

**Estudos  
Papers**



# O proxecto ‘Combinacións léxicas especializadas’ (Projecom)

## *Project ‘Specialized Lexical Combinations’(Projecom)*

Cleci Regina Bevilacqua

Universidade Federal do Rio Grande do Sul

cleci.bevilacqua@gmail.com

Resumo: Este artigo ten o obxectivo de presentar o actual proxecto que desenvolve o Projeto Terminológico Cone Sul (Termisul) (2010-2015), do Instituto de Letras da Universidade Federal do Rio Grande do Sul (UFRGS). Na súa vertente aplicada, o proxecto ten o obxectivo de construír unha base de datos de Combinacións Léxicas Especializadas (CLEs) da linguaxe legal, normativa e científica para tradutores e redactores especializados. A parte teórica analiza o proceso desa construción para propor a estrutura teórica que sustenta esa práctica. Preséntase o referente teórico, as etapas previstas, a metodoloxía aplicada e algúns dos resultados alcanzados ata o momento.

Palabras clave: terminoloxía; linguaxe xurídica; linguaxe científica; combinacións léxicas especializadas.

*Abstract: The purpose of this paper is to present the ongoing work developed by the Terminológico Cone Sul Project (Termisul) (2010-2015) at the Institute of Letters of the Federal University of Rio Grande do Sul (UFRGS), Brazil. As applied research, the project aims to build a database of Specialized Lexical Combinations (SLCs) of legal, normative and scientific language to assist specialized translators and text writers. The theoretical inquiry is concerned with the structuring of this building process in order to make up the theoretical framework that supports the practice. This paper presents the theoretical ground, the adopted methodology, further phases and some results achieved so far.*

Keywords: terminology; legal language; scientific language; specialized lexical combinations

## 1. Introducción

O proxecto de investigación que aquí presentamos constitúe unha etapa máis do percorrido de estudos e de formación en investigación do Termisul, que se fundou no último decenio do século XX. A investigación teórica e aplicada do grupo (2010-

---

Data de recepción 1.4.2013. Data de aceptación 23.7.2013.

2015) focaliza as expresións pluriverbais específicas dos textos legais, normativos e científicos<sup>1</sup>. O estudo concéntrase na identificación e na caracterización desas expresións, denominadas combinacións léxicas especializadas (CLEs)<sup>2</sup> en portugués brasileiro (PB) e dos seus equivalentes en alemán, español, francés, inglés e italiano. O seu obxectivo é reflexionar sobre a terminoloxía e a fraseoloxía, distinguindo os compostos léxicos especializados –específicos da linguaxe científica– de aqueles outros que teñen un carácter xeral e que aparecen en textos sen asociación directa con terminoloxías científicas. Esta reflexión busca xerar fundamentación teórica e metodolóxica para ofrecer produtos terminográficos *on-line* dirixidos principalmente a redactores técnicos e tradutores, así como propiciar avances significativos para o recoñecemento automático en aplicacións de Procesamento da Lingua Natural (PLN).

Coidamos que identificar as CLEs, explicar o seu significado, coñecer o seu comportamento na comunicación e suxerir alternativas para a súa descrición, categorización e tratamento son temas indispensables para o desenvolvemento dos estudos terminolóxicos. Ademais disto, cómpre coñecer as devanditas combinacións na lingua materna e nas linguas de traballo do tradutor ou redactor para producir textos correctos dende o punto de vista lingüístico e adecuados dende o punto de vista da convencionalidade da área especializada. Consecuentemente, a súa identificación e presentación en forma de base de datos, dispoñible *on-line*, é un dos obxectivos da investigación.

Resaltamos que, aínda que o destinatario que se busca son os tradutores e redactores, tamén poden aproveitar os resultados obtidos na investigación profesionais doutras áreas da lingüística e da informática, así como lexisladores, avogados, empresarios, profesionais da comunicación, entre outros, abranguendo varios sectores da sociedade. En Brasil, os estudos das CLEs, ou o que se coñece como Fraseoloxía Especializada, aínda son recentes e cómpre afondar máis neles. A carencia que se detecta no noso país pode constatarase tamén nos demais países do continente latino-americano. Nese sentido, como Brasil forma parte do Mercosur, pensamos que é importante incluír o estudo das CLEs en lingua española, dada a relevancia desa lingua na rexión e a necesidade dun estudo contrastivo co portugués, tal e como fixemos noutros traballos que elaborou o grupo. Seguindo a tradición plurilingüística do grupo, analizamos tamén as linguas alemá, francesa, inglesa e italiana, conforme dixemos máis arriba. Estas linguas proporcionan o desenvolvemento de subproxectos de investigación.

---

<sup>1</sup> Este proxecto conta coa participación das investigadoras: Anna Maria Becker Maciel, Cláudia Mendonça Scheeren, Cristiane Krause Kilian, Sandra Dias Loguércio e Patrícia Chittoni Ramos Reuillard. Conta tamén con bolseiros de iniciación científica, alumnado da carreira de Bacharelado em Letras – Tradução, e investigadores colaboradores.

<sup>2</sup> Optamos pola denominación Combinacións Léxicas Especializadas (Combinatórias Léxicas Especializadas) seguindo a proposta de L'Homme e Bertrand (2000) e L'Homme (2000). Adaptando esta proposta ao proxecto e tendo en conta tamén as ideas de Gouadec (2004), é posible abarcar varios tipos de combinacións léxicas, dende as que son sintagmas verbais e nominais deverbais ata frases. As devanditas combinacións tamén poden denominarse Unidades Fraseolóxicas Especializadas (Cabré, Lorente e Estopà, 1998) ou Fraseoloxía Terminolóxica (Tabares Plasencia e Pérez, 2007).

Pola súa complexidade, os puntos de maior interese do estudo concéntranse no propio concepto de fraseoloxía ou CLEs e na súa delimitación e, polo tanto, na súa identificación no tocante a outros tipos de combinacións. Nese sentido, a clara determinación do inicio e da fin dunha combinación depende tamén da definición ou da perspectiva teórica adoptada. Para os tradutores e redactores de textos especializados, en cambio, máis que trazar límites, interesa coñecer qué unidades lingüísticas ou recorrentes acompañan os termos de determinada área e qué combinacións son características dun dominio discursivo determinado.

Neste contexto, establécese a conexión dos estudos sobre as combinacións (no ámbito da Terminoloxía) con outras áreas do coñecemento (como a Tradución e a Linguística de Corpus). Deste modo, desexamos contribuír ao avance dos estudos terminolóxicos e ao desenvolvemento da produción e da tradución científico-técnica.

## 2. Marco referencial teórico

### 2.1. Propostas teórico-metodolóxicas

O tema das combinacións ou da fraseoloxía no ámbito da lingua xeral é bastante antigo e os estudosos preocupáronse por este dende hai moitos séculos. Dise que Santo Agostiño, no século IV, afirmaba que expresións como *in saeculum* debían tratarse conxuntamente a pesar da súa configuración pluriverbal. O seu estudo foi obxecto de investigacións lexicolóxicas e lexicográficas intensificadas dende a década de 1980 e profundadas grazas á contribución de lingüistas e lexicógrafos. En cambio, aínda se discute sobre a súa descrición; e a cuestión do seu tratamento en obras de referencia aínda non se resolveu de todo malia que sexan moitos os dicionarios de combinacións, tamén coñecidos como dicionarios de fraseoloxía, de frases feitas ou de expresións idiomáticas.

O desenvolvemento non foi igual no eido das linguaxes especializadas. Houbo que agardar ata a década de 1990 para que as combinacións ou fraseoloxías dos textos especializados comezasen a chamar a atención de lingüistas e terminólogos que, nun primeiro momento, consideraron tales asociacións sintagmáticas dende unha perspectiva lexicolóxica e lexicográfica. Apoiados na noción de *colocación* de Firth (1957) e na *análise distribucional* de Harris (1991), trataron as combinacións ou as UUFF que atopaban nos textos da ciencia e da tecnoloxía como *fraseoloxía* da lingua común, de acordo coas propostas de estudosos como Hausmann (1990) e Melčuk (1984, 1988, 1992). Consonte esta visión, as devanditas unidades denomínanse *colocacións* e compóñense basicamente de dous elementos, un deles considerado como *base* e o outro, o *colocado* ou elemento recorrente.

A base é un elemento semanticamente autónomo, mentres que o colocado está restrinxido semanticamente por esta. As unidades están semilexicalizadas, posúen un valor semicomposicional e o seu significado non pode deducirse de forma independente a partir de cada un dos elementos que a compoñen, pero si polo conxunto destes elementos. Por estas razóns, segundo os seguidores da proposta, tales unidades deben

incluirse nos dicionarios e os falantes que pretenden utilizalas deben memorizalas. Nesta perspectiva, e para axudar a recoñecelas, establécense estas estruturas morfosintácticas: substantivo + substantivo, substantivo + adxectivo, verbo + substantivo, adxectivo + adverbio e verbo + adverbio. Son moitos os autores que adoptan esta concepción para o estudo das CLEs, entre eles L'Homme e Bertrand (2000) e Tercedor Sánchez (1998).

Pero L'Homme e Bertrand advirten que, malia que as CLEs compartan algunhas similitudes coas colocacións (grupos de palabras usadas na lingua xeral), non se comportan exactamente da mesma maneira. En consecuencia, non deben describirse usando o mesmo método que habitualmente empregan os lexicógrafos<sup>3</sup> (L'Homme; Bertrand, 2000: 1).

Para estas autoras, o obxecto de estudo da fraseoloxía especializada é o conxunto das combinacións formadas por termos e polos seus colocados, entre os que se establece unha restrición de selección determinada polas especificidades dos ámbitos especializados. Como exemplos destas unidades podemos citar: *dose única* (medicina), *salvar um arquivo* (informática), *resíduos sólidos*, *proteger os recursos naturais* (ecoloxía). Como se pode observar, non hai unha distinción precisa entre o que é un termo sintagmático (*resíduos sólidos*) e o que é outro tipo de combinación especializada (*proteger os recursos naturais*). De feito, a fronteira entre o concepto de termo sintagmático e de combinación especializada é moi tenue, e establecer os seus límites é unha tarefa moitas veces incerta e discutible.

Pola súa banda, a proposta de Tercedor Sánchez (1998) céntrase no papel que desempeñan as diferentes situacións comunicativas na conformación das CLEs nos textos de cada área temática. Esta autora tamén salienta a aparición da variación das fraseoloxías. Ademais destes aspectos, afirma que na linguaxe biomédica, área estudada na súa tese, utilízanse moitas combinacións da lingua común.

Para a análise das unidades que recolle nesta área, Tercedor Sánchez baséase en dous modelos: o modelo lexemático-funcional –proposto por Martín Mignorance e seguido por Faber e Mairal (1999, 2002)– e o de Melčuk (1984, 1988, 1992), que xa amentamos anteriormente. O primeiro, formado polos eixes paradigmático, sintagmático e contrastivo, permite estruturar xerarquicamente as unidades léxicas, prever padróns de complementación para cada predicado e comparar a información léxica entre as diferentes linguas. Do segundo modelo, a autora utiliza os padróns morfosintácticos e as funcións léxicas para identificar as combinacións. A partir dos recortes teóricos que fai, clasifica as combinacións identificadas en tres grupos: a) unidades lexicalizadas terminolóxicas e da lingua xeral; b) UUFF non lexicalizadas; c) colocacións.

Outra perspectiva en relación coa concepción de fraseoloxía segue unha orientación máis terminolóxica e considera que as combinacións especializadas, denominadas Unidades Fraseolóxicas Especializadas (UFES), son unidades sintagmáticas, formadas

<sup>3</sup> No orixinal: “[...] specialized lexical combinations (word groups used in special languages) – even though they share some similarities with collocations (word groups used in general language) – do not behave exactly like them. Thus, they should not be described using the apparatuses lexicographers usually resort to”.

por dous ou máis elementos, especificando que un dos elementos é o termo e que o outro ten carácter verbal. Distínguese das perspectivas anteriores polas seguintes características: a) atribúelle un carácter verbal á unidade; b) non propón estruturas morfosintácticas para axudar a identificalas; c) non fai referencia á restrición entre os elementos que compoñen a unidade; d) establece claramente a diferenza entre termos e fraseoloxismos.

Entre os autores que adoptan esta concepción, podemos incluír a Lorente, Estopà e Bevilacqua (1998) e Bevilacqua (2004). Para estas autoras, é importante a distinción entre termos sintagmáticos e UFEs, tal e como presentamos arriba. Deste xeito, *resíduos sólidos* é claramente un termo, xa que posúe un referente específico na área ambiental, mentres que *tratar* ou *tratamento dos residuos* é unha UFE, xa que denota un proceso desta área.

Unha terceira perspectiva, distinta das anteriores, é a de Gouadec (1994). Este investigador prefere o termo *entidades fraseolóxicas*, no que inclúe un conxunto flexible de combinacións como palabras, grupos de palabras, termos, locucións, expresións, proposicións, sen a preocupación de establecer categorías nin criterios morfolóxicos. Para el o obxectivo da identificación e tratamento destas unidades é a xestión e dominio por parte do redactor e do tradutor. O autor indica catro características relevantes para o seu recoñecemento: a especialización, a repetición, o risco inherente á súa manipulación e a vantaxe de dominar o seu uso (Ibidem: 167-168).

Neste contexto Gouadec postula como criterios fundamentais para a identificación deste tipo de unidades a estereotipia, concibida como un conxunto de caracteres estabilizados e recorrentes nunha área, e a frecuencia coa que ocorren en tal área. Concibe ademais que estas unidades están formadas por elementos invariables e variables, indicando, deste modo, a posibilidade de alteracións no interior dunha mesma unidade, na que pode haber inserción e/ou supresión dun ou máis dun elemento ou alteración de orde. A partir destes criterios (estereotipia, frecuencia e conformación a través de invariables e variables), estes conxuntos de caracteres especializados asumen o carácter de matriz. A súa proposta é eminentemente funcional e relaciónase directamente coa ganancia de calidade e produción do texto no proceso tradutivo. Como exemplo de matriz temos *assumir um compromisso* entre [x] e [y], no que *assumir um compromisso* entre é a parte invariable, e [x] e [y], a parte variable. As variables poden completarse, no caso dun contrato de aluguer, por exemplo, con [*locador*] e [*locatário*].

Considerando as diferentes visións do tema e as diferentes abordaxes<sup>4</sup> suxeridas para tratalo que presentamos arriba podemos dicir que hai polo menos tres perspectivas teóricas relativas á fraseoloxía ou combinacións especializadas, cada unha coas súas especificidades: unha de perspectiva máis lexicolóxica, que se aproxima á lingua xeral;

---

<sup>4</sup> Hai varios autores que tratan a fraseoloxía especializada, entre eles Picht (1990), Blais (1993), Tabares Plasencia e Pérez (2007), entre outros. Pode verse unha revisión destes autores en Bevilacqua (2004) e Zilio (2009).

outra marcadamente terminolóxica, que enfatiza o carácter especializado; e outra de carácter pragmático, que se centra na práctica textual e tradutiva.

## 2.2. A nosa proposta

Do cadro referencial teórico máis amplo deste traballo destacamos o presuposto básico da Teoría Comunicativa da Terminoloxía (de agora en diante TCT) (Cabré 1999, 2001), que considera a linguaxe especializada como unha expresión da lingua común empregada en contextos de especialidade e que comparte todas as propiedades da lingua natural ao desempeñar as funcións comunicativas propias da área temática. Nesta perspectiva, recorreremos ao tratamento descritivo da Terminoloxía Textual (Bourigault; Slodzian 1999), que analiza a linguaxe especializada na súa efectiva realización en contexto de uso. Ao mesmo tempo, atendemos á Linguística de Corpus, que privilexia o estudo da lingua en grandes conxuntos de textos auténticos analizados por ferramentas dixitais (Berber Sardinha 2004).

O termo, na perspectiva da TCT (Cabré 1999), defínese como unha unidade léxica poliédrica, de carácter denominativo e conceptual, dotado de capacidade de referencia que denomina un núcleo ou nó na estrutura conceptual dun ámbito. Deste xeito, tanto *virus* –termo simple– como *águas subterráneas* –termo sintagmático– considéranse como termos, xa que posúen un referente específico na estrutura conceptual da medicina ou da informática, no primeiro caso, e do medio ambiente, no segundo. Os fraseoloxismos ou combinacións especializadas xa non posúen un referente específico. Son a suma de dúas ou máis unidades léxicas e denotan procesos, actividades ou accións propias dunha determinada área, tal como vemos nos exemplos *alijamento de residuos* e *entrar com um recurso*.

Considerando tales paradigmas da Terminoloxía e tendo como base os autores que presentamos no apartado anterior (2.1), definimos as CLEs como unidades sintagmáticas formadas por dous ou máis elementos que poden conter un termo. Estas unidades son recorrentes nas situacións de comunicación das áreas temáticas que revelan unha preferencia marcante polas especificidades e polas convencións propias do idioma, da área e/ou do tipo de texto no que aparecen. Son unidades semilexicalizadas pero non fixas por completo, que non se confunden coas terminoloxías, pero que caracterizan a especialización dun texto. Aínda que non compartan o carácter de termo na estrutura conceptual dun dominio do coñecemento, as CLEs son de importancia indiscutible na comunicación do saber porque denotan procesos, actividades e accións propias da área temática na que aparecen.

Con estas premisas foi posible establecer os seguintes criterios para a súa identificación:

- sintagmaticidade: deben estar formadas por máis dun elemento lingüístico;
- relativa estabilidade semántica e sintáctica: son fixas ou semifixas, é dicir, admiten pouca ou ningunha substitución dos elementos lingüísticos que as conforman, no eixo paradigmático, e inclusión de novos elementos no eixo sintagmático;



- prototipicidade: son específicas de determinadas áreas de coñecemento e caracterizan as súas formas de dicir;
- frecuencia relevante: deben ter unha frecuencia considerable nos ámbitos nos que se utilizan, o que leva a caracterizalas como prototípicas desa área. En cambio, é importante lembrar que este criterio pode flexibilizarse en determinado corpus; unha CLE que ten unha frecuencia baixa pode ser importante na área e non pode excluírse unicamente pola aplicación deste criterio;
- consensualidade: os membros da área especializada establecen o seu uso, o que lles confire un carácter convencionalizado e dun ámbito específico. Deste xeito, no corpus de lexislación ambiental, atopamos *alijamento de residuos*, pero non *exclusão de residuos*; *revogar uma lei* pero non *retirar uma lei*.

Ademais de definir os criterios para a súa selección, ao traballar con diversas linguas estranxeiras é importante tamén ter presente e fundamentarse nunha definición de equivalencia. Deste modo, adoptamos a definición de equivalencia funcional que defendeu Gémar (1998). Segundo este autor, son equivalentes as estruturas que expresan a mesma relación semántica e o mesmo efecto pragmático nos textos de partida e de chegada. Isto significa que cómpre buscar os equivalentes en lingua estranxeira en textos auténticos e nos seus contextos e analizalos co fin de establecer se cumpren a mesma función do texto de partida. Para dirimir dúbidas é primordial a consulta a especialistas.

### 3. Córpora textuais e etapas metodolóxicas

As CLEs extraeranse das bases textuais que xa construíu o Grupo Termisul ([http://www.ufrgs.br/termisul/bases\\_textuais/apresentacao.php](http://www.ufrgs.br/termisul/bases_textuais/apresentacao.php)), tendo como punto de partida os cörpera en portugués brasileiro:

- *Base Legis*: conxunto de textos de lexislación ambiental de Alemaña, Arxentina, Brasil, Estados Unidos, Francia, Italia, Paraguai e Uruguai. A lexislación de cada un dos países constitúe un corpus separado con preto de 600.000 palabras cada un. Ademais dos textos da lexislación, esta base inclúe outros textos que se utilizan como apoio na busca de contextos ou de equivalentes para as linguas estranxeiras. Estes textos son: actos internacionais (*Agenda 21*, *Declaração do Rio*, *Protocolo de Kyoto*, en alemán, español, francés, inglés, italiano e portugués, e *Convenção de Estocolmo* en español, francés, italiano, inglés e portugués), e textos aliñados<sup>5</sup> (*Declaração do Rio* e *Protocolo de Kyoto* en inglés/alemán, inglés/español, inglés/francés, inglés/italiano, inglés/portugués e *Convenção de Estocolmo* en inglés/español e inglés/portugués).

---

<sup>5</sup> Textos aliñados: conxunto de dous ou máis textos, un orixinal e a(s) súa(s) tradución(s), con segmentos equivalentes que se dispoñen paralelamente.

- *Base Gestamb*: corpus composto por textos extraídos de manuais, ademais de teses e disertacións da área de xestión ambiental, en portugués de Brasil<sup>6</sup>. Contén preto de oito millóns de palabras.

Cómpre resaltar que as bases Legis e Gestamb están constituídas por textos que o grupo empregou na elaboración dos tres dicionarios: *Dicionário de Direito Ambiental: Terminologia das leis do meio ambiente* (1998 e 2008), *Glossário Multilíngüe Direito Ambiental Internacional: Terminologia dos Tratados* (2004) e *Dicionário de Gestão Ambiental* (2007). Para as primeiras obras, seleccionáronse e utilizáronse textos da Legislação Ambiental Brasileira vixente, abarcando textos dende 1934 ata 2008; tratados internacionais bi- e multi-laterais promulgados por Brasil; ademais da Constitución e os códigos xurídicos vixentes no país. A partir destes textos seleccionáronse os termos incluídos nos dicionarios. Pola súa banda, o corpus da base Gestamb está formado por fragmentos de manuais que tratan a posta en funcionamento dos procedementos de certificación da ISO 9000, así como por disertacións de mestrado e teses de doutoramento seleccionadas en páxinas web de bibliotecas universitarias entre 2005 e 2009 a partir da palabra clave *gestão ambiental*. Por seren de dominio público, as teses e disertacións descargáronse integramente.

- *Base Tecno-Ciência*: textos científicos da área de cardioloxía (portugués, alemán, español, francés, inglés) baixados das revistas de referencia na área e dispoñibles en internet, con extensión de 600.000 palabras en cada unha das linguas.

O traballo de recollida das CLES en cada unha das bases mencionadas constitúe unha etapa diferente do proxecto, é dicir, o proxecto prevé tres grandes etapas, unha para cada unha das bases mencionadas arriba. Para cada unha destas etapas prevíronse os seguintes procedementos metodolóxicos<sup>7</sup>:

- revisión xeral dos *córpora* reunidos e sistematización das súas principais características estruturais nos diferentes idiomas<sup>8</sup>;
- extracción de candidatos a CLEs con axuda das ferramentas do Termisul –listador de palabras<sup>9</sup>, concordanciador<sup>10</sup>, xerador de *n-gramas*<sup>11</sup>, xerador de *clusters*<sup>12</sup>–, e

<sup>6</sup> Esta base só posúe textos en portugués e os seus equivalentes buscáronse noutras fontes, principalmente de internet, seguindo criterios de busca establecidos polo equipo.

<sup>7</sup> Para máis detalles sobre a metodoloxía de extracción das CLEs, véxase Bevilacqua, C.R.; Maciel, A.M.B.; Scheeren, C. M.; Reuillard, P.C.R. 2012 e Bevilacqua, C.R.; Maciel, A.M.B.; Scheeren, C.M.; Reuillard, P.C.R.; Kilian, C. K 2013.

<sup>8</sup> A descrición detallada de cada un dos *córpora* pode verse en [[http://www.ufrgs.br/termisul/bases\\_textuais/apresentacao.php](http://www.ufrgs.br/termisul/bases_textuais/apresentacao.php)].

<sup>9</sup> Lista todas as palabras do corpus por orde alfabética ou de frecuencia.

<sup>10</sup> Lista as aparicións no texto dunha determinada palabra co seu contexto inmediato.

<sup>11</sup> Secuencias de palabras que se repiten no corpus; poden ter diversas extensións dependendo das opcións que ofrezan as ferramentas. O xerador de *n-gramas* do Termisul permite xerar *n-gramas* de extensión 1 a 8 e coa opción de amosar a palabra clave nas diferentes posicións posibles na expresión (dereita, esquerda, centro ou en calquera delas). Tamén se pode escoller ou non o corte de frecuencia. Para máis detalles desta ferramenta, véxase [[http://www6.ufrgs.br/termisul/limbo/ferramentas\\_novo/Seleciona.php](http://www6.ufrgs.br/termisul/limbo/ferramentas_novo/Seleciona.php)].

<sup>12</sup> Agrupamento de palabras ao redor dunha palabra clave.

tamén as dos programas AntConc (Anthony 2008) e Multiword Expression Toolkit (Ramisch 2012);

- aplicación dos criterios para a selección das CLEs (ver 2.2) nas listas de *n-gramas* xeradas polas ferramentas;

- identificación dos equivalentes nas linguas estranxeiras, utilizando os córpora da Base LEGIS e GESTAMB e, cando for necesario, empregando fontes de confianza de internet<sup>13</sup>.

Ademais da identificación das CLEs en portugués e da busca dos seus equivalentes nas linguas estranxeiras para a Base Legis (primeira etapa), tamén se definiron os campos da base de combinacións, conforme describimos no apartado 4.2.

## 4. Resultados iniciais

### 4.1. Clasificación das CLEs da Base Legis

Na etapa actual da investigación estamos traballando coa recollida das CLEs da Base Legis<sup>14</sup>. Recollemos un conxunto significativo de combinacións e, despois de analazalas con detalle considerando os seus aspectos sintácticos, semánticos e pragmáticos (Bevilacqua 2004; Kjaer 2007), foi posible clasificalas en dous grupos:

#### 4.1.1. CLEs xurídicas

Estas CLEs empregáanse en textos legais de calquera área de dereito e poden ter diferentes funcións:

a) indicar accións específicas do poder lexislativo como emitir normas, atribuír poderes e competencias, prescribir comportamentos, etc. Ex.: *Revogam-se as disposições em contrário.*

b) indicar remisión interna (parte do mesmo texto) ou remisión intertextual (outro documento legal). Ex.: *de acordo com o disposto em...*

*Ressalvado o disposto nas Leis números 5.357, de 17 de novembro de 1967, e 7.661, de 16 de maio de 1988, a receita proveniente da aplicação desta Lei será recolhida de acordo com o disposto no ART.4 da Lei número 7.735, de 22 de fevereiro de 1989. (LEI N º 6.938, DE 31 DE AGOSTO DE 1981).*

c) establecer restricións e/ou delimitacións. Ex.: *Para os fins de...*

*Para os fins deste Decreto, entende-se por: I - aquíicultura: o cultivo de organismos que tenham na água o seu normal ou mais freqüente meio de vida; [...] (DECRETO Nº 2.869, DE 9 DE DEZEMBRO DE 1998).*

<sup>13</sup> Para realizar buscas en páxinas fiables en internet seguimos os parámetros propostos por Marins e Krieger (2005).

<sup>14</sup> Proxecto realizado co apoio de recursos no edital 18/2012 do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Ademais destas funcións, observamos que cada unha das CLEs sitúase en xeral en partes específicas da macroestrutura textual, o que permite identificar a súa función pragmática. Por exemplo, a CLE *Esta lei entra en vigor na data de súa publicación* aparece sempre nos artigos finais e ten a función de fechar a lei e indicar que terá vixencia a partir do momento en que for promulgada.

Outra característica deste tipo de CLEs é a súa estrutura matricial, é dicir, posúe unha parte fixa e outra variable:

- a) a parte fixa é estable e non sofre substitución, aínda que poida presentar variacións sintácticas: *para efectos de* [texto lexislativo ou parte de texto lexislativo] e *para os efectos de* [texto lexislativo ou parte de texto lexislativo];
- b) admite variantes (sinónimos): *de acordo com a lei* ten como expresión sinonímica *nos termos da lei*;
- c) a parte variable é a unidade terminolóxica específica do ámbito do Dereito: *para os efectos do Artigo 14, para os efectos da Lei nº 6.938/81*, etc.

#### 4.1.2. CLEs terminolóxicas

Estas CLEs posúen as seguintes propiedades:

- a) caracterizan a área do Dereito Ambiental, área interdisciplinar;
- b) refírense ás accións e/ou procesos específicos do medio ambiente, pero están circunscritas polo Dereito –a tutela;
- c) están formadas por verbos (*controlar, preservar*) ou nomes deverbais (*controle, preservação*) e un termo (simple ou sintagmático: *água; recursos naturais*). Vexamos un exemplo cos seus equivalentes nas linguas estranxeiras:

PT: *conservação dos recursos naturais*  
DE: *Schonung der natürlichen Ressourcen*  
EN: *natural resources conservation*  
ES: *conservación de los recursos naturales*  
FR: *préservation des ressources naturelles*  
IT: *conservazione delle risorse naturali*

#### 4.2. Arquitectura da base de datos

Despois de recoller as CLEs lexislativas e comezar a identificación das CLEs terminolóxicas, definimos os campos e a arquitectura da base de datos, considerando o usuario e as funcións previstas<sup>15</sup>. Así, a base terá dous módulos, un de traballo, que utilizarán os investigadores para inserir os datos, e outro de consulta, destinado aos usuarios e con diferentes recursos de busca. O módulo de consulta está en construción.

<sup>15</sup> Para a definición da arquitectura da base consultáronse algunhas bases xa existentes como IATE [<http://iate.europa.eu/iatediff/>], Termium [<http://www.btb.termiumpplus.gc.ca/>] e LexAlp [<http://lexalpapps.eurac.edu:8080/termbank/ViewDictionaries.po?SeeVolume=1000024>].

A ficha do módulo de traballo está en fase final de revisión e está constituída polos campos:

- CLE: campo para a inclusión da combinación;
- opción de escolla da lingua: permite a escolla da lingua (portugués, alemán, español, francés, inglés e italiano) a partir da que se vai completar a ficha;
- contexto: espazo para a inclusión do contexto de aparición da combinación coa indicación da súa fonte;
- outras formas: formas que presentan a inserción dalgún elemento lingüístico en relación coa CLE principal ou algunha alteración morfosintáctica. Por exemplo: *de acordo com o disposto na lei*, no que a forma *disposto* se incluíu na CLE principal *de acordo com a lei*;
- contextos doutras formas: permite a inserción de contextos de aparición de combinacións consideradas como outras formas da CLE principal coa indicación da súa fonte;
- véxase tamén: remisión das CLEs que posúen algún tipo de relación semántica coa CLE principal. Por exemplo: *de acordo com* ten remisións ás súas formas sinónimas *nos termos de, conforme a, em conformidade com*;
- equivalente nas distintas linguas: remite á ficha completa da CLE na lingua estranxeira seleccionada;
- nota: posibilita informar o usuario dalgún uso peculiar da CLE, como a súa función pragmática no texto lexislativo, ou algún aspecto particular nunha das linguas estranxeiras;
- comentarios: espazo para que os investigadores poidan indicar dúbidas, correccións e revisións que haxa que facer, visible só no módulo de traballo.

Deseguido presentamos unha ficha do módulo de traballo completa:

**Ficha de Entrada**

CLE:

Língua:

Contexto:

Seleção autoforma:

Outra forma:

Exemplo:

Seleção ver também:

Seleção equivalente de:

Seleção equivalente em:

Seleção equivalente em os seguintes de [texto legislativo ou parte de texto legislativo] Bem como:

Seleção equivalente 1:

Seleção equivalente 2:

Similite esta ficha como problemática

Comentários:

Selec os dados da CLE Associe esta CLE:

Outros assuntos:

Ficha de traballo para rexistro das CLEs

## 5. Consideracións finais

Aínda que establecemos un conxunto de criterios para a selección das combinacións, aínda hai algunhas cuestións pendentes que están en estudo. Entre estas podemos citar as CLEs formadas por varios verbos ou nomes deverbais. Por exemplo, a CLE *produção, distribuição e aplicação de agrotóxicos e afins*. Unha das solucións é analizar a frecuencia para identificar se esta estrutura é frecuente nos textos lexislativos en portugués e se ten que inserirse desa forma na base. No caso de que non sexa frecuente, haberá que dividila en tres combinacións, cada unha coa súa propia ficha: *produção de agrotóxicos e afins*, *distribuição de agrotóxicos e afins* e *aplicação de agrotóxicos e afins*.

Para resolvermos problemas coma este e mellorarmos a extracción de CLEs a partir de córpora, iniciamos a proba de extracción das CLEs coa ferramenta MultiWord Expressions Toolkit (MWE Toolkit), conforme comentamos na sección 3. Para empregar esta ferramenta, etiquetamos o corpus en portugués co TreeTagger<sup>16</sup>, que etiqueta textos

<sup>16</sup> Ferramenta desenvolvida por Schmid (1994). Dispoñible en [<http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>].

automaticamente. Despois da revisión da etiquetaxe, definimos os padróns de busca para identificar as CLEs terminolóxicas. Algúns dos padróns son: V+N; V+Det+N; N+Prep+N; N+Prep+Det+N. Estes padróns permiten extraer CLEs como: *comprar energia, fiscalizar os residuos; compra de energia, fiscalização dos residuos sólidos*.

Na etapa actual da investigación estamos analizando os primeiros resultados obtidos co MWE Toolkit e os contrastes co corpus de referencia formado por textos do xornal *Folha de São Paulo*. Estes procedementos permitirán identificar as CLEs prototípicas dos textos lexislativos en lingua portuguesa con base en padróns estatísticos e lingüísticos, aos que asociaremos os criterios mencionados no apartado 2.2, asegurando así unha maior confiabilidade no recoñecemento das combinacións especializadas.

Ademais desta etapa de extracción, a continuidade da investigación prevé para os próximos dous anos (2013-2015) as seguintes tarefas:

- conclusión da selección das CLEs terminolóxicas da base Legis en lingua portuguesa e nas linguas estranxeiras;
- inserción do conxunto das CLEs na base de combinacións;
- revisión do módulo de traballo da base;
- conclusión do módulo de consulta da base;
- publicitación da base.

Posteriormente, darémoslle continuidade á identificación das combinacións das bases Gestamb e Cardiologia, seguindo os mesmos principios e metodoloxía que presentamos aquí. Cos resultados obtidos esperamos contribuír ao avance dos estudos de fraseoloxía especializada, ofrecendo subsidios para a reflexión teórica sobre o tema, ademais dunha base de consulta de acceso libre.

## 6. Referencias bibliográficas

- ANTHONY, Lawrence (2008) AntConc (3.2.1 w) Tokyo: Waseda University. [<http://www.antlab.sci.waseda.ac.jp/software.html>]. [Data de consulta: 28.04.2011].
- BERBER SARDINHA, Toni (2004): *Lingüística de Corpus*. Barueri: Manole.
- BEVILACQUA, Cleci Regina (2004): *Unidades Fraseolóxicas Especializadas Eventivas: descripción y reglas de formación en el ámbito de la energía solar*. Tese de doutoramento. [[http://www.ufrgs.br/termisul/biblioteca/teses/tese\\_DOUTORADO\\_2004\\_BEVILACQUA.pdf](http://www.ufrgs.br/termisul/biblioteca/teses/tese_DOUTORADO_2004_BEVILACQUA.pdf)]. [Data de consulta: 30.05.2010].
- BEVILACQUA, Cleci Regina; MACIEL, Anna Maria Becker; SCHEEREN, Cláudia Mendonça e REUILLARD, Patrícia Chittoni Ramos (2012): "Combinacións Léxicas Especializadas: etapas previas para identificación e tratamento" en *Actas del XII Simposio Iberoamericano de Terminología*, v. 2. Buenos Aires: Colegio de Traductores Públicos de la Ciudad de Buenos Aires; 272-283.
- BEVILACQUA, Cleci Regina; MACIEL, Anna Maria Becker; SCHEEREN, Cláudia Mendonça; REUILLARD, Patrícia Chittoni Ramos e KILIAN, Cristiane Krause

- (2013): “Combinacións Léxicas da Linguagem Legislativa: uma abordagem orientada pelo corpus” en MURAKAWA, Clotilde de Almeida Azevedo e NADIN, Odair Luiz. (Eds.): *Terminologia: uma ciência interdisciplinar*. Araraquara: FCL-UNESP Laboratório Editorial/ Cultura Acadêmica Editora; 227-244.
- BLAIS, Ester (1993): “La phraséologie. Une hypothèse de travail” en *Terminologies Nouvelles*, 10. Bruselas: RINT, 1993 ; 50-56.
- BOURIGAULT, Didier e SLODZIAN, Monique (2000): “Pour une terminologie textuelle” en *Terminologies Nouvelles*, n.19, 2000 ; 29-32.
- CABRÉ, María Teresa (1999): *Terminología: representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- (2001): “Sumario de principios que configuran la nueva propuesta teórica y consecuencias metodológicas” en CABRÉ, María Teresa e FELIU, Judit (Ed.): *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada; 17-25.
- CABRÉ, María Teresa; LORENTE, Mercé e ESTOPÀ, Rosa (1996): “Terminología y fraseología” en *Actas del V Simposio de Terminología Iberoamericana*. Ciudad de México: Colegio de México; 67-81.
- FABER, Pamela e MAIRAL USÓN, Ricardo (2002): “Reconsidering lexical representations in Functional Grammar” en MAIRAL, Ricardo e PÉREZ QUINTERO, María Jesús (Eds): *New Perspectives on Argument Structure in Functional Grammar*. Berlin: Mouton de Gruyter; 39-94.
- (1999): *Constructing a Lexicon of English Verbs*. Berlín y Nueva York: Mouton de Gruyter.
- FIRTH, John Rupert (1969) “Modes of meaning” en FIRTH, John Rupert *Papers in linguistics 1934-1951*. London: Oxford University Press; 190-215.
- GÉMAR, Jean-Claude (1998) : *Les enjeux de la traduction juridique. Principes et nuances*. [<http://www.tradulex.com/Bern1998/Gemar.pdf>]. [Data de consulta: 04.05.2010].
- HARRIS, Zelig (1991): *Theory of language and information: a mathematical approach*. Oxford: Oxford University Press.
- GOUADEC, Daniel (1994): “Nature et traitement des entités phraséologiques” en *Terminologie et phraséologie. Acteurs et aménageurs: Actes de la deuxième Université d'Automne en Terminologie*. Paris: La Maison du Dictionnaire; 167-193.
- HAUSMANN, Franz Josef (1990): “Le dictionnaire de collocations” en HAUSMANN, Franz Josef et alii. *An International encyclopedia of lexicography*. Vol. 1. Berlin, New York: Walter de Gruyter; 1010-1019.
- KJAER, Ane Lise (2007): “Phrasemes in legal texts” en BURGER, Harald (eds.): *Phraseology: An International Handbook of Contemporary Research*. vol. 2. Walter de Gruyter & Co.:Berlin, Germany; 506-616.



- KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker; ROCHA, João Carlos de Carvalho; BEVILACQUA, Cleci Regina e FINATTO, Maria José Bocorny (1998): *Dicionário de Direito Ambiental. Terminologia das leis do meio ambiente*. 1ª. ed. Porto Alegre: Editora da Universidade.
- KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker; FINATTO, Maria José Bocorny e REUILLARD, Patrícia Chittoni Ramos (2004): *Glossário multilíngüe de Direito Ambiental Internacional: a terminologia dos tratados*. Rio de Janeiro: Forense.
- KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker; BEVILACQUA, Cleci Regina; FINATTO, Maria José Bocorny e REUILLARD, Patrícia Chittoni Ramos (2007): *Glossário de Gestão Ambiental*. São Paulo: Disal Editora.
- KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker; BEVILACQUA, Cleci Regina e FINATTO, Maria José Bocorny (2008): *Dicionário de Direito Ambiental. Terminologia das Leis do Meio Ambiente*. 2ª. ed. Rio de Janeiro: Editora Lexikon.
- L'HOMME, Marie-Claude (2000): "Understanding specialized lexical combinations" en *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*. Vol. 6, n.1. Amsterdam: John Benjamins.
- L'HOMME, Marie-Claude e BERTRAND, Claudine (2000): *Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions?* [<http://www.ling.umontreal.ca/lhomme/docs/berlhom.pdf>]. [Data de consulta 15.03.2010].
- LORENTE, Mercè; BEVILACQUA, Cleci Regina e ESTOPÀ, Rosa (2002): "El análisis de la fraseología especializada mediante elementos de la lingüística actual" en CORREA, Margarita (Ed.): *Atas VI Simpósio da Rede Ibero-americana de Terminologia. Terminología, desenvolvimento e identidade nacional*. Lisboa: Ed. Colibri/ILTEC; 647-666.
- MARINS, Diego Sttiger e KRIEGER, Maria da Graça (2005): "Busca e valoração de equivalentes em espanhol para termos jurídicos-ambientais em sites da internet" en *Anais do VIII Congresso Internacional da Associação Brasileira de Estudos Canadenses (ABECAN)*. Gramado: ABECAN. (Edición electrónica)
- MEL'CUK, Igor (1984, 1988, 1992): *Dictionnaire Explicatif et combinatoire du français contemporain*. Vol. 1, 2, 3, Montréal: Les Presses de la Université de Montréal.
- PICHT, Heribert. (1990): "LSP phraseology from the terminological point of view" en *Terminology science & research: Journal of International Institute for Terminology Research*, vol. 1, n. 1-2, 1990. Vienna: International Network for Terminology; 33-48.
- RAMISCH, Carlos (2012): *A generic and open framework for multiword expressions treatment: from acquisition to applications*. Tese de doutoramento. [[http://www.inf.ufrgs.br/~ceramisch/download\\_files/thesis/](http://www.inf.ufrgs.br/~ceramisch/download_files/thesis/)]. [Data de consulta: 30.08.2012]

- TABARES PLASENCIA, Encarnación e PÉREZ VIGARAY, Juan Manuel (2007): “Fraseoloxía terminolóxica: estado de la cuestión y exemplo de análise contrastivo” en *Revista de Filoloxía de la Universidade de La Laguna* (RFULL) 25, 2007; 565-575.
- TERCEDOR SÁNCHEZ, María Isabel (1998): *La fraseoloxía en el lenguaje biomédico: análise desde las necesidades del traductor*. Tese de doutoramento. [<http://elies.rediris.es/elies6/>]. [Data de Consulta: 03.06.2010]
- ZILIO, Leonardo (2009), *Colocações especializadas e 'Komposita': um estudo contrastivo alemão-português na área de cardiologia*. Disertación de máster. Porto Alegre: Programa de Pós-Graduação em Letras.