

Investigação e descrição de CLEs terminológicas em italiano e português: um estudo de caso

Ariane Ribeiro Marques¹
Cláudia Mendonça Scheeren²

RESUMO

A construção de um *corpus* de sua legislação ambiental possibilitou a inserção da língua italiana em um grupo de pesquisas que atualmente desenvolve um projeto de âmbito plurilíngue de unidades sintagmáticas, recorrentes e condicionadas pela língua, área ou gênero textual, denominadas Combinatórias Léxicas Especializadas (CLEs). A utilização de um *corpus* não só possibilita a análise dos modos de dizer dos diplomas legais em italiano, como também oportuniza sua comparação com a língua portuguesa e as outras línguas já contempladas pelo projeto, sendo elas: alemão, francês, espanhol e inglês. No presente trabalho, elabora-se um cotejo entre alguns dos resultados encontrados na atual etapa da pesquisa, que envolve a busca por CLEs Terminológicas: trata-se de combinatórias que contêm dentro de sua estrutura um termo e que podem ser divididas em Verbais, como é o caso, por exemplo, de “contaminar as águas” e em Nominalizadas como em “contaminação das águas”. Tendo como ponto de referência o termo “água(s)” e seu equivalente em italiano “acqua(e)”, descrevemos o processo de investigação das CLEs terminológicas nas duas línguas de trabalho, comparando em seguida os resultados obtidos em ambas no que tange à estrutura morfossintática das combinatórias, ao contexto em que estão inseridas e aos tipos de verbo que acompanham o termo, buscando, a partir disso, similaridades e disparidades entre elas. Como ferramenta de trabalho, utilizamos o programa AntConc 3.2.4, que permite buscar de forma rápida e eficiente as combinatórias existentes em nossos *corpora*. No que se refere à fundamentação teórica, baseamo-nos nas perspectivas de cunho comunicativo e textual da Terminologia, bem como nos princípios da Linguística de *Corpus*.

Palavras-chave: Terminologia, Combinatórias Léxicas Especializadas, *corpus*.

ABSTRACT

The corpus building of the environmental legislation of Italy has included the Italian language in a research group which is currently developing a multilingual research of the CLEs (specialized lexical combinatories), which are syntagmatic units, recurring and conditioned by the field, language or textual genre. The *corpus* has enabled us to analyze the language of the Italian statutes and also compare it with the other languages studied in our research project, which are: English, French, German and Spanish. In this paper, we develop a comparison between some of the results of the current stage of the project that concerns the search for Terminological “CLEs”, which are combinatories that include in their structure a term. The Terminological CLEs are divided in two types: the Verbal CLEs, which is the case of “to pollute the water” and

¹ Graduanda em Bacharelado em Letras – Tradutor Português e Italiano, Universidade Federal do Rio Grande do Sul

² Professora do Setor de Italiano do Departamento de Línguas Modernas, Universidade Federal do Rio Grande do Sul

the Nominalized ones, which is the case of “water pollution”. Focusing on the term in portuguese “água(s)” and its equivalent in italian “acqua(e)”, we describe the process of investigation of the Terminological “CLES” in both languages and then we compare the results regarding their morphosyntactic structure, the context in which they are included and the typology of the verbs that follow the term to find approximations and distances between them. As work tool we used the freeware concordance program Antconc 3.2.4, which allows us to search the combinatories in our available *corpora*. Our study was performed based on Textual Terminology; we also drawn on some theoretical and methodological principles of Corpus Linguistics.

Keywords: Terminology, specialized lexical combinatories, *corpus*.

Introdução

Neste trabalho, apresentaremos uma amostra da pesquisa realizada em língua italiana dentro do Grupo Termisul, mais especificamente, no atual projeto ProjeCOM, que visa o estudo das Combinatórias Léxicas Especializadas (CLEs) da Legislação Ambiental. O ProjeCOM é um projeto amplo que abrange as seis línguas de trabalho do Termisul, a saber: alemão, espanhol, francês, inglês, italiano e português. Dentre essas, a língua italiana foi a última a ser introduzida no projeto de pesquisas, com a construção do *corpus* da Legislação Ambiental, denominado Base Legis, no ano de 2010.

A Base Legis italiana possui uma amostragem suficiente da variedade de língua que pretendemos analisar, pois contém todos os diplomas legais italianos na área (197 textos) que foram coletados diretamente no *site* do Ministero dell’Ambiente³ até o ano de 2010, tendo um formato finito (37.981 *types* e 1.616.974 *tokens*). Todos os arquivos estão em formato eletrônico, transformados do formato .pdf para .txt, com as marcações necessárias para serem pesquisados e manipulados e serem lidos pelas ferramentas de extração utilizadas na pesquisa. O *corpus* da Base Legis pode ser consultado na página eletrônica do Projeto Termisul.⁴

Como citado anteriormente, o objeto de estudo do ProjeCOM são as Combinatórias Léxicas Especializadas (doravante CLEs) que são entendidas, segundo a noção de Bevilacqua *et al* (2011 e 2012) como unidades sintagmáticas recorrentes nas situações de comunicação das áreas temáticas que revelam uma preferência marcante pelas especificidades e pelas convenções próprias do idioma, da área e/ou do gênero de texto em que ocorrem. Acreditamos que as CLEs são expressões integrantes da competência linguística do falante nativo que transitam em um domínio especializado, mas que não fazem parte da competência do leigo - somente a repetição frequente e o consenso de uma comunidade de falantes que compartilham uma mesma especialidade justificam seu uso – por isso, precisam ser adquiridas pelo redator técnico não especialista na área, pelo tradutor, pelo professor e pelo aprendiz de línguas instrumentais.

A carência de glossários, dicionários, bases de dados mono ou bilíngues, que contemplem tais combinatórias motiva a realização desta pesquisa. Sabe-se que tradutores e redatores não encontram obras de referência disponíveis que sejam capazes de solucionar dúvidas que dizem respeito às CLEs contidas em seus textos de trabalho, principalmente no que se refere a documentos jurídicos. Opções equivocadas revelam desconhecimento das convenções da área especializada e geram significados inadequados que comprometem a qualidade do texto. As CLEs terminológicas são características da área do Direito sobre a qual o texto legislativo

³ <http://www.minambiente.it>

⁴ http://www6.ufrgs.br/termisul/bases_textuais/legis/legislacao_ambiental.php

incide, no caso em foco, o Direito Ambiental; sua função pragmática é a de referir uma ação ou processo próprios da sua área temática. Possuem essa nomenclatura por conter pelo menos um termo em sua constituição e se dividem em Verbais, como no caso de “contaminar as águas” e em Nominalizadas, como é o caso de “contaminação das águas”. É nesse grupo de combinatórias que este trabalho se foca, a partir de um recorte da pesquisa que vem sendo realizada em língua portuguesa e língua estrangeira partindo de termos relacionados ao meio ambiente – no caso em questão, o termo “água/águas” (em italiano, “*acqua/acque*”).

1. Fundamentação teórica

No que se refere à fundamentação teórica sob a qual sustentamos nosso trabalho, baseamo-nos na abordagem descritiva da Terminologia Textual (BOURIGAULT; SLODZIAN, 1999), que estuda a linguagem especializada realizada em seu contexto de uso; também nos voltamos para o pressuposto básico da teoria de cunho comunicativo da Terminologia (CABRÉ, 1998) que traz a noção da linguagem especializada vista como expressão da língua comum utilizada em contexto de especialidade, em que os recursos próprios da língua natural desempenham funções comunicativas específicas de determinada área temática.

Voltamo-nos concomitantemente para os princípios da Linguística de *corpus* (STUBBS, 1996; BIBER *et al.*, 1998), que propõe um estudo linguístico a partir de uma grande quantidade de textos reais e analisados por ferramentas digitais, em que são observados padrões reais de uso da língua. Entendemos a realização da linguagem dentro do meio jurídico tendo por base esse referencial teórico-metodológico.

2. Metodologia e desenvolvimento da pesquisa

Aos pesquisadores responsáveis por cada língua contemplada pelo projeto TERMISUL foi designado, para busca, extração e análise das candidatas à CLEs, um termo relativo à linguagem de cunho ambiental que possui frequência determinante nos textos legislativos pertencentes a nossos *corpora*. O termo é analisado inicialmente em língua portuguesa e logo após em língua estrangeira. À língua italiana foi designado o termo “água(s)” e seu equivalente “*acqua(e)*”. Nosso *corpus* de língua portuguesa é constituído de 280 textos legislativos brasileiros (totalizando 61.047 *types* e 593.593 *tokens*). Na base Legis em português o termo “água” aparece 480 vezes e “águas”, 540, totalizando 1020 ocorrências a serem analisadas. A partir desses dados, iniciamos o processo de busca e análise das candidatas a CLEs.

O primeiro passo envolveu a coleta de possíveis combinatórias: para isso, utilizamos como ferramenta o programa *Antconc* 3.2.3, *software* que nos permite listar as possíveis combinatórias existentes nos *corpora* analisados. Para permitir resultados mais pontuais e com maiores probabilidades de êxito no que envolve a listagem de CLEs terminológicas, efetuou-se a busca pelo termo “água” a partir de construções que, em língua portuguesa determinam nominalizações. Dois tipos de busca foram definidos, a fim de obter parâmetros que auxiliassem no processo de extração semi-automática e refinar os critérios de seleção das CLEs terminológicas. O primeiro tipo de busca está relacionado às preposições que antecedem o termo e que em geral fazem parte de sintagmas constituídos por nominalizações + preposição + termo. Iniciamos, então, com a busca das seguintes construções:

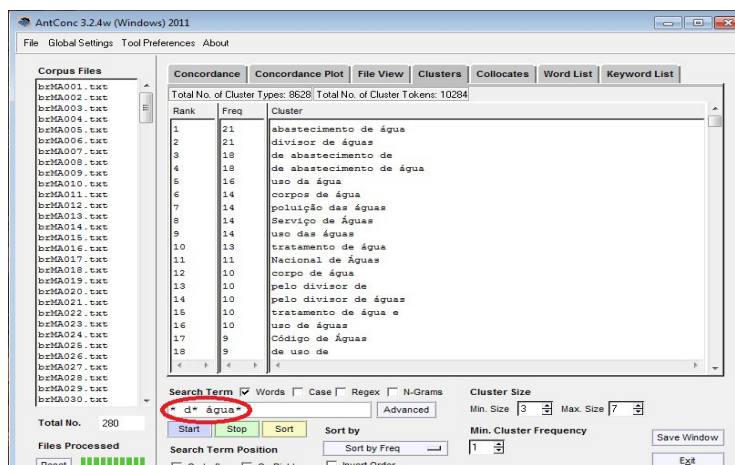
- * + d* + água* (o d* incluindo de, da, da(s) e do(s))⁵
- * + em + água*
- * + n* + água*
- * + para + água*
- * + às + água*
- * + sobre + água*

O segundo tipo de busca envolve os sufixos que formam nomes a partir de verbos em português que acompanham o termo e que formam nominalizações. Inserimos no *Antconc*, portanto, as seguintes construções:

- *ção + preposição (d* / às / em / n* / para / sobre) + água*
- *mento + preposição + água*
- *ança + preposição + água*

Após a análise dos dados da extração de candidatos a CLEs Nominalizadas no português, percebemos que seria necessário realizar um terceiro tipo de busca a partir da truncção dos verbos que geraram as nominalizações identificadas nas primeiras coletas, a fim de coletar possíveis CLEs verbais. Ou seja, após coletarmos uma combinatória como “poluição das águas”, buscamos no *Antconc* a construção “polui* * água*”, para averiguar se existe ocorrência do verbo no infinitivo, o que caracterizaria uma CLE terminológica de tipo verbal.

A tela abaixo, extraída do *Antconc*, mostra um exemplo de como é efetuada a coleta de possíveis CLEs a partir da produção de *clusters*, isto é, utilizando o termo como nóculo de agrupamentos lexicais de extensão predeterminada, com posicionamento à direita ou à esquerda da chave de busca. Determinamos a extensão dos *clusters* com o tamanho mínimo três e máximo sete e com frequência igual ou superior a um. A busca por *clusters* a partir dos dois tipos de construções acima citadas mostrou-se produtiva, já que permitiram identificar possíveis CLEs sem grande trabalho manual e produziu agrupamentos significativos e de frequência relevante que serviram de base para a segunda etapa do processo, que envolve a análise dos candidatos a combinatórias.



⁵ No programa *Antconc*, o asterisco (*) é utilizado como caractere curinga (*wildcard*), ou seja, um caractere que pode significar muitos outros. No caso de “água”, por exemplo, em que o termo no plural também nos interessa, inserimos o asterisco para obter resultados em ambas as formas.

Figura 1: Busca por *clusters* no programa *Antconc*

Após extração e listagem das possíveis CLEs, iniciamos o processo de análise individual de cada uma delas, a fim de averiguar a sua relevância e verificar seu possível caráter de especialidade. Para tal, fizemos uma listagem dos contextos em que elas estão inseridas, verificando também, a frequência com que aparecem no *corpus*. Verificamos que a frequência desse tipo de CLE é relativamente baixa – o candidato mais frequente em língua portuguesa (abastecimento de água) ocorre apenas 21 vezes, fator esse que torna necessária uma segunda verificação em um *corpus* maior. Como recurso para tal empreitada, utilizamos o *Webcorp* (*The Web as a Corpus*)⁶, ferramenta online que permite utilizar a internet como *corpus*, que filtra os resultados da rede a partir de *sites* de busca e permite indicar a área de pesquisa, trazendo resultados de concepção linguística.

Após a análise dos contextos em que se inserem as CLEs, procuramos identificar a variação da morfossintaxe entre elas a partir da elaboração de uma tabela que divide os verbos/nominalizações do núcleo terminológico. A tabela nos possibilita verificar os limites da CLE Nominalizada (já que é mais fácil delimitar sintagmas nominais em detrimento dos verbais) e a recorrência dos tipos de verbos que acompanham o termo.

Tabela 1: variação morfossintática em Língua Portuguesa

Núcleo Eventivo		Núcleo Terminológico
abastecimento	De	água
distribuição	De	águas públicas
poluição	Das	águas

Tendo em mãos o resultado preliminar da coleta por CLEs em português, iniciamos novamente todo o processo de extração, dessa vez partindo do termo em língua estrangeira, ou seja, do termo em italiano “*acqua(e)*”. O termo “*acqua*” em nosso *corpus* de língua italiana ocorre 624 vezes e “*acque*”, 1405, o que totaliza 2029 ocorrências. Refizemos todos os passos que envolveram o termo em língua portuguesa, desta vez, elaborando uma segunda e terceira tabelas: a primeira somente com as CLEs em língua italiana e a segunda, procurando estabelecer um paralelo ente os resultados obtidos nas duas línguas de trabalho. A partir dos contextos das CLEs em ambas as línguas e da elaboração da tabela comparativa, tornou-se possível cotejá-las a fim de buscar similaridades e disparidades entre elas e iniciar a busca por equivalentes.

Tabela 2: variação morfossintática em língua italiana

Núcleo Eventivo		Núcleo Terminológico
Approvvigionamento	di	acqua potabile
Fornitura	di	acqua
Tutela	delle	acque

Tabela 3: tabela comparativa português/italiano

Português	Italiano
Consumo de água	Consumo dell'acqua
Extração de água	Estrazione dell'acqua

⁶ www.webcorp.org.uk

	Estrazione di acqua
Fornecimento de água	Fornitura di acqua Rifornimento di acqua Rifornimento d'acqua

3. Resultados preliminares

Foram identificadas combinatórias que podem ser caracterizadas como núcleos terminológicos, como por exemplo, “qualidade das águas”. Foi necessário, portanto, separar essas ocorrências, já que elas não são o foco da atual pesquisa. A tabela morfossintática foi útil nesse ponto, já que sua delimitação permitiu que visualizássemos o que era termo e o que de fato se caracterizaria como combinatória.

Outro ponto relevante que observamos em ambas as línguas foi o fato de que a maior parte das candidatas a CLEs coletadas está no grupo das Nominalizadas. Foram pouquíssimas as ocorrências de CLEs Verbais e cada uma delas aparece no *corpus* apenas uma vez, o que dificulta uma possível busca por equivalentes, já que os verbos que ocorrem em português e italiano não são os mesmos. Identificamos também ocorrências de combinatórias “encaixadas”, ou seja, sintagmáticas. Abaixo, seguem dois exemplos desse tipo de CLE, o primeiro em português e o seguinte em italiano:

Auxiliam o **escoamento das águas**
Prevenire la **penetrazione di acqua**

As estruturas em negrito já caracterizam combinatórias, porém, cada uma delas está dentro de outra unidade verbal, que delimitamos aqui em itálico. São casos que, por ora, desconsideramos, já que provavelmente elas farão parte de outra categoria de CLE ainda não definida.

Observamos também que ambas as línguas aproximam-se no que tange à estrutura morfossintática das CLEs e também aos tipos de verbos que acompanham o termo. Tal semelhança, porém, não necessariamente colabora na busca por equivalentes, já que a análise prévia das combinatórias em ambas as línguas nos permitiu observar que há casos em que, para uma ocorrência de um verbo em português, há mais de uma possibilidade em italiano, como percebemos abaixo com a candidata à combinatória “utilização da água”:

Tabela 5: tabela comparativa português/italiano

Utilização da água	Utilizzazione dell'acqua
	Utilizzazioni di acqua
	Utilizzo dell'acqua
	Utenza di acqua
	Utenze di aqua

Observamos pela tabela, que temos, em italiano, cinco variações para a combinatória em português, variações essas que podemos dividir em dois blocos: o primeiro deles relativo à questão lexical, já que para o nome “utilização” temos três possíveis equivalências em italiano: “*utilizzazione*”, “*utilizzo*” e “*utenza*”. O segundo bloco refere-se à questão do número do

substantivo, já que não há a ocorrência de “utilizações” em nosso *corpus* em português, porém em italiano, com exceção de “*utilizzo*”, as nominalizações vão apresentar pelo menos uma ocorrência no plural. Torna-se vital, portanto, fazer uma análise profunda do contexto em que cada uma das ocorrências está inserida, para então verificar se cada verbo de língua italiana está sendo usado em uma situação diferente e qual deles de fato poderia equivaler à ocorrência em português.

4. Considerações finais e perspectivas

Identificar as CLEs, analisá-las e buscar equivalentes nas outras línguas de trabalho são tarefas extensas e que exigem minúcia e atenção. Estamos, ainda, dando os primeiros passos em uma pesquisa que culminará na elaboração de um glossário multilíngue disponível *online* que contemple as CLEs da linguagem jurídico-ambiental e que auxilie profissionais que transitem por essa área. Não podemos deixar de enfatizar a necessidade da colaboração de profissionais da área jurídica para o desenvolvimento da pesquisa, devido principalmente ao seu caráter multidisciplinar. A presença de especialistas é de vital importância, em especial no que se relaciona à discussão da natureza e classificação das CLEs e no reconhecimento de sinônimos e variáveis em uma mesma língua, equivalentes na língua estrangeira.

Os resultados que obtivemos até o momento são importantes no que diz respeito principalmente à aproximação de ambas as línguas na configuração das combinatórias, em suas estruturas morfossintáticas. Essa semelhança estrutural colabora na busca de equivalentes, porém, existem outros fatores em jogo que complexificam a pesquisa. Complexidade essa originada pela diversidade dos sistemas jurídicos de ambos os países, o que reflete na diversidade das funções pragmáticas das CLEs em cada um deles.

O próximo passo a ser dado é a análise profunda dos contextos em que cada uma das candidatas se insere a fim de estabelecer parâmetros que permitam corroborar seu caráter de especialidade – análise essa que em português auxiliará os pesquisadores das demais línguas de trabalho do projeto em suas futuras buscas a partir do termo “água”- e estabelecer possíveis equivalências que serão analisadas e discutidas juntamente com profissionais da área jurídica. A tarefa do ProjeCOM de determinar as combinatórias léxicas características do Direito Ambiental em uma perspectiva multilíngue e disponibilizá-las livremente foi iniciada recentemente, porém, já rende resultados importantes e determinantes para a continuidade da pesquisa.

Referências bibliográficas

- ANTHONY, L. *Antconc 3.2.3* (2008). Disponível em http://www.antlab.sci.waseda.ac.jp/antconc_index.html. Acesso em 10 out. 2012.
- BEVILACQUA, C. R. *A Fraseologia Jurídico-ambiental*. (1996). Porto Alegre: Universidade Federal do Rio Grande do Sul. Dissertação, Mestrado em Letras
- BEVILACQUA, C. R. ; MACIEL, A. M. B. ; SCHEEREN, C. M. ; KILLIAN, C.K. ; REUILLARD, P. C. R. . *CLEs da linguagem jurídica: as combinatórias discursivas do texto legislativo brasileiro*. In: II Congresso Internacional de Fraseologia e Paremiologia e I Congresso Brasileiro de Fraseologia,, 2012, Brasília. Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia. São Paulo: Pontes Editora, 2012. v. 2.

BEVILACQUA, C. R.; MACIEL, A. M. B.; SCHEEREN, C. M.; REUILLARD, P. C. R. (2012). *Combinatórias Léxicas Especializadas: etapas prévias para identificação e tratamento*. Buenos Aires: Colegio de Traductores Públicos de la Ciudad de Buenos Aires, t. 2, p. 272-283

BIBER, D. et al. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press

BOURIGAULT, D.; SLODZIAN, M. (1999). "Pour une terminologie textuelle". *Terminologies Nouvelles*, n.19, déc. 1998-juin.

CABRÉ, M.T. (1999). *La terminologia: representación y comunicación; elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra, IULA.

KRIEGER, M; G. BOCORNY FINATTO, M.J. (2004). *Introdução à Terminologia*. São Paulo: Contexto.

BERBER SARDINHA, T. (2004). *Linguística de Corpus*. Barueri, SP: Manole.

STUBBS, M. (1996). *Text and corpus analysis. Computer-assisted studies of language and Culture*. London: Blackwel.

MACIEL, A.M.B. (2001). *Para o reconhecimento da especificidade do termo jurídico*. Porto Alegre: Universidade Federal do Rio Grande do Sul. Tese de Doutorado.

WebCorp (*The Web as a Corpus*). Disponível em <http://www.webcorp.co.uk>. Acesso em 10 out. 2012.