texts.*

O objetivo deste trabalho é apresentar algumas questões referentes ao reconhecimento e extração das Unidades Fraseológicas Especializadas Eventivas (UFE eventivas) em dois corpora. Um constituído por textos da Legislação Ambiental Brasileira e outro, por textos da área da Engenharia Ambiental. Para tanto, apresentaremos, inicialmente, nossa definição de unidade fraseológica. Em seguida, apresentaremos e discutiremos as dificuldades de reconhecê-las, mesmo partindo de alguns critérios específicos estabelecidos em sua definição, como a inclusão de um termo e de um núcleo eventivo e seu caráter especializado. Desse modo, procuramos ampliar a discussão sobre os procedimentos de identificação e seleção da fraseologia especializada, dando seqüência ao trabalho apresentado no 12º InPla.

Salientamos que os resultados aqui apresentados fazem parte de um estudo realizado junto ao projeto TERMISUL (Projeto Terminológico Cone Sul, UFRGS) que pretende identificar as unidades fraseológicas do âmbito do Direito Ambiental, área sobre a qual a equipe de pesquisadores do projeto já desenvolveu dois dicionários.

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

De forma complementar, apresentaremos também dados referentes ao reconhecimento de UFE eventivas no âmbito da Engenharia Computacional subárea de Desenvolvimento de Software e de Sistemas, resultantes de pesquisa desenvolvida para a dissertação de mestrado de Ana Rachel Salgado.

Como definição de UFE eventivas adotamos a proposta de Bevilacqua (2004:16-17) que as define como:

unidades formadas por um núcleo eventivo, considerado como tal por ser de base verbal ou derivada de verbo (nominalização ou particípio), e por um núcleo terminológico (termo). Entre estes dois núcleos se estabelecem relações sintáticas, mas principalmente semânticas, determinadas pelas propriedades do texto em que são utilizadas. Portanto, são unidades que se conformam no e pelo texto em que são utilizadas. Cumprem, tal como os termos, a função de representar e transmitir conhecimento especializado.

Nesta definição, o Núcleo Terminológico (NT) representa um nó de conhecimento na estrutura ou mapa conceitual do Direito Ambiental; é de categoria nominal e possui caráter denominativo (*danos ambientais, ecossistema, meio ambiente, resíduos*, etc.). Por sua vez, o Núcleo Eventivo (NE) é de categoria verbal ou derivada de verbo (nome deverbal ou particípio), é relacional e denota processos e ações próprios da área estudada. Observamos que, para esta pesquisa, consideramos apenas as realizações das UFE eventivas em suas formas verbal e nominal, dado que a sua realização em forma de particípio requer ainda um estudo mais aprofundado.

Como corpus textual utilizamos o conjunto das leis brasileiras referentes ao meio ambiente, excetuando os decretos referentes aos acordos internacionais<sup>1</sup>. Os textos utilizados incluem, portanto, Leis, Decretos, Decretos-lei, Resoluções aprovados no período compreendido entre julho de 1934 e julho de 2004. Este conjunto está formado por 273 textos que perfazem um total de 727.674 palavras.

Para extrair unidades candidatas a UFE eventivas, utilizamos a ferramenta Extraterm iX, que faz parte do programa MultiTerm da Trados. Esta ferramenta permite:

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

- a) armazenar o corpus e obter as informações referentes ao tamanho de cada texto e o número total de palavras do corpus;
- b) estabelecer parâmetros de extração de dados com base em estatística, gerando listas de possíveis candidatos a termos ou, no nosso caso, NE e NT que conformam as UFE eventivas;
- c) gerar concordâncias para estes núcleos. No caso específico desse trabalho, veremos em seguida, que somente geramos as concordâncias para os NE.

Como critérios iniciais para a extração das UFE eventivas tínhamos os seguintes:

1. Critérios semânticos:

- a) inclusão de, no mínimo, uma unidade terminológica simples ou sintagmática referente à área do meio ambiente (*biodiversidade, água, meio ambiente, solo*, etc);
- b) inclusão de um elemento eventivo em forma de verbo ou de nominalização que representasse as ações e processos caracterizadores do Direito Ambiental (*proteger, conservar, contaminar*, etc.).

2. Critérios quantitativos:

- a) frequência igual ou superior a cinco.

3. Critérios pragmático-discursivos:

- a) fixação temática,
- b) fixação pelas propriedades pragmático-discursivas do texto especializado.

Sobre esses dois critérios teceremos alguns comentários. Em relação aos critérios semânticos, pelo trabalho desenvolvido junto ao Termisul, foi possível adquirir um conhecimento sobre os termos do Direito Ambiental bastante sólido. Aprendemos, por exemplo, que *água* é um termo da química ou da geologia, mas também é um termo legal, pois está definido como um *bem comum* e que, portanto, deve ser preservado. Da mesma forma sobre as fraseologias, dado que representam, por sua constituição, ações e processos próprios de determinada área como, por exemplo, *causar dano ambiental*,

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

*proteger a biodiversidade*, etc. (Bevilacqua, 1996; 1999; 2004). Representam, portanto, tal como os termos, conhecimento próprio de um âmbito do saber.

Com respeito à fixação temática, salientamos que está diretamente relacionada aos critérios anteriores, na medida em que é por conter um termo de uma área específica e um NE que se refere aos processos e ações dessa área que uma unidade representa o conhecimento de uma especialidade e adquire, conseqüentemente, valor especializado.

No tocante à fixação determinada pelas propriedades pragmático-discursivas dos textos especializados, acreditamos que as UFE eventivas, tal como os termos, adquirem valor especializado, não só pela temática, mas também por essas propriedades. No caso específico do Direito Ambiental temos textos que se caracterizam por terem uma superestrutura bem definida (caput, artigo, parágrafo, alínea, etc.). Além disso, tem uma função específica que é assegurar a tutela (proteção) da natureza e, para tanto, estabelece sanções e punições para quem não cumprir o que está estabelecido nos diplomas legais. Levar em consideração estas propriedades ajuda a reconhecer as UFE. Assim, selecionamos unidades como *provocar dano ambiental*, *reparar o dano ambiental*. A primeira representa uma ação que causa prejuízos ao meio ambiente (pode ser a instalação industrial em uma área de proteção, o despejo de efluentes não tratados em um rio, etc.). Esses prejuízos devem ser reparados – *reparar o dano* - de alguma forma, seja através de multas, seja através de alguma outra sanção (prisão, prestação de serviços à comunidade, etc.).

Essas constatações nos permitiram excluir as unidades fraseológicas que se referem ao discurso jurídico como um todo, tal como *de acordo com a lei*, *conforme o decreto*, *esta lei entra em vigor na data de sua publicação*, etc.

Levando em conta esse conjunto de critérios, o ponto de partida para identificar as UFE eventivas foi gerar uma primeira lista, utilizando o Extraterm. Considerando esta lista, criamos uma Stopword contendo 1.974 palavras que não se caracterizavam como termo, tais como os artigos, pronomes demonstrativos, nomes próprios, siglas, etc. (*a*, *aos*, *e*, *essa*, *essas*, *ministério*, *lei*, *artigo*, *parágrafo*, *ABNT*, *IBAMA*, *IBGE*).

Aplicando essa *StopWord*, o programa gerou uma segunda lista de mais de 6.000 prováveis termos. A partir dela, selecionamos aquelas unidades léxicas que considerávamos como NE. O trabalho, a partir de então, foi o de

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

repassar uma a uma das unidades e selecioná-las, verificando cada um dos contextos para validar uma lista de candidatos a NE e/ou verbos com potencialidade para constituir uma unidade fraseológica. Excluímos nessa etapa verbos (*dever, poder, considerar, acordar*), substantivos (*bacia, bactéria, baía, beneficiárias*) e adjetivos (*benéfica, bom, cabível*).

Com esse primeiro procedimento obtivemos 613 candidatos a NE. Após esta etapa identificamos, a partir das concordâncias geradas para esses núcleos, aquelas unidades que continham um NT pertencente ao Direito Ambiental e não ao discurso jurídico como em geral, conforme já comentamos em relação aos critérios pragmático-discursivos. Esses critérios nos permitiram selecionar unidades como *contaminar as águas, emitir efluentes, degradar o meio ambiente, minimizar os impactos ambientais* e excluir outras tais como *acarretar multa aos infratores, cometer crimes*.

O passo seguinte foi aplicar o critério de frequência igual ou superior a cinco ocorrências, o que eliminaria muitas NE e, conseqüentemente, muitas unidades fraseológicas importantes dentro do âmbito (*contaminar o meio ambiente, contaminar as águas*). Por essa razão, consideramos que outras unidades com frequência menor que cinco deveriam ser selecionadas.

Essa decisão permitiu selecionar:

- a) Unidades que, embora com frequência mínima, possuem um valor especializado dentro do Direito Ambiental (*contaminar o meio ambiente*);
- b) Unidades que aparecem unidas, formando grandes grupos de UFE eventivas como:  
*produzir, comercializar, transportar, aplicar, prestar serviço, dar destinação a resíduos e embalagens vazias de agrotóxicos; uso e ocupação do solo.*

Tais unidades não teriam nenhuma ocorrência se fossem buscadas isoladamente (*produzir resíduos de agrotóxicos, comercializar resíduos de agrotóxicos, produzir embalagens de agrotóxicos, comercializar embalagens de agrotóxicos*) e, conseqüentemente, não teriam sido selecionadas. No entanto, considerando os critérios de pertinência temática e os pragmático-discursivos, consideramos que essas unidades representam ações importantes

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

dentro do Direito Ambiental, devido as suas implicações legais, e, portanto, devem ser recolhidas.

- c) Unidades que possuem um mesmo NE e um mesmo NT, mas que podem ter outros termos inseridos:

*uso do solo*

*uso do solo rural*

*uso de solos irrigáveis*

*uso racional do solo*

*uso saudável do solo.*

Essas diferentes possibilidades de variação, que implica a inserção de outras unidades lingüísticas (adjetivos, artigos, etc.) e a flexão de número (*solos*), denominamos de variação morfossintática. Para identificar essas diferentes realizações sempre partimos dos dois núcleo que formam as UFE eventivas, considerando, em primeiro lugar, o NE representado por um verbo ou um nome e, em segundo lugar, o NT. Salientamos, portanto, que o foco principal que orienta a seleção dessas unidades sempre serão seus núcleos. Esse foco permite manter uma estabilidade que está dada pelos aspectos semânticos, pragmáticos e discursivos, ao mesmo tempo em que permite refletir sua variação morfossintática:

*evitar dano*

*evitar dano irreparável / de difícil reparação*

*evitar danos*

*evitar danos à saúde/ a terceiros*

*proteger espécies ameaçadas*

*proteger as espécies ameaçadas*

- d) Unidades que podem ser consideradas sinônimas. Essa sinonímia ocorre em relação - ao NE:

*licença de pesca / autorização de pesca*

*autorizar o aproveitamento industrial das quedas d'água / conceder*

*o aproveitamento industrial das quedas d'água*

- ao NT:

*conservação da biodiversidade / conservação da diversidade biológica*

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

Obtivemos como resultado final dessa metodologia de extração e seleção um conjunto de 1155 UFE eventivas, que fazem parte de uma base de dados da fraseologia jurídico-ambiental (BDT-Frasjure).

Apresentamos como dados complementares ao processo realizado para a identificação das UFE eventivas do Direito Ambiental, os resultados referentes ao seu reconhecimento em um corpus em língua espanhola da área de Engenharia Computacional, utilizando o *Simple Concordance Program (SCP)*. Nessa área, observamos que as dificuldades encontradas não diferiram muito das já mencionadas.

Primeiramente, foi necessário gerar uma lista de palavras, utilizando a ferramenta *Wordlist*, o que resultou em uma lista de aproximadamente 3.200 palavras. A partir dessa lista, foi realizada uma seleção manual dos candidatos a termo e possíveis núcleos de UFEs, num total de aproximadamente 400 palavras-chave.

Foi gerada, então, uma lista das concordâncias de todas as palavras-chave. Esse material foi posteriormente submetido à leitura e seleção manual para identificar aquelas unidades que poderiam constituir UFE, ou seja, unidades que tivessem pelo menos um termo entre seus componentes e um núcleo verbal ou eventivo e que possuíssem um certo grau de fixação. Para esse levantamento, a quantidade de ocorrências de uma possível UFE não foi levada em consideração, uma vez que o objetivo do trabalho é a análise semântica (qualitativa) e não quantitativa dos dados obtidos.

Por tratar-se de um trabalho preliminar, a análise foi feita a partir de uma amostragem retirada do total de possíveis UFE encontradas na leitura da lista de concordâncias.

Essa amostragem foi selecionada a partir de cinco termos que apresentaram um maior número de ocorrências: *actividad*, *datos*, *modelo*, *programa* e *software*. Neste caso, foi utilizado o critério de quantidade para proporcionar uma maior possibilidade de combinações de um mesmo termo com diferentes verbos.

Analisando a amostragem selecionada, foi possível encontrar os termos escolhidos relacionados aos seguintes verbos, como mostram os exemplos abaixo para os termos *programa* e *software*:

*matar programa*  
*crear programa*  
*ejecutar / ejecución de programa*

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

*describir / descripción de programa*  
*estructurar / estructuración de programa*  
*ampliar programa*  
*abandonar programa*  
*construir programa*

*desarrollar / desarrollo de software*  
*mantener software*

A partir das relações estabelecidas na tabela acima, foi realizada uma nova busca por contextos, inicialmente utilizando a ferramenta Concordance, do SCP. Como os contextos selecionados pelo programa eram muitas vezes insuficientes, foi necessário realizar uma busca nos textos, utilizando a ferramenta Localizar, do MSWord.

Além das dificuldades de extração automática geradas pelo programa escolhido, há também o fato de que nem sempre o termo vem imediatamente após o núcleo eventivo, podendo haver elementos (artigos, demonstrativos, etc.) entre um e outro, conforme mencionamos anteriormente e demonstramos nos exemplos abaixo:

1. *Algunas publicaciones interesantes sobre el modelado difuso son: Zvieli y Chen, que **proponen un modelo** que soporta atributos difusos en las entidades y las interrelaciones; (...)*
2. *Con el objetivo de conocer la situación actual que presentan los sistemas de información y las tecnologías de información se realizará un análisis comparativo de los contextos mundial, nacional y regional, en lo relacionado a la empresa mediana y pequeña y en el sector comercio y servicios, para ello se **propone un modelo de conceptualización sistémica** para resolver el problema planteado, este se observa en la Figura.*

Nesses dois processos de extração e identificação de UFE eventivas, podemos observar que, embora partíssemos de um conjunto de critérios, que as caracterizavam e permitiam identificá-las, e dispuséssemos de programas



BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

que nos ajudassem na parte inicial de sua extração, encontramos ainda uma série de dificuldades relacionadas:

- a) a própria constituição complexa das UFE eventivas e seu uso, principalmente no que se refere a determinação do seu valor especializado e a sua variação morfossintática e semântica;
- b) aos programas de extração de informação - Extraterm iX e SCP-, pois sua extração prévia de candidatos a “termo” é de base estatística, fato que gera muito ruído e requer um trabalho de seleção manual considerável.

No entanto, gostaríamos de ressaltar que a pesquisa desenvolvida permitiu corroborar alguns aspectos que consideramos importantes em relação à caracterização e reconhecimento das UFE. Entre eles destacamos:

- sua conformação por dois pólos semânticos – NE e NT – que lhe conferem certa estabilidade dessas unidades, mantendo a sua diversidade morfossintática;
- a caracterização das UFE como unidades fraseológicas e especializadas determinadas pelas propriedades pragmático-discursivas dos textos em que são utilizadas;
- o critério de frequência, considerado sob essa perspectiva, é importante, mas não é determinante, posto que o peso maior está dado pelo valor especializado que a unidade adquire nos textos e, conseqüentemente, no âmbito em que são utilizadas.

Embora consideremos que tenhamos avançado no estabelecimento das propriedades das UFE eventivas que permitem reconhecê-las de forma mais adequada, acreditamos que ainda devemos desenvolver estudos que permitam:

- aprofundar a descrição das propriedades morfossintáticas, semânticas e pragmáticas das unidades fraseológicas especializadas que sirvam de subsídios para sua melhor caracterização e, conseqüentemente, de parâmetros para sua avaliação;
- estudar alternativas para aprimorar a extração (semi-)automática dessas unidades através de ferramentas de extração de informação especializada em corpora textuais.

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

#### NOTAS

<sup>1</sup> Esse recorte se deve ao trabalho realizado em conjunto com pesquisadores do Colégio de México e do Instituto Universitários de Lingüística Aplicada da Universidade Pompeu Fabra, que participam do projeto BD-Frasjure, cuja finalidade é recolher a fraseologia jurídico-ambiental em português do Brasil, espanhol da Espanha e do México e catalão. Este projeto tem apoio da Rede Ibero-americana de Terminologia (RiTerm).

#### REFERÊNCIAS BIBLIOGRÁFICAS

- BERBER SARDINHA, Tony. *Lingüística de corpus*. Barueri, São Paulo: Manole, 2004, cap. 1,2,3.
- BEVILACQUA, C.R. (2004). *Unidades Fraseológicas Especializadas Eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Barcelona: Instituto Universitário de Lingüística Aplicada, Universidade Pompeu Fabra. [Tese de Doutorado].
- (1999). *Unidades Fraseológicas Especializadas: estado de la cuestión y perspectivas*. Barcelona: Instituto Universitário de Lingüística Aplicada, Universidade Pompeu Fabra.
- (1996). *A fraseología jurídico-ambiental*. Porto Alegre: Universidade Federal do Rio Grande do Sul, Curso de Pós-Graduação em Letras. [Dissertação de Mestrado]
- CABRÉ, M. T. (2002) "Textos especializados y unidades de conocimiento: metodología y tipologización". In: García Palacios, Joaquín; Fuentes, M. Teresa (eds). *Texto, terminología y traducción*. Salamanca: Ediciones Almar, p. 15-36.
- CIAPUSCIO, G. (1998). "La Terminologia desde el punto de vista textual: selección, tratamiento y variación". *Organon*, Porto Alegre, v.12, n.26: 43-65.
- (2003) *Textos especializados y Terminología*. Barcelona: Instituto Universitário de Lingüística Aplicada, Universidade Pompeu Fabra.
- ESTOPÀ, Rosa.. (2001). "Elementos llingüístics de las unidades terminológicas para su extracción automática". In: CABRÉ, M.T.; FELIU, J. *La terminología científico-técnica*. Barcelona: IULA, p. 67-79.
- KRIEGER, Maria da Graça; FINATTO, M. J. B. (2004). *Introdução à Terminologia: teoria & Prática*. São Paulo: Contexto.

BEVILACQUA, C. R.; SALGADO, A. R.; SILVEIRA, D. A. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, ISSN 1806-275X, 2006.

L'HOMME, M-C. (2000) "Understanding Specialized Lexical Combinations".  
En: *Terminology*, Vol. 6, n. 1, p. 89-110.

LORENTE, M.; BEVILACQUA, C.R. (2000). "Los verbos en las aplicaciones terminográficas". *Terminología e industrias da língua. Actas do VIII Simpósio Ibero-americano de Terminologia*. Lisboa: ILTEC /Rede Ibero-americana de Terminologia/União Latina, p. 635-653

MEYER; I.; MACKINTOSH, K. (1996). "Refining the terminographer's concept-analysis methods: How can phraseology help?". *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication*, 3-1. Amsterdam: John Benjamins Publishing Company, p. 1-26.

SINCLAIR, J. (1992). *Corpus, Concordance, collocation*. Oxford: OUP.