

## Grupo Termisul: do projeto acervo ao estudo de combinatórias léxicas

Cleci Regina Bevilacqua\*  
Maria José Bocorny Finatto\*\*  
Patrícia Chittoni Ramos Reuillard\*\*\*

**RESUMO:** Este texto apresenta um pouco da história do Grupo de pesquisa TERMISUL e relata alguns resultados da sua mais recente iniciativa de estudos sobre padrões da linguagem especializada, transcorrida entre 2006 e 2009. A nova etapa é caracterizada principalmente pela pesquisa baseada em *corpora* e pelo oferecimento de dados obtidos apenas em versão *on-line* através do *site* [www.ufrgs.br/termisul](http://www.ufrgs.br/termisul). O formato de apresentação *on-line* alia-se à temática do ensino de Terminologia, o qual se relaciona fortemente com a formação em Tradução. São descritos e avaliados os projetos desenvolvidos nesse período: *Acervo TERMISUL: padrões da linguagem legal, normativa e científica* e apresentados os subprojetos a ele vinculados, assim como alguns estudos conexos. Por fim, é caracterizado o perfil geral do novo trabalho do Grupo, que vigorará a partir de 2010 e que se centrará na pesquisa sobre combinatórias léxicas especializadas.

**Palavras-chave:** Terminologia; recursos *on-line*; padrões da linguagem especializada.

**ABSTRACT:** This paper presents a little bit of the history of the TERMISUL research group and reports some results of its latest initiative of studies on specialized languages patterns between 2006 and 2009. The new stage of the group is characterized mainly by corpus-based researches and by on-line presentation of results at the website [www.ufrgs.br/termisul](http://www.ufrgs.br/termisul). This new presentation format is related to the theme of terminology teaching, which is strongly associated to translation training. The main research project developed during this period, called *Collection TERMISUL: patterns of legal, regulatory and scientific languages*, and related sub-projects are described, as well some related studies. Finally, we characterize

---

\* Universidade Federal do Rio Grande do Sul – e-mail: [cleci-bevilacqua@terra.com.br](mailto:cleci-bevilacqua@terra.com.br).

\*\* Universidade Federal do Rio Grande do Sul – e-mail: [mfinatto@terra.com.br](mailto:mfinatto@terra.com.br).

\*\*\* Universidade Federal do Rio Grande do Sul – e-mail: [patriciaram@terra.com.br](mailto:patriciaram@terra.com.br).

As autoras agradecem ao apoio dos seguintes fomentadores: CNPq, FAPERGS e PROPESQ-UFRGS.

the general profile of the new research, which runs from 2010 and will focus on specialized multiword expressions.

**Keywords:** Terminology; on-line resources; standards of specialized languages.

## 1. Introdução

**N**o âmbito dos estudos terminológicos feitos à luz dos pressupostos teóricos de perspectiva comunicativa, textual e sócio-cognitiva, o Grupo TERMISUL, da Universidade Federal do Rio Grande do Sul – situada no Sul do Brasil –, tem prosseguido sua trajetória de investigações sobre a comunicação especializada em um percurso que se iniciou em 1991.

Os objetivos do Grupo, desde sua fundação, permanecem centrados em contribuir para a promoção dos estudos de Terminologia, fomentando a formação em Iniciação Científica, Mestrado e Doutorado. Desses objetivos, temos, como consequência, o estímulo para o planejamento e a produção de recursos terminográficos de diferentes perfis e de temáticas variadas, a partir do que temos cultivado uma inter-relação muito forte com a tradução técnico-científica. Essa inter-relação tornou-se algo natural à medida que, desde os primeiros anos do Grupo, temos atuado em meio a um Instituto de Letras que abriga um curso de graduação em Tradução, além de cursos de graduação para a formação em ensino de língua portuguesa e línguas estrangeiras.

Nesse cenário e a partir do legado do seu histórico de trabalhos, marcado pela publicação de diversos glossários e dicionários e pela orientação de várias dissertações e teses, além da produção de todo um conjunto de reflexões sobre diferentes fazeres de pesquisa, podemos dizer que o grupo TERMISUL assumiu, em meio à nossa comunidade, uma posição de referência nas áreas de Lexicologia, Lexicografia e Terminologia, ainda que o ênfase tenha sido e ainda seja, naturalmente, para a Terminologia.

Hoje nosso Grupo pode ser definido como uma soma de experiências e de aprendizagens que é fruto de uma longa trajetória. Uma trajetória que, importa destacar, já completou 19 anos e que teve início no início dos anos 90 graças às iniciativas pioneiras das nossas queridas colegas Maria de Graça Krieger e Anna Maria Becker Maciel. Essas fundadoras do TERMISUL foram e ainda são os principais alicerces para a nossa formação enquanto pesquisadoras e modelos para a nossa atuação como novas orientadoras de uma novíssima geração de estudantes e de pós-graduandos que continuamente têm se aproximado do Grupo.

Assim, sem poder deixar de relacionar-se com toda uma trajetória pregressa, este texto apresenta especificamente uma etapa recente das pesquisas do Grupo TERMISUL, transcorrida entre 2006 e 2009, caracterizada principalmente pela produção de recursos a partir de *corpora* e pelo desenvolvimento de um sistema de recuperação de dados obtidos em pesquisa no formato *on-line*. A apresentação de diferentes resultados de estudos relacionados entre si agora apenas na forma *on-line* e com acesso gratuito, conforme gostamos de pensar, transformou o *site* do Grupo TERMISUL em algo que pode ser aproximado a um ambiente virtual de aprendizagem para o usuário interessado em estudos do léxico especializado. A seguir, é descrito e avaliado o principal projeto desenvolvido nesse período, denominado *Acervo TERMISUL: padrões da linguagem legal, normativa e científica*<sup>1</sup>, como também são apresentados os subprojetos a ele vinculados, assim como alguns estudos conexos. Por fim, é apresentado o perfil geral do novo trabalho do grupo, que vigorará a partir de 2010.

## 2. Sobre o projeto acervo *Termisul*

Delineamos aqui, sinteticamente, a proposta do Projeto *Acervo TERMISUL: padrões da linguagem legal, normativa e científica* (doravante referido apenas como *Acervo TERMISUL*) e caracterizamos sucintamente o embasamento que nos serviu como orientação teórica e metodológica.

### 2.1 Proposta da Pesquisa

A proposta do Projeto *Acervo TERMISUL* previu a concepção e a publicação de uma nova apresentação para o *website* TERMISUL ([www.ufrgs.br/termisul](http://www.ufrgs.br/termisul)). A idéia básica foi agregar informações sobre usos, características e práticas textuais, em diferentes idiomas, das linguagens legal, normativa e técnico-científica<sup>2</sup>, de modo que o usuário encontrasse informação qualificada e confiável sobre essas diferentes linguagens, além de instruí-lo sobre passos básicos da pesquisa terminológica. Assim, o *site* oferece: a) amostras de textos

---

<sup>1</sup> Projeto de pesquisa contemplado no Edital de Ciências Humanas e Sociais do CNPq em 2006.

<sup>2</sup> Além das autoras deste texto, o *ACERVO TERMISUL* conta com a participação das pesquisadoras Anna Maria Becker Maciel e Cláudia Mendonça Scheeren, de bolsistas de Iniciação Científica (BIC/PIBIC) e de pesquisadores-colaboradores.

para uma percepção de um *modus dicendi* associado a cada um dos domínios tratados; b) ferramentas *on-line* para a exploração pontual desses textos; c) bibliografia e materiais técnicos produzidos para situar o usuário em meio às questões da pesquisa terminológica teórica e prática.

Ao final de 2009, o *site* já oferece, gratuitamente, toda uma série de recursos para diferentes usuários, entre os quais destacamos pesquisadores de linguagem, tradutores, terminólogos, redatores técnicos e profissionais de áreas afins, além de estudantes de graduação e de pós-graduação. Nosso usuário mais constante, entretanto, tornou-se o professor, o pós-graduando e o aprendiz de tradução, principalmente os estudantes das disciplinas de Terminologia I e II que integram o currículo do curso de graduação em Tradução.

Entre os recursos já implementados no *site*, destacamos:

- a) Bases textuais: Gestamb, Legis e Tecno-Ciência – com *corpora* de dados advindos de manuais, teses e dissertações na temática da Gestão ambiental;
- b) Ferramentas informáticas para extração de informação textual desses *corpora*: geradores de contextos mediante expressões de busca, listador de palavras, gerador de n-gramas e gerador de dados em contextos alinhados para os casos de textos originais e suas respectivas traduções;
- c) Biblioteca virtual: glossários, mapas conceituais, tutoriais para formação continuada em Terminologia, textos acadêmicos e apresentações de trabalhos em eventos.

Considerando que esse projeto dedicou-se ao estudo dos fenômenos da comunicação legal, normativa e técnico-científica, buscamos, por um lado, observar os diferentes usos da linguagem em situações comunicativas especializadas e, por outro, oferecer aos usuários recursos que lhes possibilitassem buscar informações de forma autônoma, de acordo com suas necessidades específicas de pesquisa. Por exemplo, um tradutor poderá identificar as combinatórias específicas de determinado termo e área de especialidade a partir das listas de contextos proporcionadas pela ferramenta *concordanciador*; um terminólogo, servindo-se do mesmo recurso, pode recolher contextos de uso de determinado termo com vistas a estabelecer sua definição e um estudante de pós-graduação pode ter contato com toda uma produção de pesquisas através da nossa Biblioteca Virtual. Nessa mesma Biblioteca, destacamos também o oferecimento de mapas conceituais, os quais visam situar o

usuário sobre a abrangência e as inter-relações de um dado recorte de domínio associado a uma dada Base Textual, o que é feito pela apresentação de todo um conjunto de hierarquias e conexões entre temas e termos.

## 2.2 Orientação Teórica

Os referenciais teóricos que guiaram a concepção do trabalho fundamentam-se em estudos de Terminologia de perspectiva comunicativa e textual (Hoffmann, 1998; Ciapuscio, 2003; Krieger & Finatto, 2004), na Terminologia Cognitiva (Temmermann, 2000), na Terminologia de perspectiva comunicativa (Cabré, 1999) e nos Estudos da Linguagem dedicados ao objeto texto, sobretudo os que privilegiam elementos da textualidade do texto acadêmico-científico (Beaugrande, 1980; Beaugrande & Dressler, 1981; Koch, 2001; Halliday & Hasan, 1976; Swales, 1990).

No que tange aos mapas conceituais<sup>3</sup>, são tomados como referência estudos específicos sobre a expressão de relações semânticas e conceituais em linguagens técnico-científicas (Cabré, Morel & Tebé, 1998; Sager & Kageura, 1998; Sager, 1990). Também são considerados os pressupostos teóricos da Linguística de *Corpus* (Sinclair, 1991; Berber Sardinha, 2004), que reza que, a partir do reconhecimento da configuração lexical de acervos previamente preparados e organizados, é possível detectar evidências sobre a configuração textual e a padronização léxico-gramatical da linguagem científica.

Não se trata, pois, de mera observação de terminologias mais ou menos marcadas em relação à linguagem cotidiana, não especializada, e o seu consequente registro isolado em dicionários. Muito antes disso, o que procuramos observar foi toda uma ambiência do texto especializado, consi-

---

<sup>3</sup> O desenvolvimento desse segmento do nosso trabalho esteve sob coordenação da Profa. Anna Maria Becker Maciel. A pesquisa em mapas conceituais é especialmente desenvolvida junto ao *Institute for Human and Machine Cognition* (IHMC). É um instituto de pesquisa sem fins lucrativos junto ao sistema de universidades da Flórida, Estados Unidos, que oferece uma série de ferramentas para a elaboração desses mapas e indicações para seu aproveitamento em situação de ensino e de pesquisa sobre estruturação do conhecimento. O IHMC oferece um programa que auxilia a elaboração desses mapas, ver <http://cmap.ihmc.us>. Na UFRGS, utilizando esses recursos, temos o Projeto “Mapas conceituais na Educação”, <http://mapasconceituais.cap.ufrgs.br>. Em Terminologia, mapas conceituais são denominados “árvores de domínio” ou “árvores de conceitos” e servem para ilustrar a organização de uma determinada área de conhecimento cuja terminologia se queira repertoriar.

derado em sua totalidade e condições de produção, levando em conta suas tipologias e estruturação, bem como sua constituição lexical e gramatical mais ampla, o que nos dá toda uma moldura de condicionamentos para a terminologia presente.

Acreditamos que a imbricação dessas diferentes concepções teóricas proporciona uma visão mais ampla do texto especializado, que leva em conta as suas diferentes facetas e, em consequência, permite chegar a resultados mais precisos de acordo com as necessidades de pesquisa. A seguir, apresentamos um maior detalhamento sobre os recursos já disponíveis no *site* TERMISUL.

### 2.3 Recursos disponíveis

O primeiro recurso que oferecemos ao nosso usuário é justamente um *corpus* textual. Esse *corpus* foi criteriosamente organizado de modo a espelhar uma dada linguagem. A dimensão das amostras de texto, acessíveis no seu todo ou em partes, está, entretanto, condicionada por limites de direito autoral, quando aplicável, daí porque o segmento dedicado à Legislação é o mais amplo. As Bases Textuais servem para apresentar e caracterizar os perfis e as práticas textuais das linguagens jurídica, normativa e técnico-científica. Assim, nossos *corpora* foram organizados em três blocos de Bases Textuais, a seguir caracterizadas.

#### a) Base Legis

É constituída por textos de:

- **legislação ambiental** (Alemanha, Argentina, Brasil, Estados Unidos, França, Paraguai e Uruguai), da legislação brasileira (*Código Civil, Código de Defesa do Consumidor, Código Penal*);
- **Constituições** (Alemanha, Angola, Argentina, Brasil, Cabo Verde, Espanha, Estados Unidos, França, Guiné Bissau, Moçambique, Paraguai, Portugal, São Tomé e Príncipe, Timor-Leste, Uruguai);
- **Atos Internacionais** (Agenda 21, Declaração do Rio e Protocolo de Kyoto, em alemão, espanhol, francês, inglês e português; Convenção de Estocolmo em espanhol, francês, inglês e português), **inclui textos alinhados**<sup>4</sup> (Declaração do Rio e Protocolo de Kyoto em inglês/alemão, inglês/espanhol, inglês/francês,

<sup>4</sup> Textos alinhados: conjunto de dois textos, um original e sua tradução, cujos segmentos equivalentes são dispostos paralelamente.

inglês/português, e Convenção de Estocolmo em inglês/espanhol e inglês/português).

A Base Legis, assim como as outras duas bases, contém um Catálogo com códigos específicos para país, tema, língua e numeração, além de informações sobre o texto, chamadas de anotações, que incluem registro no catálogo, título e extensão do texto, bem como informações sobre sua macroestrutura (cabeçalho, corpo do texto, anexo, entre outras).

No quadro 1 abaixo, temos uma pequena amostra do trabalho de codificação e organização da informação para o usuário, salientando-se que são documentos legais providos de diferentes países, escritos em diferentes línguas e com estatutos jurídicos distintos.

Quadro 1 – BASE LEGIS – Amostra da codificação da informação para Leis e Decretos

<b>Código</b>	<b>Título</b>	<b>Ementa</b>
arMA001 <b>(Argentina)</b>	Decreto Nro 107/2007	Biocombustibles - Actividades alcanzadas por los términos de la Ley 26.093.
brMA005 <b>(Brasil)</b>	LEI N° 2.419, de 10 de fevereiro de 1955	Institui a Patrulha Costeira e dá outras providências.
frMA001 <b>(França)</b>	Décret n° 2006-629 du 30 mai 2006	Relatif à la déclaration de projet et modifiant le code de l'environnement.
uyMA002 <b>(Uruguai)</b>	Decreto 249/000	Créase una Comisión de Evaluación de Riesgo de Vegetales Genéticamente Modificados integrada por especialistas de los Organismos que se determinan.

Cada um desses textos legais exibe também informações sobre a sua estrutura e extensão, as quais estão agregadas sob a forma de *tags* e de cabeçalhos conforme se pode ver pelo exemplo abaixo, que é apenas um trecho de uma lei brasileira. Essas marcações podem ser úteis para o processamento da informação apenas por segmentos, de modo que o usuário poderá selecionar visualizar apenas decretos ou apenas leis ou apenas textos de um determinado tamanho em termos de número de palavras, entre outras opções.

```

<head>
<name> brMA048 </name>
<title> LEI Nº 10.650, DE 16 DE ABRIL DE 2003. </title>
<summary> Dispõe sobre o acesso público aos dados e informações
existentes nos órgãos e entidades integrantes do Sisnama. </summary>
<ntoken> 756 </ntoken>
<ntype> 345 </ntype>
</head>
<body> O PRESIDENTE DA REPÚBLICA Faço saber que o Congresso
Nacional decreta e eu sanciono [...]
Art. 10. Esta Lei entra em vigor quarenta e cinco dias após a data de sua
publicação.
</body>
<date>Brasília, 16 de abril de 2003; 182º da Independência e 115º da
República. </date>

```

Por fim, fornecem-se instruções (*passo a passo*) para auxiliar o usuário, indicando como construir um *corpus* individual a partir dos textos disponíveis nessa base e explorá-la servindo-se das ferramentas oferecidas.

#### **b) Base Tecno-Ciência**

É integrada por textos científicos das áreas de Química, Informática e Medicina, oferecendo exclusivamente a consulta para trechos de artigos de periódicos especializados mediante a consulta por expressões de busca ou por listas de palavras. São oferecidos dados sobre textos de Cardiologia (português, alemão, espanhol, francês e inglês), Pediatria (português e inglês), Enfermagem (português, espanhol e inglês), Química (português, espanhol) e Informática (português e inglês). Há também a intenção de oferecer dados sobre neologia tradutória partir dos textos completos dos Seminários de Lacan conforme resultados da tese de doutoramento de Reuillard (2007). A consulta à base de textos é feita através de ferramentas de busca por palavras ou expressões.

Nessa base, os textos também foram codificados e, como envolvem direito autoral, são oferecidas apenas amostras de texto e informações sobre seus padrões macro e microestruturais. O material disponível nesse segmento do *Acervo TERMISUL* traz uma integração com recursos do Projeto TEXTQUIM/TEXTECC ([www.ufrgs.br/textecc](http://www.ufrgs.br/textecc)). Esse projeto, que é uma



derivação do TERMISUL, ocupa-se de caracterizar textos especializados, do ponto de vista lexical e discursivo mediante a utilização de metodologias e de princípios da Linguística de *Corpus* (Berber Sardinha, 2004), com especial destaque para as construções recorrentes, sejam discursivas, colocacionais, idiomáticas ou terminológicas. Essa parceria, conforme acreditamos, tem ampliado o espectro de informações do nosso usuário sobre a configuração das linguagens especializadas em suas realizações textuais.

### c) Base Gestamb

Base composta por textos da área de Gestão Ambiental, tais como manuais e obras de referência, a partir dos quais serão dadas informações sobre tópicos relativos à caracterização da linguagem da área. À Base Gestamb estará relacionada uma base de dados de combinatórias léxicas especializadas da área de Gestão Ambiental em línguas portuguesa e espanhola<sup>5</sup>. Essa base conterà as combinatórias léxicas coletadas a partir dos termos contidos no Glossário de Gestão Ambiental elaborado pelo Grupo TERMISUL<sup>6</sup>. Como exemplos já coletados, apresentamos as combinatórias abaixo, onde os negritos indicam os termos do glossário:

*tomada de **ações corretivas e preventivas***  
*operacionalização da **ACV** (*Análise do Ciclo de Vida*)*  
*consumo de **agrotóxicos***  
*controle de **agrotóxicos***  
*uso de **fertilizantes** e de **agrotóxicos***  
*alocação de **água***

O segundo recurso do Acervo TERMISUL compreende um conjunto de ferramentas *on-line* que permitem a exploração das bases textuais antes citadas. Possibilitam a extração de informação textual através da coleta de palavras ou expressões, em diferentes línguas, nos textos disponibilizados. Essas ferramentas são de quatro tipos:

---

<sup>5</sup> Identificação e Descrição das Combinatórias Léxicas Especializadas da Gestão Ambiental em Língua Portuguesa e em Língua Espanhola, projeto coordenado pela Profa. Cleci Regina Bevilacqua. O referido projeto tem apoio da UFRGS e foi contemplado no Edital de Ciências Humanas e Sociais de 2007 do CNPq.

<sup>6</sup> Krieger, Maria da Graça; Maciel, Ana Becker Maciel; Bevilacqua, Cleci Regina; Finatto, Maria José Bocorny; Reuillard, Patrícia Chittoni Ramos. 2006. *Glossário de Gestão Ambiental*. São Paulo: Disal.

**i) Concordanciador:** lista as ocorrências de uma determinada palavra com seu contexto imediato. Se o elemento de busca for seguido de asterisco\*\*, o resultado enfocará o radical. Por exemplo, *amb\** gerará uma lista das ocorrências de *ambiente*, *ambiental*, ***ambientais***, com duas ou mais palavras no seu entorno, como vemos nos exemplos abaixo:

*as preocupações relativas a meio **ambiente** e desenvolvimento e a elas  
As interações entre as políticas **ambientais** e as questões comerciais são  
e para uma melhor proteção **ambiental**. Um meio ambiente saudável, por*

Vale o inverso para uma busca com sufixos precedidos por asterisco, por exemplo, *\*çãõ*. O resultado será uma lista de contextos com todas as palavras terminadas por esse sufixo, precedidas e seguidas por dois ou mais itens lexicais:

*encontra em um momento de definição **çãõ** histórica.  
Defrontamo-nos com a perpetua**çãõ** definição histórica.  
Defrontamo-nos com a perpetua**çãõ** das disparidades existentes entre as*

**ii) Listador de palavras:** lista individualmente todas as palavras presentes no texto e/ou no *corpus* por ordem de frequência. Abaixo, as 10 palavras mais frequentes do documento *Agenda 21*:

*as=>2055  
das=>1910  
desenvolvimento=>1900  
que=>1809  
do=>1730  
com=>1482  
devem=>1025  
recursos=>978  
no=>962  
meio=>922*

**iii) Gerador de n-gramas:** lista as ocorrências de **grupos de palavras repetidos** ao longo de um texto ou *corpus*, com indicação de sua frequência. A extensão dos grupos pode ser previamente escolhida, ou seja, pode-se optar por bigramas, trigramas e quadrigramas.

Como indica o prefixo, um bigrama é um conjunto composto por duas palavras. Um trigrama, por três palavras, e assim por diante. Nos exemplos abaixo, de bigramas, o número que segue a combinação indica sua frequência no texto selecionado:

*meio\_ambiente => 598*

*desenvolvimento\_sustentável => 343*

**iv) Concordanciador alinhado: busca uma palavra em textos alinhados – envolvendo originais e respectivas traduções ou duas traduções para um mesmo original – e permite visualizar os respectivos contextos. São oferecidos materiais alinhados em português, alemão, espanhol e francês. A busca pode ser feita pela palavra em inglês ou na outra língua. Abaixo um exemplo do Protocolo de Kyoto, no sentido português/inglês e inglês/português para a palavra **emissão**:**

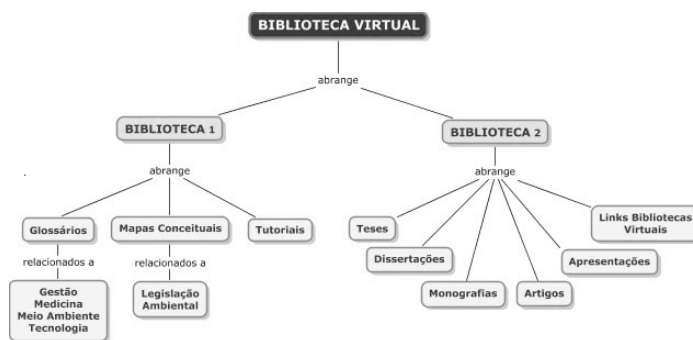
*ARTIGO 4*

*O respectivo nível de **emissão** determinado para cada uma das Partes do acordo deve ser nele especificado.*

*ARTICLE 4*

*The respective **emission** level allocated to each of the Parties to the agreement shall be set out in that agreement.*

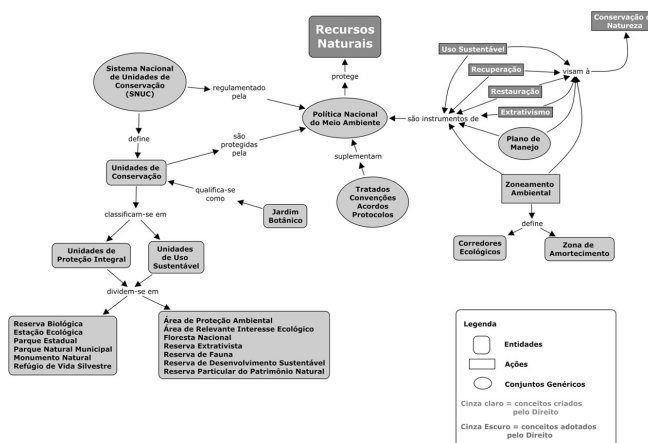
Um outro recurso importante no Projeto *Acervo TERMISUL* envolveu a organização e seleção de informações que possibilitassem o desenvolvimento e a formação em Terminologia. A ideia também foi a de criar uma maneira para que o usuário do *site* pudesse integrar, teórica e metodologicamente, as experiências que já teria tido nas Bases Textuais com as Ferramentas oferecidas. Esse recurso, que denominamos **Biblioteca virtual**, compreende uma Biblioteca 1 e a Biblioteca 2, como se pode ver a seguir:



A **Biblioteca 1** inclui:

a) **Glossários**: constituem uma pequena amostragem de trabalhos terminográficos realizados nas disciplinas de Terminologia I e II no Curso de Bacharelado – Tradução, da nossa universidade; são exemplos parciais<sup>7</sup>.

b) **Mapas conceituais**<sup>8</sup>: representações gráficas em forma de diagrama em que os termos referentes aos conceitos ocupam polígonos ou círculos ligados por vetores bidirecionais que identificam seus inter-relacionamentos. Seu uso para fins didáticos e para a indexação de conteúdos e recuperação da informação fundamenta-se no princípio de que nada é apreendido pela mente humana se não for ancorado em um conhecimento anteriormente adquirido. Abaixo o exemplo do mapa conceitual das Unidades de Conservação:



<sup>7</sup> Essas amostras podem ser vistas em <http://www6.ufrgs.br/termisul/glossarios.php>.

<sup>8</sup> A elaboração dos mapas conceituais é desenvolvida pelo Subprojeto *Interface: Terminologia/Mapas Conceituais/Legislação Ambiental*, coordenado por Anna Maria Becker Maciel.

**c) Tutoriais:** orientações para o uso de diferentes ferramentas informáticas:

- *Microsoft Access*: elaboração de bancos de dados terminológicos;
- *Corpógrafo*: ferramenta gratuita de armazenamento de *corpora*, de extração de informação textual (lista de frequência de palavras, gerador de n-gramas, extração de termos e combinatórias) e elaboração de bancos de dados terminológicos<sup>9</sup>;
- *Multiterm' 95*: elaboração de bancos de dados terminológicos;
- *WordSmith Tools 3.0*: extração de informações linguísticas (lista de frequência de palavras, gerador de n-gramas; concordância);
- *Simple Concordance Program*: extração de informação textual (lista de frequência de palavras, concordâncias).
- *AntConc*: extração de informação textual (lista de frequência de palavras, concordâncias).

Por sua vez, a **Biblioteca 2** inclui: textos acadêmicos, tais como monografias, comunicações, artigos, dissertações e teses de pesquisadores e bolsistas do Projeto, bem como de mestrandos e doutorandos da linha de pesquisa *Lexicografia e Terminologia: Relações Textuais* do nosso Programa de Pós-Graduação em Letras da UFRGS.

Além desses recursos, está previsto um canal de discussão e debate sobre questões relativas à Terminologia e à Terminografia e aos recursos disponibilizados. A continuidade do *ACERVO* até julho de 2010 ainda prevê:

- a) elaboração de tutorial para realização e registro de observação de textos paralelos;
- b) Conclusão da base de combinatórias léxicas especializadas da Gestão Ambiental.

### 3. Subprojetos relacionados e estudos afins

Como o Projeto *ACERVO* envolveu trabalho em diferentes frentes, com recursos, idiomas, tipos textuais e terminologias diferentes, a ele se relacionaram investigações desenvolvidas em paralelo e que envolveram várias participações, principalmente de estudantes de graduação na modalidade Iniciação Científica. Esse desdobramento envolveu subprojetos de pesquisa e

---

<sup>9</sup> Acessível em [www.linguateca.pt/corpografo](http://www.linguateca.pt/corpografo)

também a temática de projetos em Educação a Distância e Formação Continuada. Os projetos envolvidos foram os seguintes:

Pesquisas com participação de bolsistas de Iniciação Científica:

- 1 - *Acervo Termisul: Catálogo de Construções Recorrentes em Cardiologia alemão-português* – coordenação de Maria José B. Finatto
- 2 - *Equivalentes tradutórios para os neologismos lacanianos* – coordenação de Patrícia Chittoni Ramos Reuillard;
- 3 - *Identificação e descrição das combinatórias léxicas especializadas da Gestão Ambiental em língua portuguesa e em língua espanhola* – coordenação de Cleci Regina Bevilacqua;
- 4 - *Interface: Terminologia / Mapas Conceituais / Legislação Ambiental* – coordenação de Anna Maria Becker Maciel
- 5 - *Projeto Textquim Fase Dois: Expressões de Causalidade no Texto de Química: Coesão, Terminologias e Enunciação Científica* – coordenação de Maria José B. Finatto
- 6 - *Metáforas em Obras de Direito Ambiental* – coordenação de Maity Guerreiro

Propostas de Educação a Distância e Formação Continuada:

- 1 - *Educação Continuada no Acervo Termisul: Iniciação à Pesquisa Terminológica* – coordenação de Cleci Regina Bevilacqua, Patrícia Reuillard e Anna Maria Becker Maciel
- 2 - *ACERVO TERMISUL: um objeto de aprendizagem dos padrões léxico-gramaticais e textuais da linguagem legal, normativa e científica* – coordenação de Cleci Regina Bevilacqua, Patrícia Reuillard e Anna Maria Becker Maciel
- 3 - *Catálogo de Construções Recorrentes em Pediatria em Francês e em Italiano* – coordenação de Maria José B. Finatto
- 4 - *Catálogo de Construções Recorrentes em Textos de Pediatria para Aprendizagem de Tradução* – coordenação de Maria José B. Finatto

Dado que foram vários os projetos e estudos relacionados, em função do espaço de que dispomos aqui, apresentaremos breves informações sobre apenas três deles.

### **Equivalentes tradutórios para os neologismos lacanianos**

Pesquisa e proposição de equivalentes em língua portuguesa para 298 neologismos lacanianos já elencados e classificados a partir de um *corpus* formado pelos vinte e cinco Seminários. Para tanto, o projeto aplica uma proposta de princípios de neologia tradutória lacianiana, buscando a validação dos equivalentes propostos junto a especialistas em psicanálise lacianiana. Concomitantemente, o projeto tem a intenção de aprofundar o estudo das práticas textuais da psicanálise de orientação lacianiana, compartilhando essas reflexões através de artigos, da participação em eventos e da disponibilização de uma amostra dos resultados no *site* do Acervo Termisul: <http://www.ufrgs.br/termisul/acervotermisul>. Este projeto se insere na base Tecno-Ciência.

### **Projeto Textquim Fase Dois: Expressões de Causalidade no Texto de Química: Coesão, Terminologias e Enunciação Científica**

Geração de caracterizações linguístico-textuais para apoiar a necessária crítica do material bibliográfico geral e instrucional de Química, de modo a colaborar para o aperfeiçoamento das suas condições de redação, o que naturalmente origina subsídios para a crítica das condições de representação e de construção de conhecimentos na matéria. Pesquisa vinculada à pesquisa sobre linguagens, leitura e sobre o ensino de ciências. Objetivos: descrever e analisar, do ponto de vista linguístico-terminológico e enunciativo, formulações causais presentes em textos científicos de Química, especialmente manuais acadêmicos utilizados durante a formação de Licenciandos em Química, visando contribuir tanto para os Estudos de Texto (Teorias Linguísticas dedicadas ao Texto) quanto para os estudos de Educação Química, pesquisas sobre problemas do ensino e aprendizagem dessa ciência. Também envolve a geração de glossários e de outros repertórios disponíveis em [www.ufrgs.br/textquim](http://www.ufrgs.br/textquim), além de recursos textuais para o Projeto ACERVO TERMISUL.

### **Identificação e descrição das combinatórias léxicas especializadas da Gestão Ambiental em língua portuguesa e em língua espanhola**

Identificação e descrição as Combinatória Léxicas Especializadas (CLEs) do âmbito da Gestão Ambiental nas línguas portuguesa e espanhola. Baseados em L'Homme (2000 e 2007), entende-se as CLE como combinatórias de dois elementos ou mais, sendo um deles um termo e o outro o seu co-ocorrente. Entre estes elementos ocorre uma restrição de seleção determinada pelas especificidades do âmbito em que são utilizadas. Como objetivos

específicos derivados do geral propomos: a) a identificação das CLE do âmbito da Gestão Ambiental em língua portuguesa, partindo dos termos que fazem parte do Glossário de Gestão Ambiental e de critérios previamente estabelecidos para sua caracterização e reconhecimento; b) a identificação dos padrões morfossintáticos das unidades recolhidas a fim de repertoriar as tipologias morfossintáticas mais frequentes de constituição dessas unidades em língua portuguesa e de poder compará-las com os padrões encontrados em língua espanhola; c) a descrição das propriedades semânticas e pragmáticas de tais unidades e de suas relações com as especificidades da área em questão, tais como a situação comunicativa própria desse tipo de texto, as funções que cumprem as CLE nos textos em que são utilizadas; d) a identificação e validação dos equivalentes em língua espanhola; g) a descrição morfossintática e semântica das CLE em língua espanhola; e) a constituição de uma base de registro das unidades coletadas em língua portuguesa e seus respectivos equivalentes em língua espanhola que será disponibilizada no site do Acervo Termisul (<http://www6.ufrgs.br/termisul/baseGestambCle.php>).

Os projetos envolvidos ou relacionados com o *Projeto Acervo*, no que tange às terminologias, buscam reconhecê-las em função das suas relações no texto e com o texto, oferecendo a diferentes usuários recursos que permitam uma compreensão mais ampla, aprofundada e sustentada da organização e do funcionamento do texto especializado. Pretende-se, assim, estabelecer e mostrar a interface entre a terminologia e os textos em que são utilizadas e, mais que isso, proporcionar os recursos necessários para o conhecimento dos *modus dicendi* dos textos que conformam as bases textuais.

#### 4. Grupo termisul: do presente ao futuro

Atualmente, um dos principais interesses da pesquisa teórica e aplicada desenvolvida pelo Grupo TERMISUL é a observação das expressões multivocabulares recorrentes nas linguagens especializadas, tais como *expectativa de vida limitada*, *administrar um medicamento*, *gerenciamento de resíduos*, *revogam-se as disposições em contrário*. Seu estudo centra-se em questões referentes à sua caracterização e identificação, permitindo, por um lado, o oferecimento de produtos terminográficos que contenham tais expressões e estejam dirigidos principalmente a tradutores, redatores de textos técnicos, professores de língua e, por outro, proporcionando significativos avanços no seu reconhecimento automático em aplicações de Processamento da Língua Natural.

Denominadas combinatórias léxicas especializadas (CLEs), também conhecidas como fraseologias ou unidades sintagmáticas, são colocações que



não podem ser explicadas unicamente por exigências sintáticas, estilísticas e afinidades semânticas porque resultam de uma seleção restritiva determinada pelas especificidades e pelas convenções próprias do idioma, da área e/ou do tipo de texto em que ocorrem. Essas restrições relacionadas a questões gramaticais, semânticas, pragmáticas e temáticas estão condicionadas a especificidades dos modos de dizer característicos do uso de cada âmbito do saber. Formadas basicamente por dois elementos, um dos quais é considerado a base e o outro, o colocado ou o elemento co-ocorrente, são unidades semilexicalizadas que, na área temática, não têm um referente específico. Embora sejam distintas do termo, definido como uma unidade léxica de caráter denominativo e conceitual, dotado de capacidade de referência que denomina um núcleo ou nó na estrutura conceitual de um âmbito do conhecimento, as combinatórias léxicas especializadas são de importância indiscutível na comunicação do saber. Na comunicação especializada, as combinatórias desempenham importante papel seja porque denotam processos, atividades ou ações próprias de uma dada área, seja porque contribuem para a estruturação textual característica de um âmbito do conhecimento ou da atividade humanas. Além dessas combinatórias, também nos deparamos com aquelas combinatórias de caráter discursivo-textual, para as quais prevemos um estudo em paralelo, em parceria com o Projeto TeXTQUIM/TEXTECC antes referido.

Pela sua complexidade, os pontos de maior interesse da pesquisa estão voltados para o próprio conceito de fraseologia ou CLEs e para sua delimitação. Nem sempre é claro onde começa e onde termina uma combinatória, e definir seus limites depende também da definição ou perspectiva adotada para essas combinatórias. No entanto, para os tradutores e redatores de textos especializados, mais do que estabelecer limites, interessa conhecer que unidades linguísticas ou co-ocorrentes acompanham os termos de determinada área. Saber, por exemplo, que para o termo *água* há co-ocorrentes como *reuso* e *reaproveitamento*, que indicam processos no âmbito de Gestão Ambiental, ou ainda que se usa *preservação* ou *proteção do meio ambiente* e não *cuidado com o meio ambiente*, é fundamental para produzir textos corretos do ponto de vista linguístico e adequados do ponto de vista da área em questão. Oferecer a descrição e dados sobre essas combinatórias é proporcionar conhecimento sobre as formas de dizer específicas de determinada área.

A pesquisa em andamento se apoia na reflexão teórica sobre o tema, registrada em dissertações, teses, artigos e comunicações dos pesquisadores do Grupo Termisul e disponibilizados no seu *site*. Esse conjunto de trabalhos acadêmicos, caracterizado pela multiplicidade de áreas, línguas e tipos e gêne-

ros textuais, traz elementos para a discussão dos fatores determinantes das escolhas lexicais e permite levantar hipóteses sobre o comportamento das combinatórias nos diferentes idiomas, domínios e textos.

Assim, partindo do material armazenado no ambiente virtual do Acervo TERMISUL, iniciaram-se a identificação, a coleta e a descrição das combinatórias especializadas em língua portuguesa e em língua estrangeira (alemão, espanhol, francês e inglês) em *corpora* comparáveis na área de gestão ambiental e cardiologia e em *corpora* paralelos alinhados em tratados internacionais do meio ambiente.

#### 4.1 Projeto-Piloto para a Pesquisa de 2010

Descreveremos a seguir os materiais utilizados e a metodologia adotada para o projeto-piloto, que abrange, até o presente momento, os *corpora* dos Atos Internacionais do Meio Ambiente – Agenda 21, Declaração de Estocolmo, Declaração do Rio e Protocolo de Kyoto – em línguas portuguesa, espanhola e inglesa.

Obedecemos às etapas abaixo, seguidas separadamente para cada uma das línguas.

A primeira parte da pesquisa foi a busca de combinatórias nas três línguas, iniciando pelo português, considerado, nesta pesquisa, a língua de partida. O procedimento constou de:

- 1) constituição de um *corpus* unificado para cada uma das línguas dos quatro Atos Internacionais (português e espanhol com cerca de 200 mil palavras cada; inglês, 180 mil);
- 2) extração dos candidatos a combinatórias nas três línguas com a ferramenta ExtraCom<sup>10</sup> do ACERVO TERMISUL. Esta etapa implicou os passos a seguir:
  - a) construção de *stoplist* para eliminar itens lexicais não relevantes para a presente pesquisa, tais como palavras funcionais (a, por, em, porque, porém, the, a, and; y, la, el etc);
  - b) geração de n-gramas (com tamanho 3, 4 e 5) com corte de frequência 5.
- 3) estabelecimento e aplicação dos critérios seguintes para a seleção das combinatórias nas três línguas:

---

<sup>10</sup> Baseada no sistema *web* e desenvolvida pelos bolsistas SEAD Gabriel M. F. V. Pereira e Alessandro Dalla Vecchia.

- a) combinação de um termo e de dois ou mais co-ocorrentes. Ex: *abastecimento de água potável; abastecimiento de agua potable; drinking water supply.*
- b) existência de restrição de seleção combinatória entre esses elementos determinada pelas especificidades do âmbito temático em que são utilizados. Ex: *conservação da diversidade biológica; conservación de la diversidad biológica; conservation of biological diversity.*
- c) complexidade tradutória. Ex: *princípio de quem polui, paga; principio de que quien contamina paga; the polluter pays principle; gerenciamento dos recursos marinhos vivos; ordenación de los recursos marinos vivos; management of marine living resources.*

A segunda parte foi a busca dos equivalentes nas línguas espanhola e inglesa, a partir das combinatórias geradas na etapa anterior. Para essa etapa, foram seguidos os critérios abaixo:

- a) análise dos colocados dos termos;
- b) análise dos contextos de uso das combinatórias, a partir das concordâncias geradas com as ferramentas do ACERVO TERMISUL.

#### 4.2 Resultados Parciais

Foram identificadas 140 combinatórias léxicas especializadas em língua portuguesa e seus respectivos equivalentes em espanhol e inglês.

Iniciou-se a classificação morfossintática das combinatórias em português, sendo identificadas até o momento as seguintes estruturas:

- a) Nominalização + Sintagma preposicional (Preposição + Artigo + Termo)  
Ex: *avaliação do impacto ambiental; evaluación del impacto ambiental; environmental impact assessment.*
- b) Verbo + Artigo + Termo  
Ex: *combater a desertificação; luchar contra la desertificación; combat desertification; evitar a degradação da terra; evitar la degradación de las tierras; to prevent land degradation.*

#### 4.3 Próximas Etapas

Outras etapas devem ser cumpridas dentro do projeto-piloto:

- 1) busca dos equivalentes nas línguas alemã e francesa;
- 2) extração de candidatos a combinatórias nos outros textos que conformam a Base Legis e também em textos de Cardiologia e Pediatría;
- 3) seleção das combinatórias nesses textos;
- 4) classificação funcional das combinatórias, distinguindo-as em combinatórias do discurso especializado (*Para fins deste Protocolo*) e combinatórias léxicas especializadas (*combate à desertificação*);
- 5) classificação morfossintática das combinatórias;
- 6) criação de uma base ou catálogo de combinatórias, definindo os campos a serem incluídos, considerando principalmente o tradutor como seu usuário principal e o redator de textos, que será disponibilizada na página do ACERVO TERMISUL.

Com esse novo projeto sobre CLEs, damos continuidade às pesquisas desenvolvidas pelo Grupo, centrando-nos principalmente no oferecimento de produtos ou ferramentas que auxiliem os consulentes a conhecer os modos de dizer das áreas de conhecimento representadas no nosso *Acervo*, reiterando o viés aplicado. Da mesma forma, prosseguimos com a pesquisa teórica em temas que estabelecem a interface entre Terminologia, Linguística de *Corpus* e Tradução.

Com a apresentação desses trabalhos, mostramos a trajetória mais recente do nosso Grupo TERMISUL, ressaltando, por um lado, o desenvolvimento da pesquisa teórica e aplicada e, por outro, o oferecimento *on-line* de informações sobre usos, características e práticas textuais, em diferentes idiomas, das linguagens legal, normativa e técnico-científica. Esperamos auxiliar a consolidação dos estudos terminológicos no Brasil, compartilhando com a comunidade acadêmica nacional e internacional as pesquisas que estamos fazendo.

#### Referências bibliográficas

- BEAUGRANDE, R. de. 1980. *Text, discourse and process*. London: Longman.
- \_\_\_\_\_; DRESSLER, W. U. 1981. *Introduction to text linguistics*. London: Longman.
- BERBER SARDINHA, T. 2004. *Linguística de Corpus*. Barueri: Manole.
- CABRÉ, M. T. 1999. *La Terminología. Representación y comunicación*. Barcelona: IULA/UPF.
- \_\_\_\_\_; MOREL, J.; TEBÉ, C. 1998. Las relaciones conceituales de tipo causal; un caso práctico. In: *Actas del V Congreso de la Red Iberoamericana de Terminología*. México-DF. Paris: União Latina.

- CIAPUSCIO, G. 2003. *Textos especializados y terminología*. Barcelona: IULA/UPF.
- HALLIDAY, M. A. K.; HASAN, R. 1976. *Cohesion in English*. London: Longman.
- HOFFMANN, L. 1998. *Llenguatges d'especialitat. Selecció de textos*. Barcelona: IULA/UPF.
- KOCH, I. G. V. 2001. *A coesão textual*. São Paulo: Contexto.
- KRIEGER, M. da G.; FINATTO, M. J. B. 2004. *Introdução à Terminologia: Teoria & prática*. São Paulo: Contexto.
- \_\_\_\_\_; MACIEL, A. M. B.; BEVILAQUA, C. R.; FINATTO, M. J. B.; REUILLARD, P. C. R. 2006. *Glossário de Gestão Ambiental*. São Paulo: Disal.
- L'HOMME, M.-C. 2007. Using explanatory and combinatorial lexicology to describe terms. In: Wanner, L. (ed.). *Selected lexical and grammatical topics in the meaning-text theory*. Amsterdam/Philadelphia: John Benjamins.
- L'HOMME, M.-C. 2000. Understanding specialized lexical combinations. *Terminology* 6 n. 1. 89-110.
- REUILLARD, P. C. R. 2007. *Neologismos Lacanianos e equivalências tradutórias*. (Tese de Doutorado). Porto Alegre: Universidade Federal do Rio Grande do Sul.
- SAGER, J.-C. 1990 *A practical course in terminology processing*. Philadelphia: John Benjamins.
- \_\_\_\_\_; KAGEURA, K. 1994-1995. Concept classes and conceptual structures: their role and necessity. *Terminology* 7/8. 191-216.
- SINCLAIR, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- SWALES, J. M. 1990. *Genre analysis*. Cambridge/New York: Cambridge University Press.
- TEMERMANN, R. 2000. *Towards new ways of terminology description*. The sociocognitive approach. Amsterdam/New York: John Benjamins.