



UNIVERSIDADE DO PORTO

FACULDADE DE LETRAS

MESTRADO EM TERMINOLOGIA E TRADUÇÃO

**PESQUISA DE INFORMAÇÃO TERMINOLÓGICA:
DOS MARCADORES LEXICAIS AOS *PADRÕES SUPORTE***

UM ESTUDO DE CASO NO DOMÍNIO DOS *RISCOS NATURAIS - CHEIAS*

Alexandra Sofia de Albuquerque Morais d' Almeida

Março 2006

UNIVERSIDADE DO PORTO
FACULDADE DE LETRAS

MESTRADO EM TERMINOLOGIA E TRADUÇÃO

PESQUISA DE INFORMAÇÃO TERMINOLÓGICA:
DOS MARCADORES LEXICAIS AOS *PADRÕES SUPORTE*

UM ESTUDO DE CASO NO DOMÍNIO DOS *RISCOS NATURAIS - CHEIAS*

Alexandra Sofia de Albuquerque Morais d' Almeida

Março 2006

**Dissertação da candidatura ao grau de
Mestre, no âmbito do Mestrado em
Terminologia e Tradução, sob orientação
da Professora Doutora Belinda Maia
(F.L.U.P.).**

People do not think in English, Chinese or Apache; they think in a language of thought.

This language of thought perhaps looks rather like all of these languages [...].

Pinker (1994, citado por Teubert, 1977, s/ página)



in *Expanding the Horizons of Disaster Research*

Natural Hazards Observer Vol. XXVIII No. 4 March 2004¹

¹ Disponível em <http://www.colorado.edu/hazards/o/mar04/mar04a.html>, acessado em 2004.11.05.

ÍNDICE

Índice de quadros	vii
Agradecimentos	viii
Resumo	ix
Abstract	x
Introdução	11
I Parte – Enquadramento Teórico	15
Capítulo 1. As principais bases teóricas do estudo	16
1.1. Terminologia, Termo e Línguas de Especialidade	16
1.1.1. Introdução	16
1.1.2. Terminologia	17
1.1.2.1. Definição e caracterização	17
1.1.2.2. Teorias e principais postulados da disciplina: da Teoria às teorias	22
1.1.2.2.1. A Teoria Geral da Terminologia	23
1.1.2.2.2. As teorias contemporâneas da Terminologia	25
1.1.3. Termo	31
1.1.3.1. Definição e caracterização	32
1.1.3.2. Distinção entre termo e não-termo	34
1.1.4. Línguas de especialidade: definição e caracterização	39
1.2. Linguística de <i>Corpus</i> e <i>Corpus</i>	41
1.2.1. Breve perspectiva histórica	41
1.2.2. Linguística de <i>Corpus</i> e <i>Corpus</i> : definição, caracterização e aplicações práticas	42
1.2.3. <i>Corpus</i> de Especialidade	44
Capítulo 2. A definição	48
2.1. Definir “definição”	48
2.2. Tipologia de definições	50
2.3. A definição terminológica	51
2.3.1. Sua definição e caracterização	51
2.3.2. Sistematização das suas principais regras	53
Capítulo 3. O contexto e a extracção de informação terminológica	55
3.1. Introdução	55
3.2. Contexto: definição e caracterização	57
3.2.1. O que é o contexto?	57
3.2.2. Tipologia dos contextos	59
3.3. Extracção de informação terminológica relevante: termos, definições, contextos de uso, sistemas conceptuais e redes semânticas	61
3.3.1. As propostas de Dubuc (1992)	61
3.3.2. As propostas de Sager (1990)	64
3.3.3. A proposta de Pearson (1998)	67
3.3.4. A proposta de Conceição (2001 e 2004)	68
3.3.5. A proposta de Colussi (2002)	71
Capítulo 4. Os marcadores lexicais indiciadores de informação terminológica e os <i>padrões suporte</i>	73
4.1. Os marcadores lexicais indiciadores de informação terminológica	74
4.1.1. O uso de verbos e grupos verbais como marcadores lexicais de definição	77
4.1.2. O caso específico do verbo <i>SER</i> / (<i>to</i>) <i>BE</i>	78
4.1.3. Os marcadores lexicais indiciadores de informação terminológica e o <i>Corpógrafo V2</i>	79
4.2. Os <i>padrões suporte</i> e o <i>Corpógrafo V3</i>	82

II PARTE – Estudo Empírico	86
Capítulo 5. Metodologia	87
5.1. Etapa 1 – Definição dos objectivos do estudo	88
5.2. Etapa 2 – Escolha e descrição dos instrumentos de trabalho	88
5.2.1. O <i>Corpógrafo</i>	88
5.2.2. Escolha e introdução ao domínio conceptual	90
5.2.3. O especialista e o seu contributo	92
5.2.4. As línguas de trabalho	93
5.3. Etapa 3 – Pesquisa bibliográfica: recolha, selecção e apreciação dos textos	94
5.3.1. Recolha e selecção dos textos	94
5.3.2. Apreciação dos textos	96
5.4. Etapa 4 – Organização do domínio e representação do mapa conceptual	98
5.5. Etapa 5 – Constituição e apreciação dos <i>corpora</i> e das Bases de Dados no <i>Corpógrafo</i>	101
5.5.1. Constituição e apreciação dos <i>corpora</i>	101
5.5.2. Constituição e apreciação das Bases de Dados no <i>Corpógrafo</i>	102
5.6. Etapa 6 – Processo de extracção de informação terminológica a partir dos <i>corpora</i> para análise e apresentação dos resultados	103
Capítulo 6. Apresentação e análise descritiva dos resultados	107
6.1. A primeira possibilidade	108
6.2. A segunda possibilidade	111
6.3. A terceira possibilidade	113
6.4. A quarta possibilidade	116
Conclusões	123
Bibliografia e fontes	128
Anexos	136
Anexo I	137
Quadro 7 – Marcadores linguísticos indiciadores de informação terminológica [Língua Portuguesa - Singular]	139
Quadro 8 – Marcadores linguísticos indiciadores de informação terminológica [Língua Portuguesa - Plural]	141
Quadro 9 – Marcadores linguísticos indiciadores de informação terminológica [Língua Inglesa - Singular]	142
Quadro 10 – Marcadores linguísticos indiciadores de informação terminológica [Língua Inglesa - Plural]	143
Quadro 11 – <i>Padrões suporte</i> utilizados pelo <i>Corpógrafo V3</i> [Língua Portuguesa – Singular]	144
Quadro 12 – <i>Padrões suporte</i> utilizados pelo <i>Corpógrafo V3</i> [Língua Portuguesa – Plural]	149
Quadro 13 – <i>Padrões suporte</i> utilizados pelo <i>Corpógrafo V3</i> [Língua Inglesa – Singular]	154
Quadro 14 – <i>Padrões suporte</i> utilizados pelo <i>Corpógrafo V3</i> [Língua Inglesa – Plural]	158
Anexo II	162
Quadro 15 - Descrição do <i>corpus</i> em língua portuguesa	163
Quadro 16 - Descrição do <i>corpus</i> em língua inglesa	171

ÍNDICE DE QUADROS

Quadro 1 - Marcadores lexicais indiciadores de informação terminológica [Língua Portuguesa – Singular]	80
Quadro 2 - Marcadores lexicais indiciadores de informação terminológica [Língua Portuguesa – Plural]	80
Quadro 3 - Marcadores lexicais indiciadores de informação terminológica [Língua Inglesa – Singular]	81
Quadro 4 - Marcadores lexicais indiciadores de informação terminológica [Língua Inglesa – Plural]	81
Quadro 5 - <i>Padrões Suporte</i> utilizados pelo <i>Corpógrafo V3</i> [Língua Portuguesa]	83
Quadro 6 - <i>Padrões Suporte</i> utilizados pelo <i>Corpógrafo V3</i> [Língua Inglesa]	84
Quadro 7 - Marcadores linguísticos indiciadores de informação terminológica [Língua Portuguesa - Singular]	139
Quadro 8 - Marcadores linguísticos indiciadores de informação terminológica [Língua Portuguesa - Plural]	141
Quadro 9 - Marcadores linguísticos indiciadores de informação terminológica [Língua Inglesa - Singular]	142
Quadro 10 - Marcadores linguísticos indiciadores de informação terminológica [Língua Inglesa - Plural]	143
Quadro 11 - Padrões suporte utilizados pelo <i>Corpógrafo V3</i> [Língua Portuguesa – Singular]	144
Quadro 12 - Padrões suporte utilizados pelo <i>Corpógrafo V3</i> [Língua Portuguesa – Plural]	149
Quadro 13 - Padrões suporte utilizados pelo <i>Corpógrafo V3</i> [Língua Inglesa – Singular]	154
Quadro 14 - Padrões suporte utilizados pelo <i>Corpógrafo V3</i> [Língua Inglesa – Plural]	158
Quadro 15 - Descrição do corpus em língua portuguesa	163
Quadro 16 - Descrição do corpus em língua inglesa	171

Agradecimentos

Ao meu marido, uma homenagem como reconhecimento da sua infindável paciência e apoio para que o trabalho se concretizasse.

Este trabalho é também dedicado à minha mãe, *in memoriam*, ao meu pai e aos meus irmãos Cristina, Cláudia e João, à Elisa e a todos os familiares e amigos que sempre me apoiaram. À Cristina Queirós, pela amizade e encorajamento constante. À Maggie, pela companhia. A todos, muito obrigada.

Um agradecimento muito especial à Professora Doutora Belinda Maia pela orientação e por sempre ter acreditado no meu trabalho, ao precioso contributo da equipa da Linguateca responsável pelo *Corpógrafo*; à Professora Doutora Fantina Pedrosa, pela orientação na área da Geografia – Riscos Naturais e, por fim, à Faculdade de Letras da Universidade do Porto.

Resumo

O crescente interesse na Terminologia, bem como na Linguística de *Corpus*, têm contribuído de sobremaneira para o rápido desenvolvimento de técnicas de pesquisa e extracção de informação terminológica.

Este estudo propõe uma abordagem na pesquisa terminológica em textos de domínios de especialidade, na detecção e extracção de unidades terminológicas e de informação contextual, incluindo definições, contextos de uso, sistemas conceptuais e redes semânticas, entre outros. Esta metodologia é potenciada através da utilização de determinados recursos linguísticos em programas informáticos vocacionados para o estudo das línguas. Destes recursos linguísticos destacamos os marcadores lexicais indiciadores de informação terminológica, com especial enfoque nos marcadores lexicais de definição, assim como os *padrões suporte*, propostos pelo *Corpógrafo*.

O *Corpógrafo*, como ferramenta integrada e baseada na *World Wide Web*, permite a extracção de informação terminológica relevante para diversas aplicações práticas através do processamento computacional de textos de domínio conceptual específico. A metodologia de trabalho estabelecida para este estudo permitiu criar condições para ensaiar, no *Corpógrafo*, a viabilidade de alguns dos referidos marcadores lexicais e *padrões suporte* na tarefa de detecção e extracção de informação terminológica no domínio conceptual dos *Riscos Naturais: Cheias*. A eficiência do desempenho do *Corpógrafo* é avaliada a partir dos resultados obtidos de forma automática e semi-automática, variando com o nível de intervenção nas diversas áreas de pesquisa proporcionadas pelo *Corpógrafo*.

Numa perspectiva meramente descritiva, procedemos à observação do comportamento e à avaliação do grau de eficácia dos marcadores lexicais indiciadores de informação terminológica assim como dos *padrões suporte*. O sucesso dos resultados finais permite-nos reconhecer que o *Corpógrafo*, juntamente com os marcadores lexicais indiciadores de informação terminológica e os *padrões suporte*, constituem uma abordagem nova e eficaz em pesquisa terminológica.

Palavras-chave: “pesquisa terminológica”, “*Corpógrafo*”, “marcadores lexicais”, “*padrões suporte*”, “informação contextual”, “*Riscos Naturais: Cheias*”

Abstract

Increasing interest in Terminology, as well as in *Corpus Linguistics*, has contributed to the fast development of research into techniques for extracting terminological information. This study considers an approach in terminological research in texts in specialized domains that involve the detection and extraction of terminological units and contextual information, including definitions, contexts of use, conceptual systems and semantic networks, among other things. This methodology makes use of certain linguistic resources in data processing programs dedicated to the study of the languages. Special attention is paid to the lexical markers indicative of terminological information, with a special focus on definition lexical markers, as well as the *support patterns*, suggested by the *Corpógrafo*.

As a tool integrated and based on the *World Wide Web*, the *Corpógrafo* allows the extraction of relevant terminological information for diverse practical applications through the computational processing of special domain texts. The methodology established for this study allowed us to experiment with the *Corpógrafo*, in an attempt to verify the viability of some of the above-mentioned lexical markers and the *support patterns* in the task of detecting and extracting terminological information in the conceptual domain of *Natural Hazards: Floods*. The efficiency of the *Corpógrafo*'s performance is evaluated from the results obtained automatically and semi-automatically, and varies with the level of intervention in the diverse research areas provided by the *Corpógrafo*.

From a merely descriptive perspective, we proceeded to observe the performance and to evaluate the degree of effectiveness of the lexical markers indicative of terminological information as well as of the *support patterns*. The success of the final results allows us to recognize that the *Corpógrafo*, together with the lexical markers indicative of terminological information and the *support patterns*, constitutes a new and efficient approach in terminological research.

Key-words: "terminological research", "lexical markers", "*Corpógrafo*", "Support Patterns", "contextual information", "Natural Hazards: Floods".

Introdução

Este trabalho, realizado no âmbito do Mestrado em Terminologia e Tradução da Faculdade de Letras da Universidade do Porto (F.L.U.P.), tem como objecto de estudo a descrição e a avaliação do comportamento dos marcadores lexicais indiciadores de informação terminológica bem como dos *padrões suporte*² na obtenção de informação terminológica relevante em contextos do domínio conceptual específico dos *Riscos Naturais: Cheias*. Para além da inerente contextualização dos conceitos abordados, necessária ao melhor entendimento da hipótese por nós colocada, procederemos à delineação de algumas perspectivas teóricas que julgamos de maior credibilidade. Neste sentido, destacaremos não só as teorias e principais postulados teóricos da Terminologia, que se têm vindo a multiplicar, proporcionando, ao mesmo tempo, um aprofundamento científico no estudo desta disciplina, mas também o inestimável contributo da Linguística de *Corpus*, dos *corpora* e do estudo mais aprofundado das línguas de especialidade. Do ponto de vista empírico, com o conhecimento alcançado, procuraremos estudar o contributo, numa pesquisa terminológica, dos marcadores lexicais indiciadores de informação terminológica (extraídos manualmente do *Corpógrafo* V2 [Versão 2]), com especial destaque para os marcadores de definição, assim como dos *padrões suporte* (apresentados automaticamente pelo *Corpógrafo* V3 [Versão 3]), a partir de contextos de uma área de conhecimento de especialidade.

Foram vários os pontos de partida desta análise. Em primeiro lugar, a proposta apresentada pelo Departamento de Geografia desta Faculdade para a elaboração de um glossário sob o tema *Riscos Naturais: Cheias*. Em segundo lugar, a possibilidade de utilização da ferramenta informática disponibilizada pela *Linguateca*³: o *Corpógrafo* (versões V2 e V3). Em terceiro lugar, a percepção da forte possibilidade da obtenção de resultados encorajadores no âmbito da extracção de informação terminológica através da utilização de recursos linguísticos, como é o caso dos marcadores lexicais indiciadores de informação terminológica de diferentes características (de definição, de reformulação, explicativos, etc.), em associação com os *padrões suporte*. Partimos do

² Conforme terminologia utilizada pelo *Corpógrafo* (cf. <http://www.linguateca.pt/corpografo/>).

³ Trata-se de um centro de recursos vocacionado para o processamento computacional da língua portuguesa, conforme descrito no seu portal de apresentação disponível na Internet (<http://www.linguateca.pt/>). O seu objectivo de servir todos aqueles que se dedicam ao processamento da língua portuguesa leva a que as actividades desenvolvidas pela *Linguateca* se caracterizem pela total abertura e pelo acesso e disponibilização livres.

pressuposto que, tanto os marcadores lexicais, como os *padrões suporte* contribuiriam para a apresentação de uma nova proposta na detecção e extracção de informação terminológica (candidatos a termos, definições, contextos, etc.), bem como na obtenção de dados necessários à elaboração de definições terminológicas ou estabelecimento de ontologias, sistemas conceptuais e redes semânticas. Em quarto e último lugar, será de referir que os marcadores lexicais por nós escolhidos para extracção de informação terminológica representam uma síntese integrativa de cinco perspectivas de propostas defendidas por cinco autores da nossa escolha, tendo sempre por base o elevado grau de importância da informação contextual, com as inerentes variações dependendo das várias abordagens teóricas.

Este estudo tem como objectivo analisar e compreender os processos e mecanismos que levam a uma obtenção máxima de informação terminológica, quer em quantidade, quer em qualidade, tentando também retirar e explorar as vantagens proporcionadas pelo *Corpógrafo*, como instrumento potenciador de estudos empíricos, sobretudo os direccionados para a automatização da obtenção dos dados de carácter terminológico.

Desde logo ficou presente a ideia que a produção de uma análise sobre contextos e extracção de informação terminológica teria forçosamente de incluir um cenário de abordagem diferente daqueles que até agora foram elaborados. Se, por um lado, o conhecimento da forma de funcionamento do *Corpógrafo* ao nível da extracção de informação facilitou a compreensão de alguns dos fenómenos próprios de uma área de estudo tão diversificada, por outro lado, havia algum desconhecimento em relação ao resultado do desempenho dos marcadores lexicais indiciadores de informação terminológica para obtenção de informação terminologicamente relevante.

Apesar das várias limitações relacionadas com a readequação do nosso estudo inicialmente proposto à nova versão do *Corpógrafo*, procuraremos, neste trabalho determinar algumas orientações para tirar partido do *Corpógrafo* a partir da utilização de marcadores de informação terminológica e *padrões suporte*. Este trabalho de investigação é encarado numa perspectiva descritiva, dado o nosso objectivo ser a recolha de informação terminológica relevante a partir da observação e análise contextual e posterior descrição de fórmulas lexicais, referentes a relações entre conceitos, indiciadoras da presença desses termos, definições e outra informação terminológica. Trata-se de apresentar os resultados apurados através do *Corpógrafo*, pelo que não é nosso propósito atribuir juízos prescritivos sobre qualquer um dos

marcadores lexicais ou *padrões suporte*. A língua geral, como veículo de comunicação é normalmente prescritiva. Caso assim não fosse, não seria possível a transmissão de conhecimentos e informação. Cada sistema linguístico tem as suas regras linguísticas (lexicais, sintácticas, morfológicas, fonológicas e até semânticas) que, se não forem respeitadas, tornam o acto de comunicar difícil ou até mesmo impossível. Nesse sentido, e mesmo tendo em conta as variações geo-socio-culturais, tanto a língua geral como a língua de especialidade têm essencialmente funções prescritivas.

Os princípios teóricos e metodológicos sobre os quais assentam as investigações terminológicas mais recentes põem em evidência não só a (indispensável) utilização de *corpora* textuais, mas também de programas informáticos desenvolvidos no sentido de maximizarem a extracção de dados. Partimos, assim, do pressuposto que estas fórmulas seriam encontradas em discursos científicos, técnicos e, sobretudo, didácticos, do domínio específico dos *Riscos Naturais: Cheias*.

Apesar de algumas diferenças de opinião sobre o tipo de orientação que deve ser aplicada numa pesquisa terminológica, poderemos concluir que as mais recentes propostas de orientações terminológicas apostam claramente nos estudos efectuados a partir de *corpora* constituídos por textos reais de um dado domínio do conhecimento, produzidos com objectivos informativos e comunicativos.

Com efeito, o acentuado desenvolvimento da informática permite a constante *reinvenção* dos vários aspectos da vida humana em sociedade. A comunicação tornou-se imediata e universal, pelo que os contactos culturais são inevitáveis. Também por estes motivos, as necessidades terminológicas aumentaram rapidamente. Nesta sequência, o *Corpógrafo* corresponde a um progresso importante na detecção e extracção automática e semi-automática de termos e outros dados, devidamente contextualizados, a partir de textos reais e representativos de uma área específica do conhecimento.

O presente trabalho é formado por duas partes, cada uma delas constituída por vários capítulos, sendo na primeira apresentado o enquadramento teórico e, na segunda, o estudo empírico por nós realizado.

Nos dois primeiros capítulos que constituem o enquadramento teórico apresentamos as principais bases teóricas que caracterizam os estudos terminológicos mais recentes. Faremos igualmente alusão a teorias menos recentes, de forma a estabelecer uma base comparativa com as novas tendências da pesquisa terminológica. Nas considerações teóricas finais (Capítulos 3 e 4), salientaremos os elementos

específicos que contribuem, de forma particular, para o sucesso da extracção (automática e semi-automática) de informação terminológica (contextos, marcadores propostos por vários autores, marcadores extraídos do *Corpógrafo V2* e *padrões suporte* apresentados pelo *Corpógrafo V3*). É a partir destes recursos terminológicos que proporemos recolher pistas para compreender melhor a forma de extrair automaticamente informação terminológica relevante a partir do *Corpógrafo V3*.

Relativamente ao estudo empírico realizado, este consiste na descrição da metodologia aplicada para a execução deste trabalho, seguida da apresentação e análise dos resultados extraídos do *Corpógrafo*. No Capítulo 5, identificaremos as etapas percorridas para levar a bom termo o nosso estudo, realçando alguns dos aspectos particulares que caracterizam o nosso estudo. Destacaremos nomeadamente o *Corpógrafo* e o seu papel no estudo, assim como a descrição do processo de extracção de informação terminológica a partir dos *corpora*. No Capítulo 6, faremos a apresentação e daremos conta da análise dos resultados obtidos considerados de maior relevo para o nosso estudo.

Por fim, apresentaremos as principais conclusões alcançadas em relação ao contributo dos marcadores de informação terminológica e os *padrões suporte* em conjugação como o *Corpógrafo* numa pesquisa terminológica com características idênticas às propostas para o dos *Riscos Naturais: Cheias*. O trabalho termina com a descrição da bibliografia consultada e a apresentação dos documentos anexos.

I PARTE

ENQUADRAMENTO TEÓRICO

1. As duas principais bases teóricas do estudo

A Terminologia e a Linguística de *Corpus* constituem as duas principais bases teóricas a partir das quais desenvolvemos o nosso estudo, pelo que este capítulo é dedicado à sua definição e caracterização.

1.1. Terminologia, Termo e Línguas de Especialidade

1.1.1. Introdução

A sociabilização do ser humano permite-lhe desenvolver, entre muitas outras capacidades, linguagens com o intuito de comunicar, de transmitir conhecimentos e organizar os seus pensamentos. A expansão e o aperfeiçoamento tecnológicos na área da informação permitem, nos dias de hoje, uma comunicação de “massas” imediata e universal. Um sistema linguístico é, conseqüentemente, reflexo espaço-temporal das sociedades e práticas discursivas estabelecedoras, geradoras e recriadoras de conhecimentos. Esses constantes avanços tecnológicos, científicos e sociais, aliados a uma organização conceptual linguística das diversas áreas do conhecimento humano, têm posto em evidência a importância da terminologia no panorama mundial contemporâneo.

A realidade é que a terminologia já não se traduz apenas na criação de bases de dados ou na elaboração de listas de palavras mercê de um fim específico de uma determinada área do conhecimento. Apesar disso, por vezes, o terminólogo é ainda visto numa óptica reducionista como uma figura que se dedica a apresentar termos e respectivas definições, como que preenchendo lacunas existentes no discurso de especialidade. Contudo, a prestação do terminólogo vai mais além: o seu contributo traduz-se, igualmente, na definição de ontologias, sistemas conceptuais e de redes semânticas que permitem a organização e a categorização das estruturas do conhecimento envolvido, conforme refere Maia (2003a: 60).

Por tudo isto será consentâneo afirmar que todos os domínios de especialidade necessitam da terminologia, ou seja, a terminologia existe desde que existe a necessidade de comunicar de um modo inequívoco e preciso, quer na forma, quer no conteúdo. A terminologia é hoje em dia um valioso recurso para o estabelecimento da clareza e, sobretudo, da precisão conceptual. A precisão é, aliás, uma das condições

necessárias e essenciais à comunicação entre especialistas, não só do mesmo domínio, como também de domínios conceptuais diferentes.

A relevância destas características é tanto mais evidente se atendermos que a língua não é a realidade, mas sim a forma mais comum de representar o que o ser humano concebe como a realidade e que pode ser descrita segundo diferentes perspectivas. Com efeito, a língua tem uma vertente sócio-cultural bastante acentuada e é, desta forma, frequente a existência de várias perspectivas em relação ao termo apropriado para um mesmo conceito. Daí que a univocidade de uma língua ou de um discurso de especialidade seja considerado um fim a atingir por via das características atribuídas às unidades terminológicas “*consideradas como ideais de expressão de monorreferencialidade, de monossema e de exclusividade denominativa.*” (Krieger, 2000: 1).

1.1.2. Terminologia

1.1.2.1. Definição e caracterização

Não existe, para a Terminologia uma definição consensual e definitiva, não só devido às diferentes abordagens teóricas, mas também devido ao carácter polissémico do próprio termo, que pode designar uma teoria ou disciplina, uma actividade ou prática e, ainda, o produto gerado por essa prática. Na primeira acepção indica que se ocupa dos termos, na segunda refere-se ao conjunto de princípios que regem a recompilação das unidades terminológicas e, por fim, na terceira acepção diz respeito ao vocabulário de uma área de especialidade. Existe, porém, uma enunciação convencionada pela *International Organization for Standardization* (ISO), através da sua Norma ISO/FDIS 1087-1:2000 (3.5.1. e 3.5.2) que a define como um conjunto de designações pertencentes a uma língua de especialidade e, no âmbito de disciplina, como a ciência que estuda a estrutura, formação, desenvolvimento, uso e gestão das terminologias em diferentes domínios. No seu sentido mais restrito, terminologia designa uma “*disciplina linguística consagrada ao estudo científico dos conceitos e termos usados nas línguas de especialidade*” (Pavel *et al.*, 2002: xvii).

Sem a pretensão de sermos exaustivos, julgamos importante reconhecer, ainda, o contributo do *Pointer Final Report* que no seu capítulo “The Importance of Terminology” define terminologia como “*a structured set of concepts and their*

designations (graphical symbols, terms, phraseological units, etc) in a specific subject field” para logo de seguida apresentar, de forma sucinta, três pontos que entende como fulcrais para a clarificação do conceito geral de terminologia:

- a terminologia preocupa-se com a relação entre conceitos e entre estes e as suas designações, e não apenas com as designações ou com os objectos que representam. Este ponto é essencial, caso se pretenda atingir a qualidade, sobretudo no que se refere a sinónimos e ambientes multilingues;
- a designação não tem necessariamente que ser uma palavra ou uma frase, apesar de, com frequência, o ser. Os recursos terminológicos podem incluir símbolos, desenhos, fórmulas, códigos, etc. Este ponto é especialmente importante dada a evolução para os sistemas multimédia;
- a terminologia está inevitavelmente ligada ao conhecimento especializado e, por conseguinte, às línguas de especialidade ou línguas para fins específicos.⁴

Daí que as características apontadas como desejavelmente intrínsecas ao vocabulário e discurso de especialidade (de que é exemplo a univocidade conceptual e designativa) tenham uma importância preponderante para o bom entendimento entre os interlocutores de um domínio específico. É precisamente a aprendizagem de uma terminologia, correspondendo à apreensão das respectivas unidades terminológicas e conceitos através do discurso e do significado, que permite ao sujeito utilizador desse vocabulário de especialidade ter consciência da presença do termo nessa logoesfera específica.

Os utilizadores de terminologias, como é o caso da referente às *Cheias (Riscos Naturais)*, não são, necessária e exclusivamente, especialistas da área. Existem outros usuários de áreas conexas e de outras mais indirectas que necessitam desta terminologia como ferramenta auxiliar para o bom desempenho nas suas próprias áreas de acção. Tomem-se como exemplos as entidades agregadas à Protecção Civil, seguradoras, biólogos, geólogos, ambientalistas, tradutores, intérpretes, legisladores, entre muitos outros. Existe, assim, uma transversalidade no uso desta terminologia sendo que esse uso varia consoante o grau do conhecimento do domínio, da língua e do discurso de especialidade, por parte do utilizador. As abordagens mais recentes da terminologia englobadoras dos factores sócio-culturais, para além de valorizarem a variação

⁴ Nossa tradução. Disponível em: <http://www.computing.surrey.ac.uk/ai/pointer/report/section1.html#1> [acedido em 2005.10.15].

linguística em toda a sua dimensão, atendem efectivamente às variações acima descritas.

Um outro aspecto essencial, caracterizador da terminologia, é que esta funciona como meio de fixação da língua e da sua evolução. Caso não fossem criados neologismos e neónimos, a língua portuguesa, por exemplo, nunca teria avançado nos campos das novas tecnologias e sujeitar-se-ia ao uso de vocabulário importado de uma outra língua estranha à sua. Neste sentido, é fácil determinar a significativa relevância da terminologia no planeamento linguístico de uma comunidade, país ou estado. A consciência da importância da terminologia num plano nacional e internacional levou autores como Rondeau (1984: XV) a afirmar que “(...) *il ne peut y avoir développement sans aménagement linguistique ni aménagement linguistique sans terminologie* (...)”. Tome-se como exemplo o caso paradigmático do Canadá, país bilingue, que por questões políticas, económicas e sociais investiu fortemente num planeamento linguístico sistematizado com evidentes resultados de sucesso nos sectores onde é aplicado. Tornou-se, ao mesmo tempo, uma superpotência mundial da Terminologia (e da Tradução) nas suas vertentes teórica e prática com o objectivo pragmático de harmonizar e normalizar os elementos necessários à comunicação entre usuários das diferentes línguas, neste caso o inglês e o francês. A crescente disponibilização de produtos terminológicos ao serviço de áreas como a Tradução tem uma profunda influência na produtividade através da redução do tempo de tradução, da melhoria da sua qualidade e da poupança de esforço e de dinheiro ao minimizar a duplicação de trabalho.

No entanto, quando existe uma preocupação de planeamento linguístico a nível mundial (por exemplo, a língua portuguesa em Portugal, Brasil e nos restantes países da CPLP – Comunidade de Países de Língua Portuguesa), não é possível colocar em primeiro plano a normalização. Não é pela força da imposição do uso de uma determinada língua considerada padrão que as diferentes comunidades a adoptam no seu discurso. Pensamos que a normalização terminológica deve surgir como resultado de um consenso, respeitando sobretudo o seu uso. Neste sentido, o planeamento linguístico tem como uma das preocupações prementes a abordagem sociológica (que compreende a variante do uso), podendo definir, inclusivamente, o grau de importância dessa variante.

Como resultado há já alguns anos que se tem vindo a acentuar o relevo dado à investigação socioterminológica o que, associado ao crescente progresso da tecnologia

(sobretudo computacional) tem revolucionado a visão da Terminologia como teoria e como prática. O desenvolvimento da Terminologia e de terminologias abre, desta forma, um novo caminho ao estudo multi- e interdisciplinar numa território até então dominado quase exclusivamente por linguistas.

Tendo por base o que Wüster refere sobre a interdisciplinaridade, ao posicionar a Terminologia na intersecção de teorias e disciplinas como a Linguística, a Lógica, a Ontologia, a Ciência da Informação, a Linguística Computacional entre outras áreas do saber, Cabré (1999: 25) acrescenta que tal como qualquer outra área científica interdisciplinar, a Terminologia é definida tendo em conta as outras áreas às quais vai buscar conjuntos conceptuais específicos. Esta autora concretiza esta ideia referindo que a interdisciplinaridade da Terminologia é determinada pelas características das unidades terminológicas, que são simultaneamente unidades linguísticas (Linguística), elementos cognitivos (Lógica e Ontologia, ou seja, parte da Ciência Cognitiva) e veículos de comunicação (Teoria da Comunicação); conclui referindo que os termos surgem em comunicações especializadas (Ciência da Informação) e os computadores são normalmente utilizados nas actividades terminográficas (Ciência Computacional).

São vários os autores que apontam a multi- e interdisciplinaridade como características preponderantes da Terminologia. Nascimento (2003: 2), por exemplo, descreve, de uma forma bastante sucinta e ao mesmo tempo abrangente, a Terminologia como “(...) um campo interdisciplinar construído a partir das contribuições de uma *teoria do conhecimento*, de uma *teoria da comunicação* e de uma *teoria da linguagem* (...)”. Alsina *et al.* (1996: 85) argumentam que a Terminologia é uma matéria interdisciplinar, que na sua vertente teórica, colhe conceitos de outras teorias para elaborar a sua própria teoria e, na sua vertente aplicada, integra-se noutras disciplinas enquanto também se serve delas. Complementando esta ideia, Sonneveld *et al.* (1993: 1-3) identificam a multidisciplinaridade como uma característica da Terminologia já que esta reúne diferentes áreas de estudo possibilitando e encorajando uma constante interacção em benefício de todas as áreas envolvidas. Estes autores chamam ainda a atenção para a situação paradoxal decorrente do facto de quanto maior é a especialização científica e tecnológica, maior é a sua interdisciplinaridade. Alsina *et al.* (1996: 85) lembram, contudo, que nem todas as disciplinas relacionadas com a Terminologia utilizam os termos da mesma forma e com os mesmos objectivos. Estas autoras apontam a Linguística e a Ciência Computacional como exemplos destacados de

entre as várias disciplinas utilizadoras das funções da Terminologia (representação, transferência e investigação).

Por um lado, a Terminologia é objecto de estudo da Linguística e concretamente da Lexicologia pelo facto de os termos serem unidades lexicais e, por outro lado, a Linguística (Lexicologia) ser indispensável para a fundamentação teórica da Terminologia. A este propósito e de entre as várias conclusões apresentadas no sub-capítulo “Research Issues In Terminology”⁵ do *Pointer Final Report* destaca-se a seguinte ideia:

At a more general level, terminology research can contribute to research in language and research in aspects of knowledge growth and dissemination. Terminology research can contribute to and benefit from research in linguistic [sic], in studies of writing and innovation, in information retrieval; open questions in semantics and pragmatics can be focused by working together with well-organised terminology databases, for instance. Knowledge growth and dissemination can benefit from research in neologisms, in changes in terminology usage, and indeed, neologism growth can be used as a measure in scientometry.

- “Research Issues In Terminology”, in *Pointer Final Report*

Quanto à relação entre a Terminologia e a Ciência Computacional, ela é recíproca já que a primeira se serve da Ciência Computacional para acelerar, facilitar e melhorar algumas das suas tarefas, enquanto a segunda se serve das estruturas conceptuais elaboradas pela Terminologia para avançar no campo da Inteligência Artificial. Esta ligação tem vindo a intensificar-se trazendo ferramentas inovadoras que permitem novos métodos e abordagens ao trabalho terminológico. A colaboração é cada vez mais expressiva, sobretudo nas etapas referentes à construção de *corpora* e extracção automática de unidades terminológicas que são, porventura, as tarefas mais demoradas e esforçadas. Felizmente esse esforço tem sido progressivamente minimizado graças ao aparecimento de ferramentas informáticas cada vez mais sofisticadas e adequadas às solicitações da Terminologia. Nessa perspectiva, o *Corpógrafo* corresponde a um progresso importante para a detecção e extracção automática e semi-automática de termos e outros dados, devidamente contextualizados, a partir de textos reais e representativos de uma área específica do conhecimento.

⁵ Disponível em: <http://www.computing.surrey.ac.uk/ai/pointer/report/section5.html> [acedido em 2005.10.15]

1.1.2.2. Teorias e principais postulados da disciplina: da Teoria às Teorias

A Terminologia surgiu com a necessidade de organizar e, sobretudo, padronizar conceitos através de termos, eliminando, desta forma, as ambiguidades que proliferavam na comunicação de especialidade. Havia uma necessidade imperiosa, por parte das comunidades científicas e técnicas, de garantir que quando se referissem a um determinado conceito era apenas esse o conceito que denominavam. Este binómio de “um conceito, um termo”, designado por monorreferencialidade, é de vital importância intra- e inter-comunidades científicas e técnicas.

Desde que a Terminologia foi constituída como disciplina, a sua evolução tem sido constante. De acordo com Auger (*in* Cabré, 1999: 5) a história da Terminologia moderna subdivide-se basicamente em quatro períodos:

- a génese (1930-1960), caracterizada pela delineação metodológica a aplicar na formação sistemática dos termos;
- a estruturação do domínio (1960-1975), correspondente ao desenvolvimento da Ciência Computacional (surgem as primeiras bases de dados digitais) e das técnicas de documentação permitindo inovações de grande relevo na área da Terminologia;
- a explosão (1975-1985), condicente com o período de planeamento linguístico e projectos terminológicos;
- a expansão (de 1985-ao presente), correspondente a uma multiplicidade de iniciativas ligadas não só ao planeamento linguístico alargado à escala mundial (em larga parte devido à globalização), mas também a uma extraordinária demanda no campo da indústria da língua para vários fins, ao que a Terminologia tem sabido corresponder sobretudo devido ao contributo inestimável da Ciência Computacional.

Em nosso entender, existem basicamente duas grandes perspectivas da terminologia: a prescritiva e a descritiva (na qual se inclui a abordagem variacionista), sobre as quais apresentamos em seguida algumas considerações.

1.1.2.2.1. A Teoria Geral da Terminologia

A sistematização da teoria terminológica, de carácter assumidamente prescritivo, surge essencialmente na primeira metade do Século XX fruto de um conjunto de circunstâncias de onde sobressai a inexistência de uma comunicação inequívoca entre especialistas de determinadas áreas técnicas e científicas. A primeira perspectiva terminológica corresponde, assim, à normalização prescritiva da terminologia técnica. Esta sistematização encontra-se focalizada e identificada essencialmente em três escolas, cada uma com a sua própria abordagem: a de Viena (ou austríaca) considerava a Terminologia como uma área interdisciplinar, embora autónoma ao serviço de disciplinas técnicas e científicas; a de Moscovo (ou soviética), focada na filosofia, estava sobretudo interessada na classificação lógica de sistemas conceptuais e na organização do conhecimento; e a de Praga (ou checa), centrada na Linguística, considerava a Terminologia um sub-elemento do léxico de uma língua e as línguas de especialidade como sub-sistemas da língua geral (Cabré, 1999: 7).

Estas escolas tradicionais têm em comum a convicção da necessidade da organização conceptual dos domínios do conhecimento, bem como a normalização terminológica de forma a melhorar a comunicação especializada.

Das três Escolas citadas, a que, porventura, adquiriu maior visibilidade foi a de Viena, devido em grande parte ao trabalho desenvolvido por Eugen Wüster, geralmente apontado como o primeiro a empregar e consolidar métodos e práticas concretas promovendo a Terminologia a uma disciplina autónoma, alertando ao mesmo tempo a sociedade para a sua importância. Este engenheiro e industrial austríaco, propunha a criação de uma ordem puramente metodológica e normativa elegendo a terminologia como um instrumento que visava a eliminação das ambiguidades nas comunicações científicas e técnicas (Rondeau, 1984: 6).

Embora um pouco menos divulgado, mas igualmente importante, é o trabalho teórico desenvolvido na mesma época pelo russo D.S. Lotte apontado como o primeiro verdadeiro professor de terminologia. A sua preocupação centrava-se na atribuição de um carácter mais racional aos trabalhos terminológicos (Rondeau, 1984: 6).

A complementaridade do trabalho desenvolvido por estas duas correntes trouxe à Terminologia enormes benefícios projectando-a ao nível internacional. Em 1936 foi criada a ISA (*International Federation of National Standardizing Associations*), com o objectivo claro de favorecer as relações comerciais em franco desenvolvimento e que, depois da Segunda Guerra Mundial, é reformulada recebendo a designação de

Organização Internacional de Normalização, mais conhecida por ISO (*International Standardizing Organization*). Ainda hoje permanece um organismo internacional orientador das normas de numerosos países (Mateus *et al.* 1998:14). A ISO é composta por vários Comitês Técnicos, entre os quais destacamos o Comité Técnico 37 (TC37), designado por “Terminologia e Coordenação de Metodologias” até ao ano de 2002 e desde essa data denominado por “Terminologia e Outros Recursos”. Este Comité tem como principal função a normalização terminológica. Reuniu pela primeira vez em 1952 (presidido por Wüster), tendo publicado as suas primeiras normas sobre terminologia em 1968 (Costa, 2003: s/ página), na altura designadas como *recomendações*.

A Teoria Geral da Terminologia (TGT) traduz os fundamentos da teoria terminológica wüsteriana designando-a como uma actividade focalizada nos conceitos e respectiva sistematização, na sua ligação aos termos, bem como na normalização e prescrição dos termos técnico-científicos. De acordo com esta teoria, antes da atribuição de um termo a um conceito é necessário enumerar a totalidade dos atributos que formam o conceito. Os conceitos necessitam não só ser nomeados, como também definidos. Tal permite aos seus utilizadores compreendê-los em toda a sua amplitude e empregá-los correctamente, isto é, de um modo conciso e desambiguado.

O ponto de vista wüsteriano é, desta forma, bastante claro: a Terminologia tem um carácter normativo e prescritivo, já que os termos são primeiramente aceites, ao que se segue a recomendação do seu uso, afastando o aspecto comunicativo da linguagem. A relativa inflexibilidade dos princípios basilares da TGT de univocidade e monorreferencialidade das unidades terminológicas (nem sempre passíveis de aplicação no plano real) contribuiu para que, sobretudo a partir da década de 80, surgissem novas vozes especialmente críticas da TGT. No essencial, questionava-se a incompatibilidade dos princípios da TGT com a realidade contemporânea ao nível da relação do conceito e da sua expressão linguística. L’Homme *et alii* (2003: 153) descrevem alguns dos principais problemas que a TGT oferece, entre os quais se destacam:

- a definição de “domínio” é arbitrária e referente a objectivos definidos no âmbito do projecto metodológico;
- os conceitos aos quais os termos se referem não constituem elementos bem definidos;
- os conceitos são independentes da língua geral;
- as descrições raramente são efectuadas do ponto de vista onomasiológico;

- a univocidade: a um termo corresponde um conceito
- os termos são independentes do contexto.

Aquando do surgimento da terminologia como área de estudo, a norma de imposição wüsteriana, de certa forma simplista, de um termo corresponder a um conceito e de um conceito ser igual em qualquer parte do mundo, não tem correspondência com o mundo real. É hoje aceite que o aspecto cultural é demasiado importante para ser afastado, quer na organização dos sistemas conceptuais, quer no processo de denominação desses conceitos.

1.1.2.2.2. As teorias contemporâneas da Terminologia

A segunda perspectiva terminológica equivale à abordagem descritiva dos usos e discursos reais fornecidos através do acesso a documentos, *corpora*, etc., opção que permite encarar as línguas de especialidade no seu pleno, com todas as suas realizações variacionistas. Em traços muito gerais, fazer terminologia hoje significa aceitar uma das características fundamentais de qualquer língua viva: a variação (sempre presente, mas só recentemente reconhecida). A ideia de que uma língua não é um todo homogêneo e uniforme, mas um sistema plural, constituído por normas que evidenciam os usos reais em variação é manifestada por diversos estudiosos da língua, entre os quais Vilela (1995), que refere:

(...) o léxico de uma língua, sem ser uma manta de retalhos, não é um todo homogêneo: constitui o que costumamos designar por diassistema: as palavras de todos os dias convivem com as palavras dos especialistas, as palavras da língua falada (ou estilo coloquial) vivem lado a lado com as palavras da língua escrita (ou estilo reflectido), as palavras "velhas", ainda de uso corrente, coabitam com arcaísmos e neologismos. (...) Há também o chamado "jargão" das variedades funcionais-contextuais: línguas sectoriais de especialidade, como a linguagem do desporto, da publicidade, da informática, dos "media", etc. Há variedades geográficas e sociais. (...) Não há um limite fixo e claro entre as variedades de língua dentro de uma mesma língua histórica.

Vilela (1995: 16)

Fazer terminologia significa, também, aceitar o estatuto linguístico das denominações dos termos; encarar essas unidades lexicais como parte da significação dos termos em situação de discurso e em contexto; identificar e categorizar as variantes

terminológicas; analisar essa variação tendo em conta o contexto de produção / uso de termos e ter em conta, para além da abordagem onomasiológica, a semasiológica, sobretudo quando existe a oportunidade de trabalhar com *corpora* (Conceição, 2003: s/ página).

Reforçando um pouco a importância da variação propomo-nos exemplificar três situações nas quais ela pode ocorrer:

a) no mesmo nível da língua

- *cheia rápida*
- *cheia repentina*
- *cheia súbita*

b) entre diferentes níveis da língua

- caudal (discurso de especialidade):

“Volume de água numa secção transversal de um rio ou de um canal, por unidade de tempo.”⁶

- caudal (discurso geral):

“Volume de água que flui num rio (ou canal) numa determinada secção deste, por unidade de tempo.”⁷

c) entre diferentes sistemas linguísticos

- cheia (português de Portugal)
- enchente (português do Brasil)

Entretanto, com o objectivo de resolver os problemas criados pela aplicação das orientações metodológicas da TGT, têm sido apresentadas propostas de abordagens alternativas. L’Homme *et alii* (2003: 154) sistematizam destacando quatro:

- a Socioterminologia que se traduz no estudo das unidades terminológicas numa perspectiva social, isto é, defende a necessidade de atender à dimensão social dos termos, uma vez que estes estão sujeitos a variações consoante o contexto social em que são utilizados;

⁶ In “Glossário Internacional de Hidrologia”, disponível em: <http://www.cig.ensmp.fr/~hubert/glu/PT/GF0339PT.HTM>, [acedido em 2005.08.20].

⁷ In Glossário – Prtoteccção Civil da Região Autónoma dos Açores <http://www.srpca.pt/glossario.htm>, [acedido em 2005.08.20].

- a Terminologia Textual que, baseando-se na Linguística de *corpus*, advoga ser necessário uma melhor compreensão do funcionamento dos termos em sede de texto;
- a Teoria Comunicativa de Terminologia que enfatiza a dimensão comunicativa dos termos em conjunto com os seus aspectos cognitivo e linguístico;
- a Terminologia Sociocognitiva que sustenta a necessidade da definição de conceito baseada numa teoria prototípica, com a inclusão de uma perspectiva diacrónica e social na descrição de termos.

De uma forma muito breve, descrevemos, em seguida, cada uma destas quatro abordagens teóricas da disciplina terminológica.

A Socioterminologia

A abordagem socioterminológica estuda o termo, na sua perspectiva linguística, integrado no seu contexto social, ao invés da terminologia clássica, de carácter normativo e prescritivo. A marginalização do aspecto social, por parte da TGT, tem vindo a ser sustentadamente criticada pelos seguidores das mais recentes abordagens metodológicas descritivas, entre as quais se situa a Socioterminologia, surgida nos anos 80 e sistematizada a partir da década de 90. Com efeito, em oposição à Terminologia prescritiva, impunha-se a descrição da realidade e, ao mesmo tempo, a extinção de uma terminologia refém dos desejos dos “terminólogos” pioneiros que não aceitavam os termos como unidades linguísticas da língua natural, mas como elementos vocacionados para a normalização de um dado léxico.

A Socioterminologia entende a variação como uma característica intrínseca da língua que tanto pode ser observada sincrónica como diacronicamente. Se por um lado, um termo pode ter variações (motivadas por diferenças de cariz cultural, social, individual, educacional ou geográfico) em relação ao nível da língua quando utilizado por comunidades linguísticas distintas, sem que haja necessariamente prejuízo para o rigor científico e qualidade linguística. Por outro lado, um termo pode sofrer alterações formais e/ou conceptuais ao longo do seu tempo útil de vida.

Em relação a esta abordagem variacionista e numa perspectiva assumidamente socioterminológica, Conceição (2003, s/ página) identifica sete tipos de variação: gráfica, denominativa (sistémica e discursiva), conceptual, diatópica (espaço

geográfico), distrática (níveis de especialização, camadas socioculturais), diafásica (oralidade, escrita e outros tipos de modalidade expressiva) e diacrónica (espaço temporal). Este autor acrescenta, ainda, que pode dar-se o caso de “*não haver uma variante formal e haver uma activação incompleta de traços conceptuais consoante o tipo de discurso em que se atesta a ocorrência*” (Conceição, 2003: s/ página). Tenhamos como exemplo o termo *Riscos Naturais*. Um geógrafo e um leigo na matéria não activam da mesma forma o nível conceptual desta unidade terminológica. Este fenómeno é igualmente verificável em várias fases do processo terminológico, como são os casos de construção do texto da definição e da selecção e delimitação dos contextos associados a essa definição, mesmo que se mantenham idênticos todos os outros pressupostos, como o utilizador final e a apresentação final do produto. De facto, o grau de tecnicidade e cientificidade é muito mais elevado num glossário elaborado para especialistas, do que para o público geral. Vejamos os seguintes exemplos:

- Definições de “Cheia” propostas pelo *Glossário Internacional de Hidrologia*⁸:

PT 0460 **Cheia** *ver também* Inundação ; Maré alta ;

- 1) Subida, geralmente rápida, do nível de um curso de água até um máximo a partir do qual desce em geral mais lentamente.
- 2) Fluxo relativamente elevado medido como nível ou caudal.
- 3) Maré ascendente.

PT 0514 **Cheia**

Pequena cheia causada por precipitação intensa ou degelo.

- Definição de “Cheia” proposta pelo *Serviço Nacional de Bombeiros e Protecção Civil*⁹:

⁸ Disponível em: <http://www.cig.ensmp.fr/~hubert/glu/HINDPT.HTM> [acedido em 2005.10.15].

⁹ Disponível em:

http://www2.snbp.pt/portal/page?_pageid=35,46761,35_46771:36_35018&_dad=portal&_schema=PORTAL [acedido em 2005.10.15].

O que é uma Cheia

As cheias são fenómenos naturais extremos e temporários, provocados por precipitações moderadas e permanentes ou por precipitações repentinas e de elevada intensidade. Este excesso de precipitação faz aumentar o caudal dos cursos de água, originando o extravase do leito normal e a inundação das margens e áreas circunvizinhas. Nalgumas partes do globo as cheias podem dever-se também ao derretimento de calotes de gelo.

As cheias podem ainda ser causadas pela rotura de barragens, associadas ou não a fenómenos meteorológicos adversos. As cheias induzidas por estes acidentes são geralmente de propagação muito rápida.

De uma forma sucinta, Gaudin (2003: 11), um dos fundadores da Socioterminologia, define-a apresentando as características dos seus métodos de análise e a descrição das unidades terminológicas, tendo em conta as suas variações contextuais aos níveis terminológico, linguístico e social. Descreve-a, ainda, como uma disciplina encarregue do estudo dos termos, isto é, “(...) *des vocables servant à véhiculer des significations socialement réglées et insérées dans des pratiques institutionnelles ou des corps de connaissances.*” (Gaudin, 2003: 11); explica, também, por que motivo se operou a integração das variações (ligadas ao funcionamento de situações comunicativas) no estudo descritivo da Terminologia, funcionando em oposto aos pressupostos de monossemia e univocidade da terminologia clássica. O estudo sincrónico dos termos tem a ver com a circulação dos conhecimentos, enquanto o estudo diacrónico se refere à história das ciências, das técnicas, dos discursos socialmente fixados e, mais concretamente à história das ideias (Gaudin, 2003: 11).

A Terminologia Textual

A Terminologia Textual é, no entender de Berland (2000: 3) e Berland *et al.* (2002: 1), uma abordagem descritiva assente na análise de textos especializados (produzidos no quadro de uma prática), reunidos em *corpora* textuais, para aí pesquisar as expressões linguísticas referentes ao domínio especializado. Esta abordagem fixa a terminologia numa vertente linguística essencialmente vocacionada para a criação de recursos terminológicos, tais como, glossários, terminologias e bases de dados terminológicas. A informação textual desempenha, desta forma, um papel decisivo na aquisição terminológica a partir de *corpora*. Em total harmonia com esta abordagem encontra-se a Terminologia Textual desenvolvida por Hoffmann (1998a,b, *in*, Finatto *et alii.*, 2004: 5), que perspectiva o estudo das línguas de especialidade (técnicas e científicas) centrado no texto e não apenas nos termos. A Terminologia Textual

promove, desta forma, a associação entre o aspecto cognitivo e o aspecto linguístico inseridos no funcionamento do texto especializado.

A Terminologia Textual é sintetizada por Berland (2000: 1) em três pontos essenciais:

- *“la connaissance sur un domaine d’activité doit être cherchée dans les textes produits dans le cadre de cette activité, où elle est exprimée par des expressions linguistiques;”*
- *“il n’existe pas une terminologie pour un domaine d’activité, mais autant de terminologies que d’applications liées à ce domaine;”*
- *“un terme n’existe pas a priori, il est construit et résulte d’un choix.”*, ou seja, a construção de uma terminologia repousa na colaboração entre o terminólogo, o especialista do domínio e o utilizador da aplicação final.

A Teoria Comunicativa da Terminologia

Para a Teoria Comunicativa da Terminologia (TCT) a terminologia é essencialmente concebida como um conjunto de necessidades, de práticas e de aplicações, bem como um campo de conhecimentos (mais ou menos consolidado). A Terminologia não é vista como uma ciência pois, não se distingue como tal através do seu método podendo, contudo, ser estudada através de um método científico.

Tal como o nome indica, a TCT estrutura-se na vertente comunicativa da linguagem de especialidade, relegando para segundo plano a sua função normalizadora. Utilizando uma imagem, Cabré (2002: 186-7) explica o seu modelo de concepção terminológico, que denominou como *Teoria das Portas*. Este modelo conciliador de vários outros modelos tem como objecto central a unidade terminológica, que Cabré concebe como estando dentro de uma casa e que pode ser alcançado (de forma não simultânea) através das várias portas representativas das diversas perspectivas de visionamento do objecto: a cognitiva (o conceito), a linguística (o termo) e a comunicativa (a situação). Para a escolha de uma determinada porta existem apenas duas condições: a teoria escolhida deverá estar em consonância com a porta escolhida e a teoria não pode contrariar a multidimensionalidade do objecto. A unidade terminológica, como objecto multidimensional, tem de ter em linha de conta as variações nos planos conceptual, linguístico e comunicativo passíveis de ocorrerem e serem representados nos diversos níveis de especialização e de representação.

A Teoria Sociocognitiva

A Terminologia Sociocognitiva proposta por Temmermann (2000) baseia-se num modelo de análise de conceitos e em estudos empíricos efectuados em contextos linguísticos de uso real de comunidades científicas, nomeadamente das Ciências da Vida. Esta teoria está alicerçada em cinco princípios (Termmermann, 2000: 223-8):

- o seu ponto de partida são as unidades de conhecimento (em substituição dos conceitos), que só raramente não apresentam estrutura de protótipo;
- a compreensão é um evento estruturado. A unidade de conhecimento tem uma estrutura intra- e intercategorial e funciona em modelos cognitivos;
- dependendo do tipo de unidade de conhecimento e do nível e tipo de especialização do emissor e do receptor da comunicação, o que é informação mais ou menos essencial será variável;
- tanto a sinonímia como a polissemia assumem um papel funcional na evolução do conhecimento e, como tal, necessitam ser descritos;
- as unidades de conhecimento estão em constante mudança. Os períodos históricos da sua evolução podem ser mais ou menos essenciais para a compreensão de uma unidade. Por outro lado, os modelos cognitivos desempenham um papel no desenvolvimento de novas ideias, o que implica que os termos sejam motivados.

A Teoria Sociocognitiva sistematiza, assim, a representação da realidade social, apontando os termos como designadores de unidades de conhecimento. Contudo, embora as unidades possam ser concebidas (mentalmente), apenas algumas podem ser percebidas (na realidade).

1.1.3. Termo

Depois de definida e caracterizada a terminologia e as várias propostas para novas teorias terminológicas mais recentes em oposição à teoria clássica, propomos uma breve definição e caracterização do objecto de estudo comum a todas elas: o termo.

1.1.3.1. Definição e caracterização

Como resultado da existência de diferentes abordagens metodológicas da terminologia, surgem com alguma naturalidade diferentes definições de *termo* salientando-se, desde já, a enunciação apresentada pelo *Manual de Terminologia* que define *termo*, ou *unidade terminológica*, como “*palavra (termo simples), grupo de palavras (termo composto), sintagma, símbolo ou fórmula que designam um conceito de uma área específica.*” (Pavel *et al.*, 2002: 131).

Verificamos, deste modo, a existência de unidades terminológicas monolexémicas (*cheias* e *inundação*) e unidades terminológicas multilexémicas (*cheias repentinas* e *caudal de cheia*), sendo que, em terminologia, são mais frequentes os termos multilexémicos do que os monolexémicos (Kageura *et alii*, 2004: 8). Para tal, e a título meramente exemplificativo, refira-se que a grande maioria dos termos compostos recorrem a combinações lexicais, sendo as que a seguir mencionamos as mais comuns na língua portuguesa: nome+preposição+nome (*onda de cheia*), nome+adjectivo (*inundação urbana*), nome+preposição+advérbio (*força de arrastamento*). Uma outra constatação é a de que quanto mais extensa é a sequência lexical de um termo, maior é a sua tendência para a monossemia e monorreferencialidade (*sistema hídrico* em oposição a *sistema de gestão de recursos hídricos*).

Os termos são, desta forma, representações linguísticas que sintetizam traços conceptuais ou, por outras palavras, são denominações de conceitos. Este é, também, o entendimento de Cabré (1998: 81) no sentido em que esta autora refere que os termos reflectem a estruturação conceptual de uma matéria e, como tal, são a base da comunicação especializada. Refere, ainda, que a terminologia, para além de organizar o pensamento, serve de veículo para a transferência de conhecimentos, por parte dos especialistas, sobre uma determinada matéria, numa ou em várias línguas e para estruturação da informação em textos especializados. Esta relação existente entre termo e conceito é passível de ser observada em três situações distintas: um termo para um conceito (*hidrograma de cheia*); vários termos para um conceito (*planície aluvionar*, *planície de aluvião* e *planície de inundação*); e, por fim, um termo que comporta vários conceitos (*risco*, *cheia*, *leito* - homonímia).

Conforme já citámos anteriormente, quando nos referimos à interdisciplinaridade da terminologia, as unidades terminológicas caracterizam-se, também, pela sua multidimensionalidade, uma vez que possuem outros aspectos caracterizadores para além do seu valor intrínseco, sendo distinguidos, por Sager (1990:

14-128), l' Homme *et alii*. (2003: 156-7) e Cabré (1999: 25-55), como unidades linguísticas, cognitivas e comunicativas, conforme a seguir descrito:

- **a dimensão linguística:** “(...) *examines the existing and potential forms of the representation of terminologies;*” (Sager, 1990: 14), ou seja, o aspecto linguístico insere a unidade terminológica num contexto de discurso de especialidade, deixando de a considerar como um elemento independente da língua, como é apanágio da Teoria Geral da Terminologia. Esta dimensão inclui, desta forma, o estudo das características (morfológicas, gramaticais, etc.), sendo que o desempenho dos termos é alvo de estudo por parte de uma terminologia de cariz mais descritivo;
- **a dimensão cognitiva:** “(...) *relates the linguistic forms to their conceptual content, i.e. the referents in the real world;*” (Sager, 1990: 14). O aspecto cognitivo de uma unidade terminológica refere-se não só ao significado, mas também ao tipo de relações conceptuais que esse dado termo tem com outros da mesma área de especialidade. De acordo com l' Homme *et alii* (2003: 156-7), para além das relações conceptuais hierárquicas (tais como genérico/específico ou todo/parte), tem-se verificado um crescente interesse no estudo das relações conceptuais não-hierárquicas (por exemplo, de causa/efeito e objecto/função). Por outro lado, as designações *relações conceptuais* e *sistemas de representação do conhecimento* têm sido preteridas por alguns autores, em favor de designações como *redes* ou *relações semânticas* (l' Homme *et alii*, 2003: 156-7);
- **a dimensão comunicativa:** “(...) *looks at the use of terminologies and has to justify the human activity of terminology compilation and processing.*” (Sager, 1990: 14). É nesta dimensão que tomamos consciência da importância do elemento social do uso dos termos. Na realidade, os termos podem ocorrer em diversas situações comunicativas, mas sempre e apenas em áreas de especialidade pois, são elementos linguísticos que se encontram devidamente contextualizados.

Um outro aspecto a ressaltar é a perspectivação do termo conforme a teoria terminológica ou abordagem metodológica adoptada. É, assim, possível abordá-los de forma semasiológica ou onomasiológica. No primeiro caso, através da análise de comportamentos discursivos, onde se incluem trabalhos de carácter mais terminográfico

e, no segundo, através da decisão sobre a denominação mais correcta, onde se incluem os aspectos de criação, fixação e harmonização terminológica (Conceição, 2003: s/página).

1.1.3.2. Distinção entre termo e não-termo

Ao contrário da visão wüsteriana, que postula que o termo não faz parte da língua geral, a tendência actual é considerá-lo como uma unidade lexical reveladora de um significado específico. Cabré (1995: 8), por exemplo, descreve o termo como unidade cujo significado especializado é activado sob circunstâncias comunicativas específicas. Esta “promoção” de uma palavra a termo, ou esta redesignação é sobretudo útil quando estamos na presença de polissemias, bastante frequentes nas Ciências Sociais. Esta descrição vai aliás ao encontro da característica da multidimensionalidade que Cabré (1995: 8) atribui ao termo quando o compara com um poliedro que pode ser observado de diferentes ângulos conforme o ponto de vista adoptado no momento. Esta autora (1995: 3), advoga, também, que num enquadramento linguístico, os termos correspondem a conjuntos de signos linguísticos que constituem um sub-conjunto dentro do léxico de uma língua, isto é, o termo equipara-se à palavra.

Contudo, a sua principal diferença é a de que a palavra se refere à língua geral, enquanto o termo se refere a uma língua de especialidade (L’Homme *et alii*, 2003: 154). Uma palavra tem propriedades, mas tem também muitos significados, porquanto é um elemento do léxico. Um termo, pelo contrário, é uma palavra contextualizada num discurso tendo, conseqüentemente, um referente de interpretação. Esta afirmação não implica, no entanto, que todas as unidades terminológicas sejam absolutamente dependentes do contexto em que são enunciadas. Tome-se como exemplo o termo *bacia hidrográfica* que não necessita de contextualização específica pois, para além de ser, actualmente, uma unidade terminológica com elevado grau de utilização na língua geral, tendo experimentado um processo de vulgarização. É um termo que ao ser utilizado por diversas áreas científicas como a Hidrologia e Hidrografia, Geografia Física, Engenharia Ambiental, entre muitas outras, não modifica a sua definição e, sobretudo, o seu todo conceptual.

As questões que se colocam são as de como distinguir um termo num texto de especialidade e quais os critérios a utilizar para atingir essa distinção?

A resposta à primeira questão é teoricamente simples de dar. Seja do ponto de vista qualitativo ou do quantitativo, os termos técnicos e científicos têm características

distintivas que permitem a sua identificação. Estas características distintivas geram, aliás, graus de tecnicidade e de cientificidade dos termos em questão, dependendo da relação de proximidade de um termo com determinado tema (Chung, *in* Nation, 2001: 198-200). De uma forma sucinta e com base em diversos estudos de teóricos e investigadores nesta área, Chung (2003: 221-2) lista algumas dessas características. Do ponto de vista qualitativo, alguns termos têm um significado especializado, possuindo uma representação formal distinta (geralmente oriundas do latim e do grego) e são desconhecidos ou pouco conhecidos do discurso geral, bem como de outros discursos de especialidade (*Hidrologia, hidrograma e Potamologia*). Existem, ainda, outros termos que, devido à polissemia, podem ter diferentes acepções em diferentes áreas (*risco – natural, tecnológico, de saúde*, etc.). Do ponto de vista quantitativo, alguns termos têm um significado especializado, bem como outros significados (nem sempre relacionados) e são conhecidos tanto do discurso geral como do discurso de especialidade (*cheia e leito*), mas ocorrem com maior frequência num domínio de especialidade do que no uso geral. Podem, aliás, ocorrer com maior frequência num tópico de um texto ou num texto relacionado com um domínio de especialidade.

Contudo, a questão que muitas vezes se coloca e para a qual Walinski (2003: 78) chama a atenção é a de, por vezes, não ser possível delinear claramente a fronteira entre *termo* e *não-termo*. Tal poderá suceder devido ao facto de o significado de uma determinada unidade lexical em relação a um determinado conceito ser resultado de um princípio assumido pelo autor do texto.

Com efeito, é por vezes difícil distinguir léxico de uso geral de unidades terminológicas pois, tal como refere Bessé (*in* Rey, 1995: 3), existe um grande número de unidades lexicais que são simultaneamente palavras e termos, conforme já distinguimos e exemplificámos anteriormente. Os termos que possuem esta duplicidade de atributos são, porventura, os que geram mais confusão pois, o seu significado assenta no sentido corrente da língua de uso geral. Face a situações como esta, é imprescindível que os conceitos subjacentes sejam compreendidos em toda a sua extensão e que os mesmos sejam definidos de forma clara e precisa para que reflectam desambiguada e adequadamente o seu conteúdo. No caso da existência de várias alternativas de escolha, a decisão deve pautar-se pelas características da abordagem descritiva, ou seja, precisão, aceitabilidade e uso geral.

Tendo presente as teorias contemporâneas citadas no ponto 1.1.2.2.2., os termos ou unidades terminológicas são objecto de linguagem, de cognição e de comunicação. A

terminologia está, desta forma, presente sempre que estão presentes especialistas em determinado domínio. Existe, aliás, uma inter-relação constante entre terminólogo (não-especialista) e especialista (não-linguista). O primeiro contribuirá, principalmente, com o seu saber terminológico e linguístico, enquanto o segundo, sobretudo, com o seu saber conceptual ou nocional. Uma das etapas fundamentais de um projecto terminológico é a extracção de candidatos a termos a partir de textos de um domínio especializado previamente reunidos num *corpus*. E é precisamente perante a situação de decidir entre termo e não-termo que o terminólogo hesita, sobretudo se não for especialista na área de conhecimento em causa.

No caso de projectos terminológicos, cuja equipa é dirigida por um terminólogo, a mesma deverá integrar pelo menos um especialista na área em estudo. Nesta situação caberá ao especialista sugerir as fontes mais relevantes e representativas que servem de base à constituição de um *corpus* textual, assim como avaliar, na lista de candidatos a termo elaboradas pelo terminólogo, os termos propostos, validando-os ou excluindo-os.

Uma vez que os termos apresentam características comuns aos não-termos, quer ao nível lexical, morfológico ou sintáctico, a tarefa de distinção entre ambos é à primeira vista um pouco complicada. No entanto, se atentarmos à proposta de Castillo (1998: 99-101) sobre os cinco factores pragmáticos que este autor considera como os mais importantes para diferenciar o vocabulário de especialidade do léxico comum, esta tarefa torna-se mais acessível. São eles:

- a função básica proposta pelos termos, que é essencialmente (mas não exclusivamente) a referencial, uma vez que denominam conceitos de uma área especializada de conhecimento;
- a temática tratada, uma vez que as unidades terminológicas só adquirem esse estatuto quando fazem parte de uma área de especialidade (ciência, técnica, qualquer actividade delimitada). Em contrapartida, a palavra está presente em todo o léxico utilizado pelo seu utilizador, incluindo o de especialidade;
- os utilizadores do léxico de especialidade são os especialistas da área tratada, embora nos dias de hoje o número de utilizadores se diversifique e cresça consideravelmente, sobretudo devido à divulgação e vulgarização de certas áreas do conhecimento;
- a situação comunicativa, já que em princípio o vocabulário de especialidade é utilizado em situações de comunicação formal e não em situações de maior informalidade, nem em textos que não sejam propriamente referenciais.

Porém, esta situação nem sempre se verifica, já que, por vezes, os especialistas utilizam o vocabulário de especialidade em situações comunicativas com diferentes níveis de abstracção e formalização;

- os discursos, uma vez que os tipos de textos onde surgem unidades terminológicas são “(...) *fundamentalmente informativos: descriptivos, explicativos y argumentativos sobre todo, e, en un grado más bajo, directivos e evaluativos.*” (Castillo, 1998: 101).

Os critérios propostos para a distinção entre termo e não-termo num texto de especialidade e, conseqüente, para a selecção do termo resumem-se essencialmente a dois: o estatístico e o linguístico (semântico). Para além de uma boa selecção de textos da especialidade em estudo, que garantam uma presença considerável de termos (em quantidade e diversidade), o critério semântico, que na prática se traduz na selecção de termos e conceitos relevantes de uma determinada área conceptual, socorre-se de um conjunto de indicadores linguísticos e textuais que ajudam neste processo de distinção, entre os quais se destacam: o formal, o contextual e as fórmulas e combinatórias lexicais reveladoras de relações conceptuais.

Quanto ao indicador formal, que opera ao nível morfológico, constata-se a presença de morfemas de origem grega e latina (conforme já abordado na caracterização de termos), de que é exemplo, *Potamologia*, do grego *potamós*, rio + *lógos*, tratado; assim como de derivações morfológicas pois, conforme observado por l’Homme (2002: 3), “*Terms are related to other lexical units by formal similarity.*” E de facto, tal afirmação é corroborada por dois dos exemplos de termos mais significativos dos *corpora* constituídos para este estudo: [PT] *inundação, inundar, inundado*; [EN] *flood, flooding, to flood, flooded*. Ou seja, a partir do instante em que se valida um destes elementos, podemos deduzir que as outras também são designadas como termos. Contudo, a maioria dos termos originários da língua geral, não possuem marcas formais que facilitam a sua recolha em situação de textos especializados, como por exemplo, *marcas de cheia, superfície inundada e dique natural*. Como já foi referido, uma das maiores dificuldades na distinção entre termo e não-termo reside no facto de alguns termos simples serem igualmente utilizados no léxico geral. O mesmo se passa em relação aos termos compostos em oposição, por exemplo, às combinatórias lexicais. Uma das tarefas mais delicadas no trabalho de selecção terminológica será, porventura,

a de identificar a dimensão de um termo constituído por mais do que uma unidade, isto é, o termo composto.

O segundo indicador relaciona-se com a contextualização do termo ou da palavra. Em caso de dúvida, a apresentação do contexto é, na esmagadora maioria dos casos, de importância primordial para a percepção correcta do conceito que está por detrás, sendo decisivo para a validação ou rejeição como termo. Os estudos mais recentes de terminologia indicam que há hoje em dia plena consciência de que um termo está integrado num contexto, num esquema de conhecimento. Fora dele, deixa de ser termo; é esvaziado do seu significado especializado. Por exemplo, *risco* não é termo, mas *risco natural* já o é, ou seja, é o valor pragmático, o uso, que determina o carácter terminológico de um lexema. A autenticação de um termo está dependente do seu contexto, pelo que é uma inevitabilidade descrevê-lo com as especificidades linguísticas directamente observadas nesse contexto. Esta postura descritiva é oposta à prescritiva, de acordo com a qual o uso do termo seria imposto. Nesse sentido, o princípio metodológico enunciado por Stubbs (1996: 28) que “(...) *Language should be studied in actual, attested, authentic instances of use, not as intuitive, invented, isolated sentences.*”, reflecte a tendência actual não só para o caso do estudo da língua geral, mas também das línguas de especialidade.

Podemos concluir que a contextualização é uma das ajudas possíveis à qual o terminólogo não-especialista em uma dada área de conhecimento se pode socorrer para alcançar a definição final. Usualmente um termo *per si* não é suficiente para garantir a total compreensão de um conceito. Será necessário apreende-lo através do seu contexto significativo. Neste cenário, a possibilidade que o *Corpógrafo* oferece ao seu utilizador de aceder ao contexto original de uso onde o termo está inserido permite verificar a situação semântico-sintática em que este pode ou deve ser utilizado.

Relacionado com a análise contextual, apresenta-se um terceiro indicador relevante – as fórmulas reveladoras de relações conceptuais. Existem fórmulas e combinatórias lexicais, cujo significado permite estabelecer ligações semânticas, conceptuais e ontológicas entre as várias unidades terminológicas presentes num discurso ou num texto de uma determinada área de especialidade. Os indicadores dos tipos de relações semânticas mais comuns (hiponímia, co-hiponímia, hiperonímia, meronímia, sinonímia e antonímia) funcionam particularmente bem nos domínios das Ciências Vivas e das Ciências Exactas, onde a estruturação e a hierarquização dos elementos que as compõem se encontram bem sistematizadas.

Fórmulas e combinatórias lexicais como: *processo, resultado, princípio, técnica, método, pertence a, conjunto (é) constituído por, é parte de, é caracterizado por*, etc., ajudam a caracterizar e a perceber o posicionamento dos conceitos correspondentes à área em estudo e à elaboração do mapeamento do seu sistema conceptual com bastante precisão.

Estes indicadores, para além de revelarem a existência de relações conceptuais, revelam, como esperado, a presença e localização dos próprios termos, bem como a detecção de elementos caracterizadores que permitem não só a compreensão do conceito, mas também a elaboração da respectiva definição. Teremos oportunidade de verificar e confirmar esta afirmação nos Capítulos 5 e 6 deste trabalho.

Igualmente indissociáveis de todas as teorias de terminologia são as línguas de especialidade. Abordamos, em seguida, os aspectos essenciais que as definem e as caracterizam.

1.1.4. Línguas de especialidade: definição e caracterização

Por força da criação e aumento do número de diversos domínios de especialidade, bem como do aprofundamento do grau de conhecimento técnico e científico, tem-se assistido, com maior ou menor sistematização, a um crescente aperfeiçoamento das línguas de especialidade e de terminologias inerentes a essas áreas de conhecimento especializado, criadas a partir de ontologias, sistemas conceptuais e de redes semânticas (Sonneveld *et al.*, 1993: 1). Com efeito, o ser humano tem apreendido o mundo de forma distinta e para que o referido aprofundamento do conhecimento se concretize é necessário comunicar utilizando um discurso apropriado e eficiente, não só entre os especialistas da mesma área, mas também de outras áreas de especialidade (por exemplo, técnicas e científicas). Nesse sentido, as línguas de especialidade desempenham, no entender de Pavel *et al.* (2002: xvii), um papel fundamental sendo utilizadas “*para proporcionar uma comunicação sem ambiguidade numa área determinada do conhecimento ou da prática, com base num vocabulário e em usos linguísticos específicos desse campo*”. Neste seguimento, e destacando igualmente a dimensão comunicativa das línguas de especialidade, Cabré (1998: 23) descreve-as como instrumentos elementares de comunicação entre profissionais e, referindo-se à Terminologia, acrescenta que se trata do elemento mais importante que não só distingue

as línguas de especialidade da língua geral, como também as línguas de especialidade entre si.

Muitos são os autores que no âmbito de publicações sobre Terminologia têm dado lugar de destaque ao papel das línguas de especialidade, bem como a alguns dos seus elementos caracterizadores. Temmerman (2000: 47), por exemplo, apresenta uma perspectiva para as línguas de especialidade baseada no princípio da horizontalidade e verticalidade. A autora argumenta que a classificação horizontal se refere ao domínio específico da actividade e que a classificação vertical se baseia em quatro princípios: o nível de abstracção, o tipo de língua (a forma natural ou artificial de expressar os elementos sintácticos), o contexto envolvente (os requisitos impostos, por exemplo, pelo tipo de publicação) e, finalmente, os elementos participantes no processo comunicativo.

Um outro aspecto salientado por Hoffmann (1988, *in* Silva, 2004), relativamente à especificidade da língua de especialidade é o de que pode ser medida através da frequência. Na realidade os textos de especialidade tendem a apresentar frequências de repetição lexical mais elevadas do que textos de língua geral devido ao facto de incluírem nomenclaturas próprias. Como é facilmente constatável por métodos estatísticos, numa língua de uso geral existem palavras que são mais utilizadas do que outras. Verificamos também que o léxico adstrito a uma determinada área de conhecimento dessa língua predomina sobre o de outras áreas. Esta ideia é-nos confirmada por Yzaguirre (1996: 70), quando chama a atenção para duas evidências que possibilitam, ao mesmo tempo, a detecção automática de termos a partir de textos pertencentes a uma área de especialidade: os termos aparecem com uma frequência mais elevada nos textos de especialidade do que nos de língua geral, para além de não ser possível elaborar textos de uma área específica sem recorrer aos seus termos. Nestas condições não seria difícil imaginar que a substituição lexical acarretaria imprecisões e, possivelmente, incorrecções (Finatto *et al.*, 2004, *in* Silva, 2004: 2). Entretanto, convirá ter presente que o discurso de especialidade se caracteriza

“(...) pelos factores e critérios de coesão, coerência, intencionalidade, aceitabilidade, informatividade, situacionalidade e intertextualidade, é definido como uma estrutura comunicativa complexa na qual interferem o autor, seus objectivos de comunicação e a estratégia de comunicação visando o destinatário do texto.”

Hoffmann (1988, *in* Silva, 2004: 2)

Quer isto dizer que a língua de especialidade também possui vários níveis, dado que é possível reformular o discurso consoante, por exemplo, o seu receptor. No caso de

uma publicação, cujo conteúdo possua um elevado grau de especialização, dificilmente será compreendido na sua totalidade ou será até mesmo considerado opaco por um público com conhecimentos menos aprofundados nessa área, não por questões linguísticas, mas por questões de ordem conceptual.

1.2. Linguística de *Corpus* e *Corpora*

Those who work with computer corpora are suddenly finding themselves in an expanding universe.

(Leech, 1991: 25)

Esta citação de Leech dá-nos o mote para a próxima secção na qual destacamos o papel da Linguística de *Corpus* e o seu objecto de estudo – o *corpus* textual.

1.2.1. Breve perspectiva histórica

Existem hoje em dia inúmeras ferramentas informáticas vocacionadas para a extracção de informação, conhecimento e terminologias que, em associação com outras áreas da Linguística, permitem realizar pesquisas cujos resultados têm contribuído, de sobremaneira, para uma nova perspetivação da terminologia baseada na observação e na descrição de fenómenos contidos em textos, quer em língua geral, quer em línguas de especialidade. Uma das áreas da Linguística que mais tem contribuído para esses estudos é a Linguística de *Corpus* disponibilizando aos investigadores meios empíricos de pesquisa reveladores de resultados credíveis.

Apesar de nos dias de hoje a Linguística de *Corpus* ser uma das áreas de investigação da língua de maior interesse e investimento, o facto é que o seu reconhecimento tardava. Apesar disso, desde a compilação do primeiro *corpus* em formato electrónico no início da década de 1960 (o *Brown Corpus*, da Brown University, sob a responsabilidade de Nelson Francis e Henry Kucera), a sua evolução tem sido permanente.

Pensamos, inclusivamente, que deixou de ser necessário fazer a apologia da sua utilização, uma vez que entrou definitivamente como método ou abordagem

metodológica de suporte na defesa de teorias ou estudos descritivos das línguas, quer através da descoberta de novos factos, quer pela confirmação de hipóteses. Não obstante, a história recente da Linguística de *Corpus* tem provado que as suas potencialidades não se esgotam na análise linguística. Este ramo da Linguística possibilita a observação do uso real de toda a informação em determinados contextos, a análise da ocorrência de elementos intra- e intertextuais, igualmente contextualizados, bem como as complexas relações criadas entre si, de forma a gerar novos conhecimentos.

Para o seu sucesso não é alheia a contribuição, desenvolvimento e intensificação do uso de aplicações e programas informáticos específicos, que no início eram mais vocacionados para o estudo da língua geral sendo, actualmente, construídos de acordo com as necessidades específicas de um dado projecto que pode ser, por exemplo, terminológico.

1.2.2. Linguística de *Corpus* e *Corpus*: definição, caracterização e aplicações práticas

Por se tratar de uma área em amplo desenvolvimento, muitos autores têm avançado definições, umas incompletas ou confusas, outras mais abrangentes e outras, ainda, mais objectivas das quais destacamos a de Sardinha (2000):

A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais que foram colectados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

(Sardinha, 2000: 325)

Com efeito, o estudo da linguagem com base em *corpora* textuais combina, actualmente, aspectos de duas orientações que durante algum tempo se opuseram: por um lado, a Linguística numa abordagem baseada no racionalismo, que considera o conhecimento como fruto de princípios estabelecidos *a priori* e, por outro lado, uma abordagem baseada no empirismo, que dá lugar de destaque aos dados resultantes da observação da língua. McEnery *et al.* (1996: 87) descrevem a Linguística de *Corpus* como “(...) *a subset of the activity within an empirical approach to linguistics.*”

Na perspectiva do relatório do EAGLES – *Expert Advisory Group on Language Engineering Standards Guidelines* (1996)¹⁰:

A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.

Tendo em atenção todas estas considerações, a noção de *corpus* poderá parecer, à primeira vista, bastante elementar. Contudo, como elemento central da Linguística de *Corpus*, o *corpus* tem sido considerado como objecto heurístico, potenciador de descobertas. Os *corpora* informatizados permitem ao investigador, normalmente da área da Linguística, o acesso a certos dados mais adequados que de outra forma estariam indisponíveis. Este será porventura um dos motivos pelo qual é geralmente aceite que um dos objectivos da utilização de *corpora* nos estudos de qualquer área é, conforme refere Mellet (2002: 2), o de fornecer as bases empíricas necessárias para dar resposta a questões teóricas ou suportar hipóteses estruturais. Os *corpora* tornaram-se fontes inesgotáveis de possibilidades de estudo das mais diversas ciências e disciplinas humanas, sociais e linguísticas, entre outras. Contudo, verifica-se que é no campo da Linguística, que a noção se tem tornado mais complexa no decurso das últimas décadas devido, sobretudo, à diversidade das práticas e dos objectivos atribuídos à constituição e exploração de *corpora* (Mellet, 2002: 1). Essa diversidade traduz-se, por exemplo, na existência de uma grande variedade de tipos de *corpora*. De entre as várias classificações de *corpora* propostas por diversos autores no decorrer da relativamente curta história da Linguística de *Corpus*, caracterizaremos apenas o tipo de *corpus* que mais directamente diz respeito ao estudo proposto.

Os dois *corpora* construídos para este estudo reúnem simultaneamente as características dos *corpora* monolíngues, comparáveis e de especialidade. Monolíngues, porque cada um dos dois *corpora* são separadamente constituídos por textos pertencentes a um único idioma (português ou inglês). Comparáveis, dado que foram construídos a partir de textos de duas línguas distintas versando sobre o mesmo domínio conceptual. Esta última característica da especificidade temática atribui-lhes igualmente a designação de *corpora* de especialidade. Deste modo, e dadas as características do presente estudo, optámos pela construção de dois *corpora* combinando as características supra-referidas, transformando-os em ferramentas potentes para a extracção de

¹⁰ Disponível em: <http://www.ilc.cnr.it/EAGLES96/corpintr/node13.html>, [acedido em 2005.10.15].

informações relevantes, tais como, candidatos a termos, definições e outro tipo de informações terminológicas complementares para a pesquisa proposta.

Das características dos *corpora* monolingue e comparável, definido no relatório EAGLES (1996)¹¹ como “(...) *one which selects similar texts in more than one language or variety.*”, a tarefa de recolha de textos, embora árdua, revelou-se mais profícua, uma vez que cada texto continha apenas uma língua (a original). Tal como lembra Wallinski (2003: 86), existe o pressuposto de que a comparabilidade de textos de uma sub-língua, ou língua de especialidade, implica a existência de uma terminologia comum. A sua construção pautou-se, desta forma, por critérios gerais de comparabilidade entre as duas línguas de trabalho, nomeadamente pela restrição ao domínio e sub-domínio propostos (*Riscos Naturais* e *Cheias*), pela delimitação ao mesmo género e tipologia textual e discursiva, bem como pela sua analogia cronológica, já que devem pertencer ao mesmo período. A preferência dada ao *corpus* comparável para a execução deste projecto, em detrimento do *corpus* paralelo, é ainda mais facilmente compreendida se atentarmos à seguinte afirmação contida no relatório EAGLES (1996)¹²:

The possibilities of a comparable corpus are to compare different languages or varieties in similar circumstances of communication, but avoiding the inevitable distortion introduced by the translations of a parallel corpus.

Esta ideia é complementada por Maia (2002a: 30), ao referir que os *corpora* comparáveis mais procurados são os de domínio específico.

Dado não ser intuito deste trabalho elencar a tipologia completa dos *corpora* existentes, propomos, em seguida, descrever e caracterizar apenas o *corpus* que consideramos ajustado ao estudo: o *corpus* de especialidade.

1.2.3. *Corpus* de Especialidade

A designação de *corpus* de especialidade deve-se ao facto de este *corpus* ser constituído por textos referentes a uma língua de especialidade pertencente a um

¹¹ Disponível em <http://webdeptos.uma.es/filifa/personal/amoreno/teaching/cl/corpus-typ.html>, [acedido em 2005.10.15].

¹² Disponível em: <http://www.ilc.cnr.it/EAGLES96/corpus-typ/node21.html>, [acedido em 2005.10.15].

domínio ou sub-domínio do conhecimento. Neste contexto, uma língua de especialidade é utilizada, de acordo com as palavras de Pavel *et al.* (2002: xvii), “ (...) *para proporcionar uma comunicação sem ambiguidade numa área determinada do conhecimento ou da prática, com base num vocabulário e em usos linguísticos específicos desse campo*”.

Nesta altura, julgamos interessante e oportuno fazer, ainda, duas breves referências. A primeira ao *corpus* para fins específicos (*special purpose corpus*), proposto por Pearson (1998: 48), que muito linearmente o descreve como um *corpus* cuja composição é determinada pelo fim específico para o qual se destina, o que na realidade se aplica a este projecto. A segunda referência, que temos como particularmente semelhante ao *corpus* supracitado, é o *corpus* situacional que Temmermann (2000: 53-54) propõe para o conjunto de documentos e fragmentos de textos avaliados, seleccionados e organizados mediante um critério específico.

A construção de *corpora* de especialidade para extracção de informação terminológica em línguas de especialidade é proposta como uma das muitas possibilidades de estudo e aplicação prática que o campo da Linguística Computacional oferece nos dias de hoje. É comumente aceite que a qualidade de um *corpus* de especialidade, no qual se baseia um estudo ou um trabalho de investigação, determina a qualidade do seu resultado final. Esta constatação reflecte a importância da observação de certos preceitos na construção de um *corpus*. A sua estrutura deve ser architectada consoante a finalidade da pesquisa e deve ser representativa de uma determinada língua de especialidade. Embora não exista um conjunto fixo de procedimentos para a elaboração de um *corpus*, a realidade é que há elementos estruturantes de carácter mais generalista que devem ser pré-definidos e observados na sua construção, tais como são os propostos por McEnery *et al.* (1996: 21-6): leitura automática, extensão, amostra e representatividade e, por fim referência padrão, sobre os quais fazemos em seguida algumas breves considerações.

Em relação à leitura automática ela é hoje em dia praticamente indissociável da descrição de *corpus*. As contribuições no campo da programação computacional, que através das suas ferramentas informáticas se tornam elementos potenciadores da análise e exploração de *corpora*, têm sido determinantes para o seu rápido reconhecimento como área de estudo da Linguística Aplicada. A importância e utilidade do uso de computadores é salientada por Biber *et al.* (1998:4-5) quando refere: “(...) *computers provide consistent, reliable analyses – they don’t change their mind or become tired*

during an analysis. Computers can also be used interactively, allowing the human analyst to make difficult judgements while computer takes care of record-keeping.”

O segundo elemento referido pelos autores McEnery *et al.* (1996: 21-6), diz respeito à extensão do *corpus*. Contudo, por se tratar de uma amostra, seja ela de cerca de 157.000 átomos (por língua), de que é exemplo o Projecto de *Riscos Naturais: Cheias*, ou de 100 milhões de palavras como é o caso do *British National Corpus* (BNC), o seu tamanho é sempre limitado. Uma vez que não é esperado que um *corpus* de especialidade atinja a dimensão de um *corpus* geral ou de referência, torna-se fundamental investir na qualidade e representatividade dos textos seleccionados. É, aliás, por esse motivo que Johansson (1991: 306) defende que os *corpora* pequenos têm a vantagem de poderem ser cuidadosamente construídos e exaustivamente analisados sob as mais variadas formas. Este critério de exigência funciona como garantia da fiabilidade e do sucesso dos resultados do estudo. De qualquer forma, focalizar a importância de um *corpus* no seu tamanho é ter, como refere Leech (1991: 10), uma visão *naïve*, pois como é facilmente perceptível, “(...) *a collection of machine readable text does not make a corpus.*”

A extensão de um *corpus* encontra-se geralmente associada à representatividade, tido como um dos aspectos mais importantes na sua construção. Não existe, no entanto, uma medida exacta para a representatividade, embora seja ideia generalizada que um *corpus* de especialidade não necessita ser tão extenso como um *corpus* geral ou de referência. O que está demonstrado por autores como Sardinha (1999) e Biber *et al.* (1998), é que quanto mais extenso for, maior será a probabilidade de nele se encontrarem representados registos de baixa frequência. Num *corpus* de especialidade, a classificação de representatividade não significa que este seja absolutamente exaustivo nessa mesma área de especialização, mas que funcione como uma amostra significativamente expressiva de uma determinada realidade ideal para pesquisas específicas. Neste sentido, e na medida do possível, o *corpus* deve conter textos completos pois, com apenas pequenos excertos e amostras corre-se o risco de se perder, por exemplo, a riqueza terminológica, contextual ou conceptual de um texto.

Para além da representatividade máxima de uma língua ou sub-língua de especialidade é crucial o controlo da tipologia dos textos a incluir num *corpus*. Conforme lembra Hassan (*in* Sardinha, 2000: 348), “*Para serem adequados, os corpora devem ser afinados com os objectivos da análise.*” É, desta forma, importante ter em consideração que os textos de um mesmo domínio devem ser seleccionados tendo em

conta dois aspectos determinantes, sendo o primeiro a sua autenticidade, isto é, os textos não foram propositadamente criados para esta situação, à excepção, por razões óbvias, de textos em formato de glossário. O segundo aspecto refere-se às suas afinidades semânticas fazendo ressaltar, por isso, a sua homogeneidade ao nível do conteúdo. Maia (2002a: 30) refere que este tipo de textos pode incluir vários níveis de especialização, embora a tendência seja a de utilizar textos com elevado grau de densidade e especificidade lexicais e terminológica que, quando elaborados correctamente contribuem para um maior sucesso da pesquisa e extracção de informação terminológica.

É, pois, necessário dominar o mais possível o funcionamento dos *corpora*, já que pelas provas dadas ganharam o estatuto de ferramentas indispensáveis para a realização de um qualquer trabalho terminológico, pelo que o nosso estudo não é excepção. Os *corpora* de especialidade oferecem a possibilidade de fornecer dados terminológicos tais como termos, bem como definições e contextos de uso, como teremos oportunidade de verificar nos Capítulos 2 e 3, deste trabalho.

2. A definição

Juntamente com a unidade terminológica, a definição é um elemento chave nos estudos terminológicos. Sager, numa perspectiva mais abrangente, refere no Prefácio do seu livro “*Essays on Definition*” que a definição “(...) *occupies a central place in all sciences and is a fundamental tool in logic, philosophy of ideas and semantics* (...)” (Sager, 2000: vii). Por extensão, podemos referir que o texto que comporta a definição ocupa um lugar de relevo nos estudos terminológicos e no trabalho terminográfico.

Uma vez que a constituição de terminologias advém, essencialmente, de necessidades comunicativas dos agentes de um dado domínio de especialidade, o texto definicional deve corresponder ao objectivo do produto terminográfico, isto é, deve haver uma correlação entre a escolha do léxico que compõe as definições terminológicas de um dicionário, glossário etc., e o público ao qual se destinam.

Não é objectivo desta dissertação elencar todos os modelos de definição que têm vindo a ser propostos desde a Antiguidade até aos nossos dias, nem pretendemos abordar a problemática dos conceitos da filosofia e da lógica que subjazem aos termos *definir* e *definição*. Contudo, em coerência com a distinção entre termo e não-termo, (apresentada no ponto 1.1.3.2), consideramos igualmente indispensável para o nosso estudo enveredarmos por um caminho mais estreito que conduz apenas a uma breve descrição da definição no seu sentido mais lato, concretizando no sub-capítulo 2.3. a descrição de definição terminológica.

2.1. Definir “*definição*”

Rey (*in* Sager, 2000: 7) argumenta que todas as discussões sobre o conceito geral de *definição* giram em torno da confusão generalizada das diversas abordagens filosóficas, não se devendo culpabilizar cientistas, epistemólogos e lógicos ou lexicógrafos por esse facto pois, resulta da popularização do conhecimento e de uma percepção inadequada da originalidade de cada processo. O ponto de articulação, em termos de definição, tal como em outros, é pois a terminologia, ela própria vítima de uma base teórica menos adequada, tal como a lexicografia.

Contrariando um pouco a ideia da importância da definição expressa por Rey (1995: 40) quando refere que “*Definition and description are probably the core issues of terminology.*”, o Glossário do Manual de Terminologia descreve, no nosso entender, de uma forma algo vaga, o termo “definição” como “1 – *Fórmula lexicográfica que explica o conceito designado por um termo.* 2 – *Em uma ficha terminológica, tipo de prova textual que permite estabelecer a equivalência textual em várias línguas ao descrever os traços semânticos distintivos de um conceito.*” (Pavel et al., 2002: 119).

O que é, então, uma definição? Na resposta a esta questão existe um ponto de consenso pois, definir é, antes de mais, fazer um exercício de análise conceptual, descrevendo, delimitando, distinguindo e relacionando conceitos, bem como a sua representação através de signos linguísticos, sob a forma de léxico de uso geral no caso da lexicografia e léxico de especialidade (termos), no caso da terminografia. É apresentar atributos e características específicas para que não possam ser confundidos com outros.

Sager (2000: vii) lembra que devido à existência de diversos pontos de vista no que se refere à definição, bem como ao vasto leque da utilização das definições, surge uma grande diversidade de interpretações sobre o significado de *definir* e do seu produto - a *definição*. Daí que Rey (*in* Sager, 2000: 1) afirme que “*Definition can be defined in various ways.*” No entanto, é também nosso entendimento que uma definição não deverá ter uma abordagem diferente conforme a óptica de quem a produz. Não deverá constituir um resultado previamente estipulado, a obter a qualquer custo, cedendo a interesses que não sejam técnicos ou científicos, pelo que todo o cuidado é pouco, quando se trata de definir o que quer que seja.

De forma a determinar a necessidade das definições, Sager (1990: 45-8) propõe a análise de dois aspectos: as necessidades dos utilizadores e a natureza dos conceitos a definir. Este autor apresenta, assim, três necessidades básicas para a definição:

- a fixação inicial da equação termo-conceito¹³;
- a identificação de um termo através da verificação da existência de uma definição independente¹⁴;
- A explicação do significado do conceito para uso dos especialistas de base de dados terminológicas tal como tradutores e especialistas e, possivelmente de leigos¹⁵.

¹³ Nossa tradução.

¹⁴ Idem.

2.2. Tipologia de definições

Dependendo das relações entre os conceitos, existem duas formas diferentes de determinar o significado dos termos. A classificação que, porventura, recolhe maior consenso entre os vários autores da área da terminologia, dos quais destacamos Sager (1990: 24), Cabré (1999: 98), Temmerman (2000: 79), Wright *et al.* (1997: 30), é a seguinte:

- a definição por intensão (também designada por definição intensional, específica ou por compreensão) consiste em determinar os aspectos caracterizadores do conceito proposto para definição, ou seja, determinar a compreensão do conceito, mencionando o conceito genérico mais próximo (entretanto já definido ou conhecido de outra forma);
- e a definição por extensão (também designada por definição extensional ou genérica), que se baseia na enumeração dos elementos individuais e pertencentes ao mesmo nível conceptual e aos quais o conceito faz referência ou aos conceitos específicos que lhe são imediatamente subordinados. Não é uma enumeração exaustiva e é mais facilmente observável do que uma definição específica. Por conseguinte, pode ser útil completar uma definição específica com uma definição genérica.

Não raras vezes, a interligação entre estes dois tipos de definições baseia-se numa relação proporcionalmente oposta. Por exemplo, à medida que aumenta a especificidade de um termo, através da pormenorização dos seus elementos caracterizadores, diminui a amplitude do termo, dado que passa a existir um menor número de elementos (conceitos, objectos, etc.) caracterizadores de índole mais geral que cabem nessa descrição.

Este é, sem dúvida, o tipo de classificação mais genérica e, que por esse motivo, se enquadra perfeitamente num qualquer trabalho terminológico. Existem, porém, muitas outras possibilidades de tipologias de definição, propostas por diversos autores, o que julgamos perfeitamente admissível atendendo aos já referidos diferentes pontos de vista e interpretações para um mesmo objecto ou conceito (*cf.* ponto 3.3). No entanto,

¹⁵ Nossa tradução.

dados os objectivos e características do nosso estudo, focalizaremos o nosso interesse apenas na definição terminológica.

Concluimos com as palavras de Rey (*in* Sager, 2000: 14), quando refere que as dificuldades encontradas em definir *definição* exemplificam a complexidade e a necessidade de definir, que se tornou um pré-requisito para a produção e compreensão das manifestações mais importantes do discurso social.

2.3. A Definição Terminológica

2.3.1. Sua definição e caracterização

O termo é um instrumento de representação e de comunicação comum a todas as línguas. Por outro lado, o conceito existe *per si*, é extra-linguístico, e a sua designação pode variar entre as diversas comunidades linguísticas e culturais. Como tal, não será demais perceber a *definição* como um fenómeno linguístico e comunicativo pois, ambos são produto de escolhas e de usos com raízes sócio-culturais de uma dada comunidade linguística. A definição terminológica permite, por assim dizer, a fusão dos aspectos conceptual, linguístico e comunicativo, uma vez que possui várias funções que ocorrem simultaneamente, descrevendo os elementos caracterizadores de um dado conceito e delimitando-o a um sistema conceptual específico permitindo, ao mesmo tempo, o estabelecimento de relações.

Na perspectiva terminológica, o conceito é “*uma unidade abstracta criada a partir de uma combinação única de características*” (Lara, 2004: 92). A definição terminológica tem, justamente, como função transmitir a imagem mental de um conceito numa área de conhecimento de especialidade, através da delimitação precisa e da sistematização dos traços semânticos caracterizadores e ao mesmo tempo diferenciadores de outros conceitos do mesmo domínio, para que a apreensão dos conteúdos semânticos se faça de uma forma inequívoca. Esta ideia é defendida por Sager (2000: vii) quando refere que, tanto o acto de definir, como as definições, desempenham um papel fundamental na terminologia e na lexicografia pois, são o meio convencional para o estabelecimento do significado de unidades lexicais, ou dito de outra forma, para ligar o conceito à palavra ou ao termo que representa. Nesse sentido, a definição (terminológica, ou outra) é utilizada como um meio para eliminar a

ambiguidade. Contudo, a forma condensada do texto de definição lexicográfica deixa, por vezes, espaço para que surjam situações de falta de clareza e de precisão, o que é entendido como inaceitável numa definição terminológica.

De acordo com a proposta avançada por Pavel *et al.* (2002: 23-4) no *Manual de Terminologia*, a definição terminológica clarifica o significado do texto, assunto ou considerações contextuais e determina o significado que se quer atribuir a um termo. Comparativamente, a definição lexical é mais abrangente e o conceito que representa é passível de ser compreendido por um maior número de pessoas (mesmo de outros sistemas linguísticos). Este tipo de definição, que pretende descrever os diversos sentidos que um vocábulo possa deter, tendo em conta os usos estabelecidos, está, também, mais sujeito a fenómenos linguísticos incontornáveis da língua geral, tais como a polissemia e a ambiguidade.

Independentemente do tipo de definição, a sua principal funcionalidade é, conforme já referido, clarificar significados. Contudo, a realidade encarrega-se de demonstrar que nem sempre é possível colher uma definição clara, devido a inúmeros factores envolvidos no processo definicional, de que é exemplo a dificuldade de delimitação de conceitos ou noções de cariz mais subjectivo.

Apesar de se encontrarem bem determinados quais os elementos a constar numa definição (mesmo considerando os diversos pontos de vista), o facto é que como já constatámos anteriormente, definir *definição* continua a ser uma tarefa intrincada. Tomemos como exemplo a definição terminológica. Dependendo do produto terminológico assim dependerá o conteúdo deste tipo de definição: deve focar os pontos essenciais, evitar a circularidade, corresponder à extensão do termo, evitar linguagem figurativa ou opaca e ser positiva e não negativa.¹⁶ A aplicação destas e de outras regras resulta numa definição bem elaborada, com a ausência de ambiguidades, polissemias ou qualquer outro vício próprio da língua de uso geral, como se pretende num produto terminológico. Esta é, pelo menos, a situação ideal. Contudo, a prática revela que a elaboração de uma boa definição terminológica é difícil de atingir pois, por vezes assistimos à existência de factores que impedem que tal aconteça, como é o caso das particularidades da língua geral que acabámos de mencionar e das várias teorias referentes à escolha das características que compõem um termo e que a respectiva definição deverá incluir.

¹⁶ Disponível em: <http://www.philosophypages.com/lg/e05.htm> [acedido em 2005.10.15].

Para além do aspecto linguístico da observação das regras na construção de definições terminológicas para a estruturação da língua de especialidade, existe um factor extra-linguístico de igual importância. Trata-se da organização dos termos no sistema conceptual em foco, uma vez que é fundamental que a área de conhecimento esteja bem estruturada. A terminologia não se limita a fornecer termos pertencentes a determinadas áreas de especialidade, “(...) *mas referenciais metodológicos para sustentar a estruturação dos campos lógico-semânticos* (...)”, conforme sugere Lara (2004: 91).

2.3.2. Sistematização das suas principais regras

Não obstante todas as considerações teóricas e de forma a prevenir o tipo de situações que acabámos de descrever, existem regras objectivas referentes à estrutura física do texto terminográfico, que quando observadas ajudam a evitar erros de definição:

- captar apenas a extensão correcta dos elementos caracterizadores que aquele conceito possui naquele domínio de especialidade. Como tal, não necessita ser demasiado lata (ao incluir elementos que não deveriam constar da descrição), ou demasiado restrita (ao não incluir elementos que deveriam constar da descrição) embora deva reflectir o conhecimento e satisfazer as necessidades do utilizador do produto terminográfico (Kemerling, 2001: 3);
- possuir o grau de tecnicidade e complexidade adequado aos utilizadores finais;
- ser esclarecedor, rejeitando a linguagem figurativa ou opaca, evitando, dessa forma, uma definição de mais difícil compreensão do que o próprio termo ou conceito e conservando o seu objectivo de clarificação (Kemerling, 2001: 3);
- evitar a tautologia e a circularidade, excluindo da definição os elementos que se remetem reciprocamente e os próprios termos a serem definidos, embora a utilização do seu hiperónimo imediato, ou de termo mais genérico, seja admissível pois, permite estabelecer uma relação entre os dois e trazer clareza e equilíbrio à definição;
- utilizar afirmações positivas em vez de negativas (Kemerling, 2001: 3);
- evitar a conflitualidade, evidenciando contrariedades (Kemerling, 2001: 3);
- sempre que possível, apresentar a definição em apenas uma frase (sem prejuízo para a ordem sintáctica natural).

Independentemente das diversas propostas de regras para a construção de uma boa definição, pensamos ser unânime a opinião de que a definição terminológica deve situar o conceito de forma clara e precisa no respectivo sistema conceptual.

As experiências terminológica e, sobretudo, terminográfica dizem-nos que não existe uma única forma de atingir a definição terminográfica pois, existem inúmeros factores que concorrem para o estabelecimento do seu texto final. Factores como o domínio de especialidade, os utilizadores finais, o objectivo do produto final, entre outros, levam o terminólogo a eleger uma metodologia adequada e adaptada a essas premissas.

O conhecimento não é estático, pelo que as definições também o não devem ser e, como tal, é importante expandir o horizonte da pesquisa buscando os novos limites do conceito e do termo e os seus novos contextos de uso. No caso específico do nosso trabalho, precisamos compreender o processo de elaboração de uma definição terminológica para sabermos, com precisão, quais as informações úteis a extrair dos contextos.

No capítulo seguinte propomo-nos comentar a importância dos contextos nas línguas de especialidade, bem como a extracção de informação terminológica relevante.

3. O contexto e a extracção de informação terminológica

Neste capítulo iremos destacar a importância do contexto no âmbito da pesquisa terminológica, bem como as propostas de cinco autores no que se refere à detecção e extracção de informação terminológica relevante que para além de termos, definições, sistemas conceptuais e redes semânticas, inclui os contextos.

3.1. Introdução

Os textos de especialidade não se limitam a fornecer unidades terminológicas, conforme já tivemos oportunidade de referir no ponto 1.1.4. (*Línguas de especialidade: definição e caracterização*). O seu conteúdo incluiu dados e informações de particular relevo para a elaboração de Fichas Terminológicas, bem como a constituição de produtos terminológicos criados a partir dessas mesmas fichas. Um texto especializado, para além de definir, descrever, explicar, contextualizar e apresentar usos empíricos associados às unidades terminológicas, proporciona, igualmente, termos relacionados, sinónimos, antónimos e todo um conjunto de dados que concorrerem para a compreensão do significado dos termos.

Do ponto de vista dos não-especialistas, estudantes ou simples leigos de uma dada área de especialidade, existem textos cujo grau de opacidade é bastante acentuado (nomeadamente em contextos comunicativos entre especialistas) e onde ocorrem unidades terminológicas cujo significado é polissémico, acarretando a indesejada ambiguidade. Esta dificuldade de descodificação da informação terminológica, devido a fenómenos como a polissemia e a opacidade, levam o investigador a socorrer-se de outros meios de pesquisa, sejam eles de análise terminológica, contextual ou outros. A análise das unidades terminológicas em contexto, bem como a análise dos respectivos conceitos traz consigo inúmeras vantagens, nomeadamente de clareza e de atestação de uso para compreensão e futura utilização noutros contextos situacionais.

É consensualmente aceite que as noções de *termo* e *conceito* estão intimamente relacionados com a de *definição* pois, o termo representa o conceito e a definição descreve e caracteriza o conceito. Por sua vez, o contexto situa tanto o termo como o conceito numa realidade conceptual e linguística (o seu referente). O contexto ajuda,

desta forma, não só a definir as relações entre termo e conceito, mas também a formular definições.

Embora nem sempre com as fronteiras bem definidas, contexto e definição são portadores de informações distintas, tal como defende Blanchon (1997: 169). O contexto possibilita apenas o fornecimento de informações incompletas sobre o conceito (podendo induzir a um determinado fluxo conceptual e prejudicar a compreensão), ou, pelo contrário, fornecer elementos suficientes ou bastantes para a total compreensão do conceito em causa. A definição, pelo contrário, descreve o conceito num determinado contexto e finalidade e situa-o com precisão no sistema conceptual de especialidade do qual faz parte. Aliás, como já tivemos oportunidade de verificar no Capítulo 2, a definição permanece como um dos tópicos mais debatidos pelos terminólogos, tendo em conta os inúmeros manuais, estudos, respectivas recomendações e regras surgidos ao longo dos últimos anos.

A relevância do contexto numa pesquisa terminológica é incontestável. No entanto, tal como em qualquer trabalho de investigação, é necessária alguma cautela pois, o recurso à análise contextual apresenta também algumas desvantagens. Por exemplo, antes de tomarmos como certo um determinado contexto de definição, ou até mesmo uma definição, convirá perceber se estamos perante uma abordagem consentânea dos especialistas daquela área ou se, por outro lado, estamos perante uma abordagem ou interpretação individual de um especialista que não recorreu aos mesmos critérios. Dubuc (1992: 75) lembra que é prudente compreender que, por vezes, também existe alguma irregularidade na apresentação final dos produtos terminológicos que versam sobre o mesmo domínio. Para essa irregularidade concorrem inúmeras causas, entre as quais, a adequação do produto ao público-alvo pois, o seu grau de especialização varia consoante o público a servir. A abordagem do investigador dependerá da sua óptica, das suas escolhas, dos seus métodos e de um conjunto de constrangimentos que sirvam os interesses, por exemplo, de quem financia o projecto. Mesmo quando as investigações estão livres de obrigações previamente estipuladas, quando se procura um termo inserido no seu contexto natural é, mais uma vez, aconselhável prudência na selecção da informação terminológica.

Todos sabemos que, tanto o conhecimento como os sistemas linguísticos não são estáticos, nem tão pouco os significados dos termos e os contextos que os envolvem. Esta é uma das razões pela qual Berdet (1999: 1), a propósito da influência ideológica do autor de uma obra lexicográfica, afirma que “(...) *los ejemplos del diccionario son,*

por una parte, el reflejo de la sociedad de fijación, representan los tópicos culturales de su época, a los que explicitan mediante un texto (citado o inventado) de referencia, actuando como notários de su realidad social.” Estas influências ideológicas sucedem não só em obras lexicográficas, mas também nas terminográficas e no discurso de especialidade que, em alguns casos, tendem a expressar uma determinada corrente ideológica ou, simplesmente, a fornecer informação terminológica pertencente a determinado grupo social, profissional, regional, etc.

É também verdade que um termo que ocorre num texto raramente contempla a sua definição terminológica em conformidade com o que é exigido numa Ficha Terminológica. Por este motivo socorremo-nos de recursos linguísticos, que, na esmagadora maioria dos casos, funcionam como valiosos fornecedores de informações relevantes, objectivas e úteis para as definições, sistemas conceptuais e todo um conjunto de informações específicas terminologicamente relevantes.

Por informação terminológica relevante entendemos toda a informação que contribui de maneira eficaz para a definição e caracterização de uma unidade terminológica. Esse contributo pode ocorrer de forma mais directa ou indirecta, consoante se tratem de elementos caracterizadores de definição, explicação ou associação de um dado termo.

3.2. Contexto: definição e caracterização

Iremos abordar, em seguida, os vários conceitos ligados à questão do contexto, tendo sobretudo por base a proposta de Dubuc (1992) apresentada no *Manuel Pratique de Terminologie*.

3.2.1. O que é o contexto?

A unidade terminológica não existe *per si*, conforme lembra Cabré (1999: 124). Ela é fruto do uso de uma dada unidade lexical num determinado contexto expressivo e situacional. Ou seja, o termo não ocorre de forma isolada, mas sim de forma combinada com outros elementos linguísticos. Assim também sucede com a definição do termo *contexto*, modelada consoante a situação em que é empregue. De acordo com o

*Dicionário Priberam de Língua Portuguesa On-Line*¹⁷, o termo “*contexto* (s.m.), do Latim *contextu* (s.m.), significa “(...) *encadeamento das ideias de um texto; contextura; entrecho.*” Gramaticalmente significa, “(...) *enquadramento sintagmático de uma unidade do discurso; situação de comunicação; argumento*”. No campo mais restrito da Terminologia, *contexto* é descrito no *Glossário do Manual de Terminologia*¹⁸ como: “1. *Parte de um texto ou enunciado, em que está incluída uma unidade lexical, e que contribui para determinar o seu significado.* 2. *Em uma ficha terminológica, prova textual que fornece informação sobre os traços semânticos de um conceito ou sobre o uso de um termo. Exemplos: contexto definatório; contexto explicativo; contexto associativo*”. Mas antes de avançarmos para o desenvolvimento explanatório destas três categorias contextuais, importa referir como a Norma Internacional ISO/FDIS 1087-1:2000 (E/F), no seu ponto 3.6.10 do capítulo dos “*Aspectos do trabalho terminológico*”, define sucintamente *contexto* como “*text which illustrates a concept (3.2.1) or the use of a designation (3.4.1)*”. No capítulo seguinte, 3.7 Produtos terminológicos, da mesma norma, o termo *contexto* volta a ser mencionado quando no ponto 3.7.6, que define *Concordância terminológica* refere “*ordered list of terms (3.4.3) extracted from a corpus (3.6.9) together with a context (3.6.10) and a source reference*”.

No âmbito do trabalho terminológico existem autores, como Dubuc (1992: 46) e Pavel *et al.* (2002: 48), que distinguem três tipos de contextos conforme o seu conteúdo semântico: de definição, explicativo e associativo. Conforme referido no parágrafo anterior, Pavel *et al.* (2002: 48 e 118) incluem os contextos de definição, explicativo e associativo (aos quais daremos especial destaque em 3.2.2), num conjunto de quatro principais tipos de provas textuais, que funcionam como complemento possibilitando ao utilizador a concepção mental do conceito (Pavel *et al.*, 2002: 48). Em relação aos restantes três tipos de provas textuais, designadas pelas autoras como *observações*, *citações* e *unidades fraseológicas*, gostaríamos apenas de referir que as primeiras “(...) *fornecem informação sobre as nuances do conceito ou sobre o uso dos termos que designam o conceito*”; as segundas devem “(...) *respeitar os direitos autorais e remeter ao código da fonte correspondente*”; e por último, a “*unidade fraseológica ou vocabulário de apoio dos termos que designam o conceito definido (...)*” (também

¹⁷ Disponível em: http://www.priberam.pt/dlpo/definir_resultados.aspx, [acedido em 2005.05.02].

¹⁸ Disponível em:

http://www.translationbureau.gc.ca/pwgsc_internet/fr/publications/gratuit_free/man_termino/glossaire_po_r.htm#c, [acedido em 2005.05.02].

denominados termos coocorrentes), ilustram o uso de um termo no discurso especializado (Pavel *et al.*, 2002: 48-9).

3.2.2. Tipologia dos contextos

Tal como acabámos de referir, para autores como Dubuc (1992) e Pavel *et al.* (2002), em Terminologia, reconhecem-se basicamente três tipos de contextos: definicional, explicativos e associativos. Cada um destes contextos contém *descritores* ou *traços semânticos*, na terminologia de Dubuc (1992: 45), em quantidade e qualidade variáveis determinando a sua classificação.

Contextos definicionais

Os contextos definicionais são os que contém os descritores em quantidade e qualidade suficientes de modo a formar uma imagem precisa do conceito em causa, conforme ilustrado pelo seguinte exemplo:

*As cheias são fenómenos naturais extremos e temporários, provocados por precipitações moderadas e permanentes ou por precipitações repentinas e de elevada intensidade.*¹⁹

Como podemos verificar, trata-se de uma descrição razoavelmente completa do fenómeno, não se tratando, contudo, de uma definição, até porque, neste caso, o próprio contexto inclui a unidade terminológica a definir.

Contextos explicativos

Os contextos explicativos fornecem alguma informação acerca de certos aspectos do conceito ao qual o termo em estudo se refere. Atentemos, desta forma ao seguinte exemplo:

As cheias extraordinárias do Douro caracterizam-se por um grande volume, rápida propagação, forte elevação do nível das águas (sobretudo nos troços mais estreitos) e por uma curta duração de 2 a 3 dias, dado que a descida do nível das águas se faz de um modo

¹⁹ Disponível em: http://inseguranca.no.sapo.pt/pc_cheias.html, [acedido em 2005.10.15].

*relativamente rápido, tal como se pode observar na fig. 4, com dados relativos ao ano de 1995/96, ano da última cheia.*²⁰

Com efeito, neste contexto o conceito não se encontra definido ou descrito com precisão, embora forneça um conjunto suficiente de características para o distinguir e identificar.

Contextos associativos

Os contextos associativos oferecem uma imagem algo ténue do conceito ao qual o termo se refere ou até mesmo não fornecem qualquer informação. Em concreto, apenas evidenciam, sem nenhuma intenção explicativa, a sua ligação ao domínio do conhecimento sob investigação. Na prática, estes contextos não só comportam associações menos evidentes entre termos, como também fornecem dados que permitem circunscrever o significado. Vejamos o seguinte exemplo:

*A recolha e compilação de dados históricos das cheias, se bem que na sua maioria empíricos, constitui uma base indispensável para o estabelecimento das tendências do perigo ou dos desastres relacionados com as inundações.*²¹

Este tipo de contexto é também designado pelo *Manual de Terminologia*, de Pavel *et al.* (2002: 121), como *exemplo de uso*, referindo-se a uma “*Breve citação que ilustra, em uma ficha terminológica, o uso de um termo em uma área temática, sem fazer referência aos traços semânticos do conceito designado.*”

No nosso estudo não procurámos uma distinção formal entre estes três tipos de contextos pois, consideramos que a conjugação das características distintivas de cada uma das diferentes tipologias dos contextos acima mencionados contribuem para a aquisição de uma melhor informação terminológica. Contudo, privilegiámos a descrição dos marcadores lexicais indiciadores de informação terminológica (sobre os quais nos debruçaremos no Capítulo 4) presentes justamente em contextos de definição.

Os contextos terminológicos são extremamente valiosos pois, contribuem para uma melhor compreensão das unidades terminológicas e ajudam a clarificar o

²⁰ Disponível em: <http://www.lettras.up.pt/geograf/APEQ/p11.html>, [acedido em 2005.10.15].

²¹ Disponível em: <http://www.lettras.up.pt/geograf/APEQ/p11.html>, [acedido em 2005.10.15].

conhecimento sobre sistema conceptual em foco fornecendo, igualmente, uma perspectiva actualizada dos termos desse mesmo sistema conceptual.

3.3. Extracção de informação terminológica relevante: termos, definições, contextos de uso, sistemas conceptuais e redes semânticas

São muitos os contributos de autores que se debruçam especificamente sobre este tema. A nossa escolha recai sobre cinco autores (cujas propostas de métodos e técnicas de extracção de informação terminológica descreveremos em seguida) e baseia-se na hipótese de viabilidade destas mesmas propostas serem testadas no *Corpógrafo*. Trata-se, pois, de uma abordagem apriorística sem qualquer intenção de controlo de resultados, para além do facto de querermos gerar material que produza informação interessante ao nível terminológico.

Passamos agora a descrever as propostas dos cinco autores: Dubuc (1992), Sager (1990), Pearson (1998), Conceição (2001 e 2004) e Colussi (2002).

3.3.1. As propostas de Dubuc (1992)

A “Análise Terminológica” (Dubuc, 1992:43-5)

Dubuc (1992), que dedica o Capítulo VII do *Manuel Pratique de Terminologie à Analyse Terminológica* (1992: 43-7), refere que esta é a essência da pesquisa terminológica temática compreendendo dois aspectos: selecção dos termos e a análise dos contextos. Em relação à selecção dos termos, eles representam o conceito, conforme já havíamos definido no Capítulo 2, existindo vários processos para a sua identificação e selecção. Dubuc (1992: 43-5) destaca dois: os determinantes e os critérios de selecção.

A análise da relação entre determinante (por exemplo, *o(s)*, *a(s)* e *a*, *an*, *the*) e determinado (por exemplo, o termo que lhe sucede), pode favorecer a identificação de termos, dado que este é resultado de uma situação de comunicação específica. Nesta matéria Dubuc (1992: 43) advoga que na relação entre determinante e determinado existem dois tipos de determinantes – os acidentais e os essenciais. Os determinantes acidentais não interferem no sentido do determinado, apesar de trazerem alterações ao

nível de aspecto ou de circunstância. Os determinantes essenciais, por sua vez, são elementos modificadores da natureza dos determinados que, devido à sua estreita relação, podem mesmo causar uma mudança de sentido dos determinados, caso sejam suprimidos.

No que se refere aos critérios de selecção, Dubuc (1992: 44-5) propõe quatro: lexicalização, estabelecimento de uma classificação ou de uma oposição, coocorrência e, por fim, artifícios tipográficos.

O critério da lexicalização, ou melhor, do grau de lexicalização, traduz-se na coesão entre as partes do sintagma que faz com que essas partes sejam indissociáveis no seguimento de uma longa utilização. Por exemplo, a ausência do artigo definido (na sua forma combinada) *Riscos Naturais* em vez de *Riscos da Natureza*.

Em relação ao segundo critério, caso a caracterização permita o estabelecimento de relações de oposição entre os conceitos em estudo, podem surgir exemplos como no domínio das Cheias: as *cheias rápidas/repentinas* opõem-se às *cheias lentas*. Desta forma ao detectarmos um destes termos é-nos permitido concluir que existe a forte possibilidade de o outro vir também a ser validado como termo. Da mesma forma que se o determinante permite classificar os conceitos em estudo, atribuindo-lhes uma categoria, podemos estimar que ele faz parte da unidade terminológica. Isso mesmo podemos verificar nos exemplos que, simultaneamente, classificam tipos de Hidrograma e constituem termos: *Hidrograma de Cheia*, *Hidrograma de Escoamento* e *Hidrograma de Carga e Descarga*.

O terceiro critério, a coocorrência, está presente quando uma mesma combinação de palavras ocorre com uma certa frequência no interior de um domínio. Este fenómeno pode indicar a presença de uma unidade terminológica. Num *corpus* de especialidade do domínio das Cheias é frequente encontrar o termo *inundações* associado ao adjetivo *rápidas*, constituindo, por isso, uma coocorrência, facto que o coloca na posição de candidato a unidade terminológica.

Por fim, o quarto critério enunciado por Dubuc (1992: 45), os artifícios tipográficos, tais como o itálico, o negrito, as maiúsculas pequenas, sublinhados e aspas, e que podem ser indício da presença de unidades terminológicas. Por vezes o autor de um texto recorre a estes artifícios para evidenciar uma noção de base do domínio. Dubuc (1992: 45) aconselha que devemos estar atentos a estes sinais sem, contudo, lhes conferirmos valor absoluto, dado tratarem-se apenas de indícios.

A “Análise Contextual” (Dubuc, 1992:45-7)

A análise contextual proposta por Dubuc (1992: 4) consiste na delimitação do contexto circunscrevendo os elementos portadores de significados que aí se encontram, dando-lhes o nome de *traços semânticos* ou *descritores*. Dubuc (1992: 4) acrescenta que, dependendo do número e da qualidade desses descritores, o contexto será de definição, explicativo ou associativo, conforme veremos em *Tipologia de contextos* (cf. ponto 3.2.2).

Um grande proveito a retirar da análise contextual é o de encontrar outros termos a ele associados como definições, descrições ou explicações, sobretudo dos neónimos; enfim, o máximo de informação de interesse terminológico por vezes não se encontra descrita em dicionários de especialidade ou outros produtos terminológicos e terminográficos. Neste sentido, a análise contextual apresentada por Dubuc (1992: 45) tem por objectivo trazer precisão ao conceito, representado pela unidade terminológica, através da identificação dos traços semânticos veiculados pelo contexto. Estes traços semânticos podem manifestar-se sob diversas formas, nomeadamente de relações conceptuais. Todas elas, individualmente ou em associação, fornecem um conjunto de informação essencial contribuindo, por exemplo, para o estabelecimento do desenho da árvore conceptual do domínio de especialidade em investigação. Dubuc (1992: 45) atribui, ainda, mais dois objectivos à análise contextual: delimitar o contexto representado na Ficha Terminológica e oferecer a possibilidade de confirmar a pertença do termo ao domínio em estudo reconhecendo, também, o(s) seu(s) sinónimo(s).

A identificação dos traços conceptuais, a delimitação de contextos e a identificação dos tipos de contextos são, aliás, os três aspectos essenciais salientados por Dubuc (1992: 45-46). Com efeito, as vantagens do recurso à análise contextual são inúmeras. Os trabalhos terminológicos, sobretudo os multilingues, são extremamente facilitados sempre que o investigador puder associar um termo de uma língua ao de outra língua por via da análise contextual. Para além deste recurso, e tal como foi abordado no ponto 1.1.3.2. (*Distinção entre termo e não-termo*), existem determinadas características que definem um termo, distinguindo-o, por inerência, de um não-termo. Da mesma forma, existem indicadores, como é o caso do contexto, da coocorrência e da pontuação, que ajudam a determinar se estamos perante um produto terminológico, isto é, se estamos perante candidatos a termos e outro tipo de informação terminológica ou apenas vocábulos pertencentes à língua geral. Pode, ainda, dar-se o caso de estarmos perante um termo que, apesar de estar presente nos textos do domínio específico sobre o

qual incide a investigação, se confirma pertencer igualmente a outro ou outros domínios que lhe estão próximos. Podemos dar o exemplo concreto do termo *caudal de retorno*, que além de ocorrer em textos relacionados com a área das *Cheias (Riscos Naturais)*, ocorre igualmente em áreas como a Hidrologia, a Hidrogeologia e a Gestão e Desenvolvimento de Recursos Hídricos.

Uma outra vantagem do recurso ao contexto e à análise contextual é o facto de ambos serem especialmente úteis no processo de desambiguação e distinção entre termos e não-termos, de que são exemplo as sobreposições lexicais de vocábulos de uso específico e de outros de uso geral. Apesar de ao nível das suas características lexicais não existirem diferenças, o mesmo não sucede ao nível semântico. É através da percepção do conceito que está por detrás do (candidato) a termo, que o incluímos, ou não, na lista de termos. A análise contextual ajuda a circunscrever os significados específicos e, conseqüentemente, a produzir definições mais precisas.

De forma a evitar contrariedades que possam surgir durante a análise contextual, Dubuc (1992: 52) propõe a adopção de alguns procedimentos. Um deles é a delimitação de contextos. Nesta tarefa, que deve constar na Ficha Terminológica, deve-se ter em conta os elementos descritores que o contexto encerra, sobretudo os mais significativos, já que existe um constrangimento de número de caracteres. Um outro procedimento seguinte refere-se à identificação de traços conceptuais numa unidade terminológica, que é em si um processo incontornável já que a existência de um termo se deve ao conceito que lhe é inerente. É nesse contexto conceptual que se encontram os elementos reveladores do conceito, que Dubuc (1992: 45) designa como descritores ou traços semânticos.

3.3.2. As propostas de Sager (1990)

“Definitions in context” (1990:44-5)

Antes de apresentarmos a proposta de Sager relativa à definição em contexto, pensamos ser útil lembrar o entendimento deste autor sobre o processo de definição. Reconhecendo a existência de vários métodos de definição, Sager (1990: 42-3) destaca sete: definição por análise (*genus et differentia*), sinonímia, paráfrase, síntese (através da identificação de relações ou da descrição), implicação (usando a palavra num contexto explicativo), denotação (fornecendo exemplos, por extensão) e demonstração

(definição ostensiva, que pode traduzir-se na forma de desenhos, fotografias, mapas, fórmulas, etc.). Sager (1990: 43-4) admite, ainda, que estes métodos podem servir de base a combinatórias, criando novos métodos de definição (análise e descrição, sinónimo e descrição, sinónimo e análise, etc.). Admite, também, que alguns destes tipos de definição não fazem mais do que listar informação, pelo que as definições terminológicas geralmente não pertencem aos tipos denotativo, contextual ou demonstrativo. Contudo, sabemos que, em situação discursiva real, não é usual os textos conterem as definições terminológicas na sua versão final como produto terminológico, pelo que é necessário recorrer precisamente ao tipo de informação que contém os traços semânticos que permitem a elaboração do texto de definição terminológica.

No que se refere às definições em contexto, Sager (1990: 44) postula que as definições terminológicas não ocorrem isoladas, funcionando como parte de uma especificação semântica. Com a expressão *especificação semântica de um termo*, o autor infere que a actualização do significado dos termos advém do seu uso real. Existem, entretanto, diversas categorias de fontes documentais de referência que podem atestar o uso real dos termos: tema de classificação, definição, relação com outros termos, contexto e uso. Cada uma destas categorias pode ser estudada em detalhe, mas nunca isoladamente, já que na realidade elas se complementam.

A delimitação de contextos é um processo semântico, pois avalia a estrutura do conhecimento, sendo possível estabelecer as relações que um termo tem em relação aos outros. Na concepção de Sager (1990: 45) as definições podem estar relacionadas com o contexto, uma vez que este pode conter termos ou expressões que ocorrem na definição podendo, mesmo, fornecer uma definição por implicação ou exemplo (usando o termo num contexto explicativo). Uma definição pode, igualmente, de acordo com o mesmo autor (1990: 47), diferenciar de forma clara o novo termo a partir dos termos que o rodeiam, sobretudo os seus co-termos, sejam eles genéricos, partitivos ou relacionados de qualquer outra forma. É, pois, consentâneo que a definição serve para explicar a intensão de um termo e que as relações terminológicas estabelecem a sua extensão.

“Relationships” (Sager, 1990: 29-35) e “Terminological definitions and relationships” (Sager, 1990: 51-3)

Igualmente consideradas como excelente fonte de recrutamento de mais termos, definições e outro tipo de informação terminológica são as relações conceptuais existentes entre os termos de um domínio de especialidade. Sager (1990: 51) recorda

que, por norma, a definição reúne não só as características essenciais que um conceito tem em comum com outros, mas também as características que o diferenciam de outros conceitos. Também por norma, mas em contraste, as relações indicam o tipo de ligação que um conceito tem com outros conceitos no sistema. Estes dois tipos de informação complementar auxiliam na identificação de termos equivalentes. Nas palavras de Sager (1990: 52-3), a complementaridade entre definição e relação vai mais além já que a maioria das definições refere outros termos, normalmente descritos como hiperónimos (termos genéricos), hipónimos (termos específicos) e co-hipónimos (termos associados). Tais definições apresentam estas relações dentro de uma dada área do conhecimento especializado.

Sager (1990: 29-35 e 52) enuncia algumas relações conceptuais que considera relevantes para a pesquisa terminológica entre as quais destacamos as que contêm elementos e definição:

- relações genéricas, estabelecadoras de uma ordem hierárquica, identificam conceitos pertencentes à mesma categoria, onde existem conceitos do mais genérico, ou superordenado, ao mais específico, ou subordinado(s).

Este tipo de relação é expresso através de fórmulas linguísticas como:

X is a type of A.

X, Y and Z are types of A.

A has the specific concepts X, Y and Z.

A has the subtype X.

(Sager, 1990: 30)

- relações partitivas, também apelidadas de relações *todo-parte* indicam a ligação entre conceitos, consistindo por mais do que uma parte e as suas partes constituintes.

As fórmulas linguísticas que as expressam são:

X is a constituent part of Y.

X, Y and Z are constituent parts of A.

A consists of X.

A consists of X, Y and Z.

Sager (1990: 32)

Expressões como “*x é um tipo de y*” (termo genérico), “*x é parte de y*” (termo específico) “*x está relacionado com y*” (termo associado) são passíveis de ser detectadas de forma automática por programas informáticos que convertem este tipo de informação numa definição simples (Sager, 1990: 53). Este princípio é igualmente aplicável a uma grande variedade de relações.

3.3.3 A proposta de Pearson (1998)

“Retrieval of candidate terms” (1998: 125-8) e “Refining the term identification process” (1998: 128-34)

Pearson (1998: 128-134) propõe um mecanismo de padrões sintácticos de restrição como critério de refinamento para identificação e extracção de candidatos a termos combinando a *generic reference* (*referência genérica*²²) com os *linguistic signals* (*marcadores linguísticos*²³). Este processo pode ser utilizado para confirmação do candidato a termo como termo, bem como para ajudar a detectar candidatos a termo menos óbvios. A referência genérica é entendida pela autora como um critério obrigatório, ou seja, o candidato a termo tem que ter referência genérica. Nesse caso o termo, que a autora denomina como “*flagged term*” (*termo reconhecido*²⁴), deverá ser precedido por um artigo definido ou mesmo não ser precedido por qualquer tipo de artigo. Em contraponto, Pearson (1998: 128) apresenta o “*unflagged term*” (*termo não reconhecido*²⁵), que pode ser precedido por qualquer um dos determinantes, com a excepção do artigo indefinido, sendo provável que, nestes casos, a referência seja específica.

Esta autora (1998: 134) conclui que o recurso a este critério traz resultados positivos, pois permite identificar alguns candidatos a termos, embora outros fiquem por identificar. A autora propõe, então, a associação da referência genérica a marcadores linguísticos específicos e adequados à pesquisa terminológica, de forma a anular as insuficiências do primeiro critério.

Os marcadores linguísticos, apontados por Pearson (1998: 125-34) como elementos sugestivos de definição, ocorrem sob diversos moldes. Podem apresentar-se

²² Nossa tradução.

²³ Idem.

²⁴ Idem.

²⁵ Idem.

na forma nominal, de que são exemplo “(...) *system, procedure, function, method* (...)” (Pearson, 1998: 125), aos quais podemos acrescentar *técnica, processo, propriedade, classe* e *aparelho*. Como exemplo, a autora (1998: 131-4) descreve e analisa alguns marcadores linguísticos associados à referência genérica, já que Pearson é bastante assertiva quando refere que a conjunção destes marcadores juntamente com a referência genérica, confirma a presença de um termo.

Entre esses marcadores Pearson (1998: 131-4) destaca:

- *called (is/are called, is/are often called, also called, usually called, generally called, sometimes called);*
- *known as (also known as, sometimes known as, generally known as, commonly known as known as);*
- *e.g.*, que tanto pode aparecer no corpo principal do texto, como entre parênteses.;
- *the term;*
- “...”. A este marcador é assinalada a dupla particularidade de ajudar a identificar e extrair um neónimo de um texto de especialidade, mas por outro lado pode significar que o termo utilizado entre aspas não é o termo correcto, mas apenas o seu sinónimo da língua corrente, mais acessível ao público em questão.

Em forma de conclusão, podemos referir que os dois critérios sugeridos por Pearson (1998: 125-134), isoladamente, possibilitam a detecção de candidatos a termos. Contudo, a sua coocorrência traz mais segurança e precisão em relação à quantidade e qualidade de candidatos a termos apresentados. A lista das combinatórias possíveis pode ser relativamente longa, sobretudo no caso português devido às desinências que lhe são intrínsecas.

3.3.4. A proposta de Conceição (2001 e 2004):

“Termes et Reformulations” (2001) e “Sequências de reformulação” (2004)

Tendo por base as ideias postuladas na sua tese de Doutoramento “*Termes e Reformulations*” (2001), Conceição (2004) apresenta sucinta e especificamente, no

Seminário “*Corpora, Sequências de Reformulação e Extracção Terminológica*” o que entende por sequência de reformulação:

Sequência discursiva composta por uma formulação, um conector (marcador) reformulativo e por uma segunda formulação em que a segunda formulação: fixa os limites de significação da primeira; contribui para o processo dinâmico que é a aquisição de conceitos (ajuda à apropriação de conceitos ou de denominações); mostra relação entre interlocutores.

Conceição (2004: s/ página)

Este autor aponta essencialmente a existência de quatro causas que estão na base da necessidade de reformulações (para além do facto de contribuírem para a coesão e coerência textuais): quando o discurso não se encontra adequado ao universo referencial, na *impossibilidade de repetição*, na *necessidade de retoma* e pelo facto de funcionar como *explicação, contestação e exemplificação* (Conceição, 2004: s/ página).

Sintomaticamente, o recurso a reformulações coloca em destaque algumas diferenças, que Conceição (2004: s/ página) sintetiza da seguinte forma, dando exemplos:

- “no interior do discurso: como diz X, no sentido de X;
- entre as palavras e as “coisas”: designação incorrecta
- as palavras e o significado mais esperado: em sentido lato
- o enunciador e o receptor: vulgarmente chamados”

Na mesma linha de Sager (1990: 30-2), também Conceição (2004: s/ página) chama a atenção para as relações conceptuais ao afirmar que há “*semantismos comuns*” e “*semantismos diferenciais*”, dando o exemplo de relações simples como $[X=Y]$, $[X>Y]$, $[X<Y]$ e $[X\neq Y]$. Um dos efeitos deste tipo de *redes semânticas* é a construção de conceitos em discurso ao que nós acrescentamos, extrapolando, que contribuem, igualmente, para a construção de definições terminológicas.

Conforme apontado por Conceição (2004: s/ página) em situação de discurso, as reformulações concretizam-se através de recursos linguísticos como as “*anáforas, catáforas (o termo retomado preceder o termo em expansão), paráfrases, repetições totais ou parciais, comparações, enumerações e metáforas*”. Nestas situações contam-

se exemplos como: *assim como, bem como, isto é, ou seja, apesar de, por outras palavras, dito de outra forma*, etc.

Neste sentido, ao procedermos à análise de reformulações, estamos, na realidade a analisar parte da moldura contextual com dois objectivos essenciais: a “*extracção de termos*” e a “*extracção de informação com interesse terminológico*”. Essa análise processa-se detectando “*indícios de reformulação, marcadores de reformulação, marcas da construção do discurso, marcas da construção / representação dos conhecimentos no discurso*” (Conceição, 2004: s/ página). Em relação à informação com interesse terminológico, Conceição (2004: s/ página) avança a seguinte tipologia: informações sobre construção discursiva, denominação e conceitos.

Da informação do tipo de construção discursiva fazem parte “*verbos que introduzem definição, expressões que introduzem outros verbos no discurso e explicação*” (Conceição, 2001: 265 e 2004: s/ página), como por exemplo, os verbos *ser, chamar, designar, denominar, significar, definir*, etc. Relativamente à informação do tipo de denominação fazem parte a “*presença de variantes ou de abreviaturas, a explicitação de mudanças conceptuais ou denominativas, a não assunção de responsabilidade das afirmações, reivindicação de autoridade e informações relacionadas com o nível de língua*”. Finalmente, da informação do tipo conceptual fazem parte as “*informações sobre o processo de categorização / conceptualização e sobre relações lexicais, informações sobre harmonização e informações sobre o domínio ou sobre domínios conexos*”.

Concretamente sobre a utilização da informação de interesse terminológico, a mesma representa um importante contributo em áreas como “*a estruturação conceptual*”, através de relações lexicais; a “*didáctica*”, através do nível de língua e vulgarização; e, por fim, as “*redacções técnica e de definições*”, através da precisão conceptual e terminológica e nível de língua (Conceição, 2004: s/ página).

Conceição (2004: s/ página) advoga que, na moldura terminológica, as sequências de reformulação desempenham seis funções: “*categorização, verbalização, representação, referenciação, regulação e comunicação*”, contribuindo para uma “*concepção dinâmica da significação*”, uma “*abordagem onomasiológica e semasiológica dos dados linguísticos e terminológicos recolhidos em corpora*” e “*encarar a terminologia no âmbito das ciências da linguagem*”.

3.3.5. A proposta de Colussi (2002)

“Recursos da reescritura da informação científica” (2002: 40-9)

Na sua Dissertação de Mestrado *“A reescritura da informação científica em textos de popularização da Ciência”*, Colussi (2002) avança alguns recursos linguísticos utilizados com o objectivo de facilitar a compreensão dos textos pelo público-alvo.

Os marcadores linguísticos (nomeadamente os de definição) contribuem, de acordo com Colussi (2002: 38-40), de forma explícita para a coesão das relações semânticas, sobretudo quando o conhecimento sobre o conceito em causa é escasso ou nulo. É, assim, possível, nos textos de popularização da ciência haver, para além de sinónimos com menor grau de especialização e de fenómenos linguísticos como a abundância de conjunções que inferem relações lógicas, expressões de definição de termos e metáforas que permitem ao usuário conjugar a informação que acabou de adquirir com outra que já detinha. A este propósito, Colussi (2002, *in* Pagano, 1998: 58-60) chama a atenção para estes elementos cuja função comunicativa serve de elo de ligação entre registo técnico e científico dirigido a especialistas e a sua versão de registo mais acessível dirigido ao público em geral. Assim é o caso dos artigos de especialidade e dos manuais universitários, em que na primeira situação o nível de informação é muito mais detalhado sem, contudo, haver tanta necessidade de explicar os conceitos, enquanto na segunda, essa necessidade existe pois de outra forma os textos não serão compreendidos pelo público-alvo.

Tendo por finalidade confirmar a utilização de estratégias de reescrita para tornar mais acessível ao leigo o discurso científico, Colussi (2002: 42-8) apresenta quatro recursos de reescrita: a metáfora, a personificação, a glosa e o aposto, aplicando-as aos nove textos que constituem o seu corpus estudo.

No estudo de Colussi (2002: 42) a metáfora é definida como *“um significado ou ideia usada para representar um segundo significado ou ideia, com o objectivo de sugerir uma analogia ou semelhança entre os dois.”* Esta autora (2002: 42) avança essencialmente duas razões para o uso de metáforas na Popularização da Ciência: economizar palavras, isto é, significar mais por menos palavras e familiarizar o leitor não-especializado com o conteúdo científico tornando-o, por exemplo, num discurso mais concreto e acessível.

Quanto à personificação, ela é apresentada por Colussi (2002: 46, *in* Lakoff & Johnson, 1980: 33) como *“uma metáfora onde o objecto físico é largamente*

especificado como sendo uma pessoa.”, ou seja, um objecto ou uma abstracção adquire características próprias dos seres humanos.

No que se refere à glosa, Colussi (2002: 47, *in* Kopple, 1985: 82) entende que esta funciona como uma anotação, observação ou comentário, prestando auxílio ao leitor para que este atinja os significados apropriados presentes no texto. *“Glosa é a palavra usada para definir, explicar ou delimitar, desde que não se amplie o conteúdo proposicional do texto, mas auxilie o leitor a entendê-lo e interpretá-lo. Portanto, a glosa tem uma função metatextual, ou seja, a capacidade de se auto explicar.”* (Colussi, 2002: 47, *in* Kopple, 1985: 82). Neste sentido e no enquadramento do seu trabalho, Colussi (2002: 47) denomina glosa a todas os lexemas ou expressões explícitos de que são exemplo *“tal/tais como” / “such as” e “ou seja” / “that is”*.

Por último, o aposto corresponde a uma informação ou um conjunto de informações acessórias que, embora não sendo indispensáveis, contribuem para o esclarecimento do enunciado ao determinar ou caracterizar o conceito de um termo. Este recurso linguístico pode apresentar-se representado numa frase ou fazendo parte de uma frase. Neste último caso, o aposto situa-se entre vírgulas, assumindo um carácter explicativo, enumerativo ou até recapitulativo, com a finalidade de expandir ou explicar um conceito científico ao público-alvo.” (Colussi, 2002: 48).

Conforme nos é dado a compreender através das propostas dos cinco autores por nós escolhidos e apresentados ao longo do sub-capítulo 3.3., existem inúmeras fórmulas que permitem detectar a presença de informação terminológica sob a forma de termos, bem como alguns dos seus elementos caracterizadores. A partir da reflexão proposta neste capítulo, julgamos ter ficado patente a importância atribuída aos contextos, sobretudo no que concerne à obtenção de informação terminológica relevante.

No próximo Capítulo pretendemos apresentar a nossa proposta relativamente aos marcadores de informação terminológica como meio de extracção de conteúdos terminologicamente relevantes.

4. Os marcadores lexicais indiciadores de informação terminológica e os padrões suporte

Conforme constatámos no sub-capítulo 3.3. (*Extracção de informação terminológica relevante: termos, definições, contextos de uso, sistemas conceptuais e redes semânticas*), a detecção e extracção automáticas ou semi-automáticas de informação terminológica através de um instrumento como o *Corpógrafo* pode ser potenciada através da utilização de recursos linguísticos empregues no discurso de especialidade. O nível de adequação dos modelos de comunicação ao público-alvo (por exemplo, especialista/especialista, especialista/não-especialista,...) por parte dos produtores de textos pode suceder através de diferentes classificações linguísticas, modelos, fórmulas ou combinatórias que se encontram disponíveis em cada sistema linguístico, regulando, por exemplo, o grau de tecnicidade da sua produção textual. No caso de um não-especialista ou simples leigo, será necessário desenvolver uma competência terminológica para além das competências linguística, discursiva e comunicativa, que necessariamente já deve dominar. Da mesma forma que a competência linguística significa dominar o código e as regras de funcionamento da língua geral, também a competência terminológica se traduz no domínio do código e das normas de uma língua de especialidade.

No âmbito da terminologia e mais especificamente da análise e extracção terminológicas, podemos definir os marcadores lexicais de informação terminológica e os *padrões suporte* como recursos linguísticos que têm como objectivo principal a obtenção de informação terminologicamente útil, rigorosa e consistente num enquadramento de língua de especialidade. Essa informação contribui para a descrição, definição, explicação e associação de um termo a um conceito. Esse conceito poderá ser mais ou menos definido, consoante seja um objecto, processo, acontecimento, ou qualquer outro tipo de descrição conceptual. Tanto na língua portuguesa, como na inglesa, estes marcadores e padrões permitem perceber a presença de informação terminológica variada, o que configura, aliás, uma situação ideal para pesquisas terminológicas. Podemos, no entanto, apontar uma pequena diferença entre os dois: os *padrões suporte* têm como principal função detectar termos e definições, enquanto, os marcadores de informação terminológica têm uma função mais alargada: não só a detecção de termos e definições, mas também de contextos de uso e relações semânticas ou conceptuais.

Neste sentido, julgamos pertinente colocar a seguinte questão: de que forma, ou em que medida, os marcadores de informação terminológica e *padrões suporte* contribuem para a detecção de candidatos a termos, definições e outro tipo de informação terminológica? É o que veremos seguidamente.

4.1. Os marcadores lexicais indiciadores de informação terminológica

Uma das conclusões que podemos expressar após a análise do conjunto de propostas defendidas por Dubuc (1992), Sager (1990), Pearson (1998), Conceição (2001 e 2004) e Colussi (2002), que considerámos adequado destacar a propósito do conteúdo do nosso trabalho, é a de que o investigador, para adquirir o maior número possível de informação terminológica, dispõe de valiosos recursos linguísticos existentes nos próprios textos de especialidade.

Os marcadores lexicais indiciadores de informação terminológica, nomeadamente os que antecipam a probabilidade da presença de elementos concorrentes à elaboração de uma definição (doravante *marcadores lexicais de definição*, como abreviatura) constituem, porventura, o recurso linguístico ao nível dos marcadores, que mais contribuem (qualitativa e quantitativamente) para a obtenção de informação necessária a estudos e a produtos terminológicos.

Antes de avançarmos com a nossa exposição sobre este tipo de marcadores, achamos importante neste ponto clarificar a diferença que entendemos existir entre os denominados marcadores linguísticos e os marcadores lexicais. Enquanto dos últimos fazem apenas parte os elementos lexicais de um sistema linguístico (palavras, expressões, etc.), nos primeiros se incluem, para além dos elementos lexicais, elementos não-lexicais, como é o caso da pontuação e dos sinais tipográficos, assim como elementos pictóricos. Uma vez que o *Corpógrafo* não está preparado para apresentar resultados de pesquisa, baseados em *padrões suporte* exclusivamente de pontuação, conforme sugerido, por exemplo, por Pearson (1998: 131-4), cf. ponto 3.3.3., termo entre parênteses ou entre aspas.

Este tipo de marcadores lexicais são também especialmente úteis quando considerados como assistentes de uma primeira abordagem numa pesquisa terminológica, proporcionando ao utilizador resultados bastante frutíferos. Hoje em dia, um leigo poderá iniciar o estudo de um determinado tema ou domínio de especialidade,

via *World Wide Web*, digitando num qualquer motor de busca um único termo representativo desse domínio em associação a um marcador lexical indiciador de informação terminológica para, numa questão de segundos ser apresentado no seu monitor um número de termos suficiente que lhe permita efectuar um primeiro esboço do mapa conceptual do domínio em estudo. A um nível mais especializado, o mesmo acontece se utilizarmos um *corpus* textual criteriosamente construído para que, juntamente com um programa informático adequado, possamos proceder à extracção de informação, sob a forma de unidades terminológicas, definições, contextos de uso ou qualquer outro tipo de informação que a ferramenta permita extrair. A partir desse primeiro esboço, o utilizador poderá partir para o aprofundamento e refinamento da sua pesquisa.

Conforme já referimos anteriormente, um termo pode ser utilizado em associação com os marcadores lexicais de definição com o objectivo de representar um conceito, esclarecendo e delimitando o significado de um determinado termo, podendo assumir a forma de palavras, expressões, sintagmas ou até fórmulas e combinatórias linguísticas.

Como combinatórias lexicais recorrentes, apresentam-se geralmente sob formas fixas, mas sobretudo, semi-fixas. Damos como exemplo:

“Denominados por *_termo_*”

“Designada por *_termo_*”

Pelo facto de uma das características destes marcadores lexicais ser a recorrência, eles são, até certo ponto, previsíveis facilitando a sua detecção pois, existem diversas possibilidades de fórmulas, sobretudo as indiciadoras de definição, que variam mais na escolha lexical, do que no significado. A recorrência dos marcadores, combinada com a sua previsibilidade, tornam mais perceptível a probabilidade de existência de candidatos a termos e/ou definições. Disto são exemplo:

“*_termo_* pode definir-se como ...”

“Aquele a que se chama de *_termo_* ...”

Existem marcadores que são indiciadores de relações conceptuais ou semânticas, o que os torna preciosos contributos na detecção de termos, estabelecendo simultaneamente o seu lugar no sistema conceptual. Vejamos os seguintes exemplos:

“As *_termo_* são constituídas por ...”

“Os *_termo_* constituem um componente de ...”

“O *_termo_* é formado pelo ...”

“A *_termo_* deriva de ...”

Sobre a constituição dos marcadores lexicais de definição, podemos facilmente verificar que, de uma forma geral, são compostos por um núcleo, podendo adquirir formas expandidas aumentando o número de elementos que as compõem. A consequência, porém, é que o seu significado pode apresentar-se alterado. Por exemplo, a frase “As *cheias são um Risco Natural.*”, não tem exactamente o mesmo significado da frase “As *cheias podem ser um Risco Natural.*” Ou até mesmo “As *cheias poderiam ser consideradas um Risco Natural.*”, pelo que na utilização de marcadores para extracção terminológica devem ser escolhidos aqueles que apontam assertividade.

Os marcadores de definição podem, assim, ser compostos por mais do que uma palavra, assumindo formas de verbos, locuções verbais, sintagmas verbais ou nominais, etc.²⁶ Funcionam como identificadores de termos, por vezes de forma bastante directa e precisa, outras de forma mais indirecta, vaga ou até redundante. No universo terminológico, poderíamos comparar os marcadores lexicais de definição a pistas de um mapa que nos guiam ao “local x” onde está o tesouro: os termos, os seus sinónimos ou até mesmo às suas definições. Estes tipos de marcadores são especialmente válidos em situações de maior especificidade ou opacidade textual e terminológica, já que fornecem informação verdadeiramente útil quando o investigador se vê confrontado com um vocábulo que numa primeira análise vacila entre considerá-lo termo ou não-termo, em casos como *Hidrograma de cheia* e *Hidrograma da cheia*.

É, aliás, nossa convicção que o grau de sucesso na determinação de termo ou não-termo aumenta significativamente se os marcadores lexicais de definição se encontrarem devidamente contextualizados. Esta afirmação vai ao encontro da ideia de que um termo devidamente contextualizado dá mais garantias do seu valor terminológico, para além de atestar o seu uso. Podemos afirmar que os marcadores lexicais de definição ou indiciadores de definição funcionam como uma forma de refinamento do processo de identificação de termos, definições e outro tipo de informação contextual, de que são exemplo as relações entre conceitos. Pensamos que quanto mais específico e conceptualmente preciso for um domínio, maior será o número

²⁶ Dado a classificação morfológica não ser um dos objectivos deste trabalho, a mesma não será aprofundada. No ponto 4.1.1. será feita referência especial aos verbos e grupos verbais.

de relações lógicas e ontológicas (hierárquicas, associativas e partitivas), partilhadas entre eles.

Uma ideia parece certa: nas línguas vivas, como é o caso do português e do inglês, os marcadores lexicais de definição acompanham as variações semânticas dos conceitos que, na sua vertente linguística, podem ocorrer de uma forma súbita (pela imposição de uso) ou mais prolongada no tempo (pelo uso corrente). Com o aprofundamento da especificidade do conhecimento, estas alterações semânticas tendem a tornar-se mais específicas por via da intensão, que no aspecto linguístico se traduz num aumento efectivo no número de elementos que compõem a unidade terminológica. Como exemplo ilustrativo do que acabámos de referir, atentemos ao seguinte termo (válido para as línguas portuguesa e inglesa) retirado do *Glossário de Internacional de Hidrologia*²⁷:

PT 1052 Torrencialidade de uma cheia por fusão da neve

Velocidade de subida e descréscimo do nível das águas em cheias provocadas pela fusão da neve na Primavera.

EN 1052 rush of snowmelt flood

Rate of water rise and fall during spring snowmelt flood.

No caso do nosso estudo, os marcadores lexicais de definição fazem parte de um conjunto de recursos linguísticos susceptíveis de serem utilizados na detecção e extracção de candidatos a termos, definições e outro tipo de informação terminológica, facilitando a compreensão do texto pelo público-alvo. No Capítulo 6 (*Análise e Apresentação de Resultados*) é nosso objectivo demonstrar que, numa pesquisa terminológica semelhante à proposta neste trabalho, constitui uma vantagem a utilização dos marcadores lexicais indiciadores de informação terminológica e conseguir obter, simultaneamente, candidatos a termos, elementos para as respectivas definições, contextos de uso, relações conceptuais ou semânticas, através do recurso a textos e contextos de especialidade.

4.1.1. O uso de verbos e grupos verbais como marcadores lexicais de definição

Na definição de Cunha *et al.* (1999: 377), o verbo é “(...) *uma palavra de forma variável que exprime o que se passa, isto é, um acontecimento representado no tempo*

²⁷ Disponível em: <http://www.cig.ensmp.fr/~hubert/glu/PT/GF1052PT.HTM>, [acedido em 2005.08.20].

(...)”. No caso dos marcadores lexicais indiciadores de definição que se apresentam sob a forma de verbos e grupos verbais, podemos afirmar (tal como pretendemos demonstrar no Capítulo 6) que, tanto no sistema linguístico português, como no inglês, existem várias possibilidades de combinatórias ou sequências lexicais que permitem ao investigador perceber se está perante um termo e respectiva definição ou, pelo menos, se está perante informação suficiente que lhe permita conferir estatuto de candidato a termo.

No caso específico da língua portuguesa “*O verbo apresenta as variações de NÚMERO, de PESSOA, de MODO, de ASPECTO e de voz.*” (Cunha et al., 1999: 377). Estas formas flexionadas ou de significação adicional oferecem possibilidades combinatórias mais extensas do que na língua inglesa.

Tanto os verbos, como os grupos verbais, são os elementos lexicais mais explicitamente reveladores da existência de termos e de definições, justamente devido à sua carga referencial. Quando utilizados com o propósito comunicativo e informativo, são por norma esclarecedores, desambiguadores e precisos.

Se tivermos em consideração exemplos como os verbos *ser / (to) be*, *definir / (to) define* ou *significar / (to) mean*, e de grupos verbais como *ser chamado / (to) be called*, *ser composto por / (to) be composed of*, *poder ser descrito como / can be described as*, apercebemo-nos que estamos muito perto da informação pretendida. Tal acontece devido ao facto de, salvo algumas excepções, os verbos e os grupos verbais não se apresentarem de forma independente ao seu referente. Por este motivo, são portadores de informação que relacionam a realidade linguística à conceptual, prevenindo, ao mesmo tempo, fenómenos como a polissemia (existência de mais de um objecto ou conceito para um mesmo termo).

4.1.2 O caso específico do verbo *SER / (to) BE*

Tendo presente o que anteriormente referimos sobre o contributo dos verbos e grupos verbais para o sucesso na detecção e extracção de termos, definições e demais informação terminológica, pensamos que a resposta à segunda questão colocada no início deste Capítulo (quais os marcadores lexicais indiciadores de informação terminológica de definição mais eficazes) se encontra no verbo *ser / (to) be*. Conceição (2001: 263) assevera que o verbo *ser* é o mais representativo na relação entre a denominação e os elementos de definição textual.

Como indiciador da presença de informação terminológica, o verbo *ser / (to) be* pode actuar:

- sozinho, como verbo principal:
é/ são, is/are
- em associação, como verbo auxiliar, constituindo expressões ou sequências verbais
 - locução verbal: verbo auxiliar + verbo principal
é/ são defined[a/as/o/os], is/are defined
pode(m) ser definid[a/as/o/os], can be defined
 - grupo verbal:
verbo auxiliar + verbo principal + preposição
é/ são constituíd[a/as/o/os] por, is/are composed of
verbo principal + preposição
constituíd[a/as/o/os] por, composed of

Em relação a esta última hipótese de acção como verbo auxiliar, enunciada, aliás por Conceição (2001: 264-5), o verbo *SER / (to) BE* é omitido, isto é, não se encontra representado no discurso, pelo que a expressão ou sequência verbal funciona explicitamente apenas com o particípio passado do verbo principal seguido, ou não, de uma preposição.

4.1.3 Os marcadores lexicais indiciadores de informação terminológica e o *Corpógrafo V2*

Nas páginas seguintes incluímos quatro quadros onde elencamos os marcadores lexicais indiciadores de informação terminológica mais comuns, detectados em contextos naturais de uso nos textos que constituem os *corpora* do nosso estudo.

Apenas como nota de esclarecimento, referimos que os elementos que se encontram dentro dos parênteses rectos representam as várias possibilidades combinatórias com os restantes elementos que compõem o que designamos por marcador lexical. Temos, assim, os seguintes quadros de marcadores lexicais indiciadores de informação terminológica:

MARCADORES LEXICAIS INDICIADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA PORTUGUESA - SINGULAR]	
1.	[a o] <i>termo</i> [apresenta constitui contém corresponde define deriva tem]
2.	[o a] <i>termo</i> é [o a um uma]
3.	[o a] <i>termo</i> é constituído/a [por pela]
4.	[o a] <i>termo</i> é formado/a
5.	[o a] denominado/a <i>termo</i> , caracterizado/a por
6.	a noção de <i>termo</i>
7.	[aquele aquela] a que se chama de <i>termo</i>
8.	[conhecido/a] como <i>termo</i>
9.	designa-se [de por] <i>termo</i> [ao à]
10.	distinção entre <i>termo</i> e <i>termo</i>
11.	é [designado/a denominado/a] [de por] <i>termo</i>
12.	[o a] <i>termo</i> [apresenta contém corresponde deriva é]
13.	[o a] <i>termo</i> é constituído [pelo/a por]
14.	[o a] <i>termo</i> é definido como
15.	[o a] <i>termo termo</i> encerra

Quadro 1 – Marcadores lexicais indiciadores de informação terminológica [Língua Portuguesa – Singular]

MARCADORES LEXICAIS INDICIADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA PORTUGUESA – PLURAL]	
1.	<i>termo</i> [constituem definem são têm]
2.	<i>termo</i> podem definir-se como
3.	[os as] <i>termo</i> [apresentam contêm correspondem derivam recebem]
4.	[os as] <i>termo</i> são
5.	[os as] <i>termo</i> são [as os]
6.	[os as] <i>termo</i> são constituídos/as [por pelos/as]
7.	[os as] <i>termo</i> são definidos/as como
8.	[os as] denominados/as <i>termo</i> ,
9.	conhecidos/as como <i>termo</i>
10.	Consideram-se <i>termo</i> [os as aqueles/as]
11.	designam-se [de por] <i>termo</i>
12.	[Entenda-se Entende-se] <i>termo</i> como [os as uns umas]
13.	chamados/as <i>termo</i>
14.	que são denominados/as <i>termo</i>
15.	são designados/as [de por] <i>termo</i>

Quadro 2 – Marcadores lexicais indiciadores de informação terminológica [Língua Portuguesa – Plural]

MARCADORES LEXICAIS INDICIADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA INGLESA – SINGULAR]	
1.	<i>term</i> can be [characterised defined as described as regarded as]
2.	<i>term</i> [conducts contains controls has occurs represents plays]
3.	<i>term</i> is [a an the]
4.	<i>term</i> is a [kind of / part of]
5.	<i>term</i> is also known as
6.	<i>term</i> is [composed of determined by part of]
7.	<i>term</i> is [located in made in the made [up] of used to refers to serves as tells about function as]
8.	is called [a the] <i>term</i>
9.	It can be described as <i>term</i> .
10.	is known as <i>term</i>
11.	the <i>term</i>
12.	the <i>term</i> can be defined as
13.	the <i>term</i> [conducts contains controls has results]
14.	the <i>term</i> is [composed of consists of constituted by function as located in made of results from serves as]

Quadro 3 – Marcadores lexicais indiciadores de informação terminológica [Língua Inglesa – Singular]

MARCADORES LEXICAIS INDICIADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA INGLESA – PLURAL]	
1.	<i>term</i> are [a an the]
2.	<i>term</i> are a system of
3.	<i>term</i> are [composed of constituted by defined as determined by initiated [also] known as located in made in the made [up] of part of released the those parts of occur [when] used to]
4.	<i>term</i> can be ... [characterised defined as described as regarded as]
5.	<i>term</i> [conduct consist of contain control function as have are where]
6.	<i>term</i> [refer to represent result from serve as]
7.	are called [the] <i>term</i>
8.	[are] known as <i>term</i>
9.	called <i>term</i>
10.	in [the] <i>term</i> [a an the]
11.	like (<i>outro termo</i>), [the] <i>term</i>
12.	the <i>term</i> can be [defined described] as
13.	<i>term</i> [occur when result from serve as tell about]
14.	these <i>term</i> are termed <i>term</i> .

Quadro 4 – Marcadores lexicais indiciadores de informação terminológica [Língua Inglesa – Plural]

Queremos ainda salientar que temos consciência que a nossa descrição de marcadores lexicais indiciadores de informação terminológica não é exaustiva, mas podemos assegurar que estes são os marcadores mais comuns obtidos em contexto terminológico da área de especialidade das *Cheias: Riscos Naturais*.

Os marcadores lexicais acima listados não representam a sua totalidade, uma vez que apenas se referem à realidade específica deste estudo, ou seja, correspondem aos marcadores lexicais manualmente extraídos por nós a partir dos contextos de uso fornecidos pelo *Corpógrafo V2*. Uma lista mais completa de marcadores linguísticos (incluindo os lexicais), poderá ser consultada no Anexo I. De referir que esta lista é aberta, dado que permite uma constante actualização (sobretudo inclusão) de marcadores.

4.2 Os Padrões suporte e o Corpógrafo V3

Apresentamos em seguida duas listas de *padrões suporte* utilizados pelo *Corpógrafo V3* na detecção de termos e definições em contextos naturais de uso nos textos que constituem os *corpora* do nosso estudo.

Temos, desta forma, os seguintes quadros de *padrões suporte*:

PADRÕES SUPORTE UTILIZADOS PELO CORPÓGRAFO V3 [LÍNGUA PORTUGUESA]	
1.	(^ ^o ^a ^um ^uma o a um uma) <i>termo</i> é (uma um o a)
2.	(^ ^entende-se ^designa-se ^denomina-se entende-se designa-se denomina-se)(por de como) <i>termo</i>
3.	(^ ^denomina-se ^designa-se ^chama-se denomina-se designa-se chama-se) <i>termo</i>
4.	(^ ^é ^é)(o a) <i>termo</i> \.
5.	(designad denominad chamad)(o a) <i>termo</i>
6.	(^ ^no ^na ^em no na em) <i>termo</i> (o a um uma)
7.	(^ ^os ^as os as) <i>termo</i> (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuam funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
8.	<i>termo</i> : .*
9.	(chamad designad denominad)(os as) <i>termo</i>
10.	(^ ^os ^as os as) <i>termo</i> que (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuam funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam se dão originam causam acontecem trazem resultam segregam englobam descrevem induzem se devem exprimem consistem crescem circulam permitem fazem transmitem dependem)
11.	- <i>termo</i> -
12.	(^ ^desencadeia ^causa ^provoca desencadeia causa provoca)(os as) <i>termo</i>
13.	, (os as) <i>termo</i> \.

Quadro 5 – Padrões Suporte utilizados pelo Corpógrafo V3 [Língua Portuguesa]

PADRÕES SUPORTE UTILIZADOS PELO CORPÓGRAFO V3 [LÍNGUA INGLESA]	
1.	([^] [^] called [^] named called named) <i>termo</i>
2.	([^] [^] the [^] a [^] an the a an) <i>termo</i> is (initiated protected secreted calculated considered spread defined released produced described regarded viewed)
3.	([^] [^] the [^] a [^] an [^] each the a an each) <i>termo</i> (results releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings consists corresponds leads depends takes receives derives plays)
4.	([^] [^] like like).* , the <i>termo</i>
5.	<i>termo</i> : (a an the)
6.	([^] [^] causes causes)(a an the) <i>termo</i>
7.	([^] [^] the [^] a [^] an the a an) flood <i>termo</i> (that which also) (prevents is holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives releases plays)
8.	([^] [^] during during)(the a an) <i>termo</i>
9.	([^] [^] groups [^] sets [^] types [^] classes [^] categories groups sets types classes categories)of <i>termo</i>
10.	([^] [^] the the) <i>termo</i> (are can be should be may be) (initiated protected secreted calculated considered spread surrounded constituted formed composed characterised characterized defined described regarded viewed used located expressed situated placed protected bathed released produced understood seen divided)
11.	([^] [^] the the)floods (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
12.	([^] [^] like like).* , the <i>termo</i>
13.	([^] [^] are are)the <i>termo</i>
14.	([^] [^] when when) <i>termo</i> occur
15.	([^] [^] cause cause) <i>termo</i>

Quadro 6 – Padrões Suporte utilizados pelo Corpógrafo V3 [Língua Inglesa]

À semelhança do que foi referido para os marcadores, os *padrões suporte* acima listados não representam a totalidade dos *padrões* previstos pelo *Corpógrafo*, já que apenas se referem à realidade específica deste estudo, ou seja, são os *padrões suporte* utilizados automaticamente pelo *Corpógrafo* para detectar e extrair informação terminológica dos *corpora* por nós criados no domínio do conhecimento específico por nós escolhido. No Anexo I poderá ser consultada uma lista mais completa de *padrões*

suporte extraída do *Corpógrafo* V3 em 2006.01.30. Esta lista é aberta, dado que permite uma constante actualização (sobretudo inclusão) de *padrões*.

Neste Capítulo apresentámos a nossa proposta para obtenção de candidatos a termos, definições, contextos de ocorrência, sistemas conceptuais ou redes semânticas, entre outros. através da utilização de marcadores lexicais indiciadores de informação terminológica em geral e de marcadores lexicais de definição em concreto, a partir de pesquisas automáticas proporcionadas pelo *Corpógrafo*. O nosso objectivo não foi o de listar exhaustivamente os marcadores lexicais existentes, mas tão só elencar alguns dos marcadores mais representativos no processo de obtenção de informação terminológica.

Entramos em seguida na II Parte do nosso trabalho, o **Estudo Empírico**, onde começaremos por descrever as etapas metodológicas percorridas para a realização da nossa pesquisa ao que se seguirá a análise e apresentação dos resultados obtidos.

II PARTE

ESTUDO EMPÍRICO

5. Metodologia

Na componente prática deste estudo pretendemos conhecer a extensão do contributo, numa pesquisa terminológica, dos marcadores lexicais indiciadores de informação terminológica, com especial destaque para os marcadores lexicais de definição, assim como os *padrões suporte* propostos pelo *Corpógrafo*. Para tal, entendemos como necessário estabelecer uma metodologia de trabalho que permitisse criar condições para ensaiar, no *Corpógrafo*, a viabilidade de alguns dos referidos marcadores e *padrões suporte* na tarefa de extracção de termos, suas definições e outras informações de importância terminológica, tais como contextos de uso, redes semânticas e sistemas conceptuais. Conforme já referimos na Introdução, o nosso trabalho apresenta uma descrição comparativa entre as reais possibilidades de extracção de informação terminológica a partir dos meios e elementos que o *Corpógrafo V2* (ou seja, Versão 2, disponível até Setembro de 2005) oferecia em relação ao *Corpógrafo V3* (a sua nova Versão 3, em funcionamento desde Setembro de 2005). A Versão 2 possibilitava a extracção manual de marcadores de informação terminológica, enquanto a Versão 3 apresenta, de forma automática, na área da pesquisa de definições, o que designa por *padrões de suporte*. É a partir destas evidências que iremos tentar compreender e avaliar o contributo dos marcadores e dos *padrões suporte* na pesquisa terminológica.

Apresentamos, desde já, as seis etapas que percorremos, de forma a atingir os objectivos inicialmente propostos neste estudo:

- ETAPA 1 - Definição dos objectivos do estudo
- ETAPA 2 - Escolha e descrição dos instrumentos de trabalho
- ETAPA 3 - Pesquisa bibliográfica: recolha, selecção e apreciação dos textos
- ETAPA 4 - Organização do domínio e representação do mapa conceptual
- ETAPA 5 - Constituição e apreciação dos *corpora* e das Bases de Dados no *Corpógrafo*
- ETAPA 6 - Processo de extracção de informação terminológica a partir dos *corpora* para análise e apresentação dos resultados

Descreveremos, em seguida, o conteúdo de cada uma das etapas do nosso estudo.

5.1. Etapa 1 – Definição dos objectivos do estudo

O primeiro passo traduziu-se no estabelecimento dos objectivos a atingir neste estudo empírico. Em seguida, assumimos que o nosso instrumento base de trabalho seria o *Corpógrafo*, dado termos verificado com satisfação que se tratava da ferramenta capaz e adequada ao nosso estudo. Com a ajuda do *Corpógrafo* procedemos à construção de dois *corpora* (um em português de Portugal [PT], outro em inglês [EN]) exclusivamente com textos do domínio dos *Riscos Naturais* e *Cheias / Natural Hazards* e *Floods*, de forma a tirarmos proveito das opções *Centro de Conhecimento: Pesquisar definições em corpora, usando restrições lexicais* e *Pesquisa: Concordância Frase*.

Tomando como ponto de partida os métodos de obtenção de informação terminológica (termos, definições, contextos e relações conceptuais e semânticas) propostas pelos autores Dubuc (1992), Sager (1990), Pearson (1998), Conceição (2001 e 2004) e Colussi (2002), bem como a possibilidade de utilização do *Corpógrafo*, avançámos com a nossa própria proposta dos marcadores lexicais indiciadores de informação terminológica. Neste sentido, começámos por delinear a metodologia para atingir os objectivos propostos, sob a orientação da Professora Doutora Belinda Maia (Departamento de Estudos Anglo-Americanos) e da Professora Doutora Fantina Pedrosa (Departamento de Geografia), da Faculdade de Letras da Universidade do Porto.

5.2. Etapa 2 – Escolha e descrição dos instrumentos de trabalho

5.2.1. O *Corpógrafo*²⁸

Ao longo dos últimos anos temos assistido a um crescimento exponencial na disponibilidade de recursos informáticos dedicados à vertente empírica da investigação linguística, com importantes repercussões na forma como é efectuada a pesquisa terminológica. Disso mesmo é exemplo o *Corpógrafo* como ferramenta integrada e baseada na *World Wide Web* (WWW), desenvolvida pelo Núcleo do Porto da Linguateca²⁹, associada à Faculdade de Letras da Universidade do Porto. O seu intuito inicial foi o de criar uma estrutura sistematizada que permitisse a recolha e armazenamento de textos sob a forma de *corpus* com a possibilidade da sua

²⁸ Disponível em: <http://www.linguateca.pt/corpografo/>

²⁹ Para informação mais detalhada consultar: <http://www.linguateca.pt/corpografo/>

reutilização. O seu acesso é livre e o utilizador não necessita possuir qualquer tipo de aptidão técnica especial ou *software* adicional para desenvolver o seu trabalho de investigação (Sarmiento *et al.*, 2004: 1).

A “*Introdução à Utilização do Corpógrafo – Um Pequeno Tutorial*” define o Corpógrafo como “*uma plataforma destinada à análise e ao trabalho sobre corpora com o objectivo essencial de produzir diversos recursos linguísticos em suporte informático.*”³⁰ Este documento chama a atenção para o facto de o Corpógrafo reunir num único ambiente de trabalho várias funcionalidades que habitualmente se encontram dispersas, entre as quais destacamos, a título de exemplo, a compilação e organização de corpora, a pesquisa de concordâncias, bem como a extracção de candidatos a termos a partir de Bases de Dados previamente constituídas e de outro tipo de informação terminológica relevante (coocorrências, contextos com informação conceptual e semântica).

O Corpógrafo é composto por quatro áreas de trabalho:

1. *Gestor*, que se subdivide em *Ficheiros*, *Corpora*, *Autores*, *Organizações* e *Repositório multimédia* que tem como funções principais a preparação e organização dos textos, bem como a gestão dos corpora;
2. *Pesquisa*, que permite o estudo dos corpora através de três tipos de *Concordância* (*Frase*, *Janela* e *KWIC*) e o *Estudo de N-Gramas*;
3. *Centro de Conhecimento*, que se subdivide em *Base de Dados Terminológicas* e *Gestor de Relações*. É a primeira destas opções que possibilita a pesquisa terminológica (candidatos a termos) e a realização de inúmeras operações como, gestão da informação terminológica (listar e editar termos), pesquisa de definições (candidatos a definições de termos), e ainda a gestão e organização do conhecimento através da pesquisa e gestão de relações semânticas.
4. *Centro de Comunicação*, que contém toda a documentação existente sobre o Corpógrafo (ajuda, tutoriais e publicações).

De especial valor são as três grandes possibilidades oferecidas pelo Corpógrafo: reutilização dos recursos informáticos nele armazenados, produção de documentação e, ainda, auxílio a actividades em torno da língua (como é o caso da tradução). As bases de dados e as listas de termos entretanto criadas permitem armazenar e gerir uma rede

³⁰ Para informação mais detalhada consultar: <http://www.linguatca.pt/corpografo/>.

complexa de elementos que representam a sistematização do conhecimento linguístico e conceptual de uma área de conhecimento específico, conforme nos é explicado no tutorial.

Na realidade, o *Corpógrafo* detecta e destaca unidades lexicais pertencentes a textos de especialidade denominados *candidatos a termos*. Só posteriormente, mediante um processo de selecção e confirmação de um determinado conjunto de características (executado pelas partes envolvidas no processo terminológico), é que podemos apresentar as referidas unidades lexicais como termos.

No presente estudo, parte dos marcadores lexicais indiciadores de informação terminológica, em especial os de definição, foram identificados pelo *Corpógrafo* na opção *Centro de Conhecimento: Pesquisar definições em corpora - Usando restrições lexicais* ao mesmo tempo que listava candidatos a termos e a definições. Como complemento e aprofundamento da nossa pesquisa, recorreremos igualmente à opção *Pesquisa: Concordância Frase*, já que verificámos, após alguns testes, ser a opção complementar mais adequada para atingir os objectivos propostos no nosso estudo. A partir destes elementos fornecidos automaticamente, procedemos a uma análise mais detalhada do contributo dos referidos marcadores para obtenção de termos, definições e outra informação terminológica, conforme teremos oportunidade de demonstrar no Capítulo 6, deste trabalho.

5.2.2. Escolha e introdução ao domínio conceptual

Antes de entrarmos na descrição propriamente dita do domínio conceptual proposto para o nosso estudo, apresentamos algumas breves considerações teóricas sobre o conceito de *conceito*.

Consideramos que a Terminologia encara os conceitos e os termos como esferas independentes: os conceitos são vistos como unidades pré-linguísticas de conhecimento, isto é, podem existir sem que lhe correspondam previamente símbolos ou signos linguísticos. Embora os conceitos sejam possuidores dos traços semânticos que os caracterizam, eles não representam o significado. É, por isso, compreensível que o terminólogo, numa abordagem onomasiológica, parta do conceito de forma a fazer-lhe corresponder o(s) termo(s) que se lhe ajustem (ao contrário do lexicólogo que busca o conceito a partir da unidade linguística, optando assim pela abordagem semasiológica).

Existe, no entanto, uma diferença entre conceito da língua geral e conceito da língua de especialidade. Enquanto no primeiro caso os elementos semânticos

caracterizadores da sua representação linguística (a palavra) são, regra geral, polissémicos e conotativos, no segundo caso é praticamente o oposto. Em terminologia, os conceitos são unidades de conhecimento caracterizadas pela factualidade, tecnicidade e cientificidade, sem hipótese de inclusão de elementos conotativos e emotivos.

Quanto à escolha do domínio de especialidade, *Riscos Naturais: Cheias*, que a norma ISO/FDIS 1087-1:2000 (3.1.2) descreve como sendo um ramo especializado do conhecimento, surgiu no seguimento de uma proposta conjunta de colaboração entre o Departamento de Estudos Anglo-Americanos / Mestrado de Terminologia e Tradução (2002/2003) e o Departamento de Geografia da Faculdade de Letras. Desta forma, como introdução aos conceitos da área do conhecimento do projecto, assistimos a uma apresentação feita pela orientadora da área de Geografia, Professora Doutora Fantina Pedrosa. Mais tarde, foram realizadas reuniões, uma geral com o grupo constituído para o Projecto dos *Riscos Naturais* e outras individuais para esclarecimentos de dúvidas entretanto surgidas.

Foi-nos assim explicado que se tratam de fenómenos que resultam de processos naturais relacionados com a geodinâmica interna (sismos, erupções vulcânicas, *etc.*), ou externa (cheias, incêndios, tempestades, *etc.*) característicos do planeta Terra geralmente causadores de danos, variando em grau a sua extensão e a sua intensidade. O seu processo de desencadeamento pode ser previsto, mas nunca controlado, embora seja possível criar infraestruturas e outro tipo de medidas para minimizar o seu impacto. Nesse sentido são implementadas medidas de prevenção e mitigação dos riscos passando inicialmente por um processo de avaliação da probabilidade de ocorrência do fenómeno natural e da vulnerabilidade. Antes (caso seja possível prever), durante e imediatamente após o fenómeno, existem medidas que poderão ser tomadas no enquadramento de um processo de gestão articulado entre várias instituições e organizações e que inclui um alerta (com diferentes níveis), o desencadeamento da actuação de toda uma estrutura organizada de protecção, a avaliação dos prejuízos (humanos e materiais) e a reabilitação e reconstrução das áreas atingidas.

Uma das principais características da área dos *Riscos Naturais* é o seu carácter interdisciplinar. Com efeito, este domínio da Geografia Física está intimamente relacionado não só com áreas como a Hidrologia, a Climatologia/Meteorologia e a Geomorfologia, mas também com disciplinas fora do âmbito da Geografia, como é o caso da Hidráulica (Engenharia) e de diversas Ciências Sociais, cujo envolvimento surge sempre que a Natureza interage com o Homem. Na realidade, todas as áreas

estudadas pela Geografia Física estão incondicionalmente ligadas umas às outras pelo simples motivo de fazerem parte da própria geodinâmica da Terra. No caso mais específico das *Cheias*, trata-se de um fenómeno natural que apesar de sempre ter existido, tem apresentado causas e, sobretudo, consequências cada vez menos pacíficas e provocadas por uma crescente intervenção humana no meio natural.

5.2.3. O especialista e o seu contributo

A apreciável complexidade das relações entre domínio, sub-domínio e áreas associadas, colocou alguns embaraços na percepção geral da rede intrincada de funcionamento entre os vários domínios conceptuais. Por esse motivo considerámos que, logo neste ponto, a comunicação com um especialista era fundamental para o sucesso do trabalho. Como tal, e como garante do pleno domínio dos dados terminológicos da área do conhecimento em estudo foi desenvolvido um trabalho de parceria com um especialista na área em questão, cujo conhecimento especializado não foi reduzido à simples complementaridade informativa.

Com efeito, a escolha do domínio dos *Riscos Naturais* e do sub-domínio das *Cheias* como tema de estudo desta dissertação surgiu na sequência de uma proposta de colaboração efectuada pelo especialista desta área, no decurso do primeiro ano do Mestrado em Terminologia e Tradução, tendo em vista a elaboração de um glossário terminológico multilingue de *Riscos Naturais*. Esta cooperação foi desde logo assumida não só pela nossa parte, como também por parte da Professora Doutora Fantina Pedrosa, Professora Auxiliar do Departamento de Geografia da Faculdade de Letras da Universidade do Porto, nas áreas científicas de Geografia Humana, de Estudos de População e Desenvolvimento e de Riscos Naturais. Foi absolutamente necessário conhecer minimamente o funcionamento dos fenómenos nas suas vertentes geográfica, técnica, organizacional, social e temporal para, posteriormente, definir o seu posicionamento com vista a organizar e hierarquizar o sistema conceptual em estudo. O seu contributo, marcado pelo rigor científico indispensável num trabalho terminológico e terminográfico, como o pretendido, foi determinante na orientação ao nível bibliográfico e, sobretudo, ao nível conceptual, sem esquecer outros elementos indispensáveis a uma percepção mais alargada da área em estudo.

5.2.4. As línguas de trabalho

As línguas de trabalho escolhidas para compilação de textos são o português de Portugal [PT] e o inglês [EN], dado o interesse da utilização de sistemas linguísticos diferentes, nomeadamente ao nível lexical e sintáctico.

O facto de termos optado exclusivamente pela utilização da variante do português de Portugal assenta apenas em dois critérios.

O primeiro prende-se com o facto de o país que mais contribui com textos livremente acessíveis na Internet – o Brasil – optar, por vezes, por uma escolha lexical e terminológica, bem como por uma construção sintáctica diversa do português de Portugal, sobretudo em domínios técnicos e científicos especializados, conforme observado por Maia (2003b: s/ página). A título de exemplo, refira-se a situação dos termos *cheia* (para o caso de Portugal) e *enchente* (para o caso do Brasil): os dois termos são sinónimos e existem tanto no português de Portugal, como na sua variante brasileira, no entanto, cada país privilegia o seu uso conforme indicado. A autora refere, ainda, como factor de afastamento entre as duas variantes, a influência do inglês (devido à proximidade dos Estados Unidos da América) na adopção de termos, sobretudo mais recentes, e posterior adaptação à língua portuguesa (Maia, 2003b: s/ página). Este processo de nomeação de conceitos diversos nos dois países deve-se, sobretudo, ao facto de cada país ter a sua própria perspectiva das ciências e técnicas.

O segundo motivo surge numa óptica de equilíbrio de representatividade das variações da língua portuguesa. Seria, pois, relevante para a recolha terminológica a inclusão de todas as variantes linguísticas que fazem parte da esfera da Comunidade dos Países de Língua Portuguesa (CPLP), caso o objectivo final deste trabalho fosse, por exemplo, a elaboração de um glossário de *Riscos Naturais: Cheias* em língua portuguesa. Porém, à parte os casos de Portugal e Brasil, a produção e disponibilização de textos no domínio pretendido são escassas ou até mesmo inexistentes.

Esta situação já não é tão premente na língua inglesa, apesar de existirem algumas diferenças nomeadamente a nível ortográfico, fonológico e até certo ponto lexical e sintáctico, mas que não têm influência na análise dos dados no presente estudo. Na realidade, na língua inglesa, não foram detectadas diferenças ao nível terminológico do domínio escolhido.

5.3. Etapa 3 – Pesquisa bibliográfica: recolha, selecção e apreciação dos textos

Nesta terceira etapa concentrámo-nos na realização das tarefas ligadas à recolha, selecção e apreciação dos textos nas línguas portuguesa e inglesa para constituição dos respectivos *corpora*.

5.3.1. Recolha e selecção dos textos

Por norma, a transmissão de conhecimento especializado não se faz exclusivamente através da utilização de terminologias (de que são exemplo os glossários). Os textos de especialidade, como é o caso dos técnicos e científicos, englobam não só os termos utilizados numa determinada área de conhecimento, mas também os respectivos contextos de uso. Estes textos obedecem a regras de construção discursivas que, para a finalidade de uma pesquisa terminológica, aportam uma dupla funcionalidade: para além de fornecerem informação terminológica ao nível quantitativo e qualitativo em elevado grau de concentração, por oposição aos textos de uso corrente, prestam um inestimável auxílio na detecção de informação terminológica menos óbvia ou directa.

No âmbito de uma língua de especialidade é, desde logo, fundamental perceber e iniciar o domínio dos conceitos-chave envolvidos. Nesse sentido, a pesquisa bibliográfica correspondeu, numa primeira fase, à leitura de textos de Linguística Geral e sobretudo de Linguística Aplicada, Lexicologia/Lexicografia e Terminologia/Terminografia, com a consequente aquisição de conhecimentos necessários à fundamentação teórica e metodológica do estudo, a partir dos quais foi possível formular a hipótese apresentada neste trabalho.

A pesquisa bibliográfica para construção dos dois *corpora*, traduziu-se, fundamentalmente, na recolha de textos de edições e publicações disponíveis na Biblioteca Central e no Departamento de Geografia da Faculdade de Letras e, ainda, através de pesquisas realizadas na *World Wide Web* (WWW), de forma a obter material terminologicamente válido para os *corpora* especializados que nos propusemos construir. Como resultado, foram encontrados textos “tradicionais” (Maia, 1997: s/ página) impressos, tais como livros, manuais de estudo, publicações periódicas, artigos científicos, técnicos e académicos, relatórios técnicos, glossários e documentos de

informação ao público disponíveis na WWW, geralmente relacionados com a Protecção Civil.

A tarefa de recolha de textos na língua portuguesa revelou-se particularmente difícil, não só devido à escassez de textos de produção nacional, publicados ou disponíveis na WWW, adequados à construção dos *corpora*, mas também à necessária filtração qualitativa. Reconhecemos que, por estes dois motivos, no que se refere aos textos reunidos em língua portuguesa, apesar de provirem de fontes credíveis, são pouco variadas. Tal deve-se ao facto de não existir material disponível desta área, seja em bibliotecas gerais e especializadas, seja na WWW. Apesar deste contratempo, foi possível reunir um pequeno conjunto de textos ainda assim suficientes e, sobretudo, representativos da área conceptual para levar a bom termo o estudo proposto. Logo à partida verificámos que a esmagadora maioria dos textos recolhidos de maior interesse e qualidade para o nosso estudo não se encontravam disponíveis no formato digital, pelo que foi necessário proceder à sua digitalização, tarefa que não se revelou fácil e que consumiu bastante tempo.

De acordo com Maia (1997: s/ página), tentar seleccionar um conjunto representativo de textos é, em geral, um dos primeiros objectivos de quem constrói um *corpus*. No presente caso, a escolha dos textos teve por base critérios como abundância terminológica (qualitativa e quantitativa), abundância de definições e de contextos (bem como todo o tipo de informação terminológica interessante e relevante para o estudo), inclusão de textos apenas na língua original (excluindo, por conseguinte, traduções) e reconhecimento qualitativo dos autores, entre outros.

A recolha do material escrito foi direccionada para obtenção de textos exclusivamente representativos do sub-domínio das *Cheias* e em menor quantidade dos *Riscos Naturais*, com diferentes graus de especialização, aptos a fornecer o maior número possível de informação terminológica.

Em relação à tipologia dos textos seleccionados, houve da nossa parte, uma forte aposta na inclusão de textos científicos e técnicos retirados de obras didácticas, como manuais académicos, artigos de divulgação e comunicações de conferências da especialidade, bem como de textos procedentes de enciclopédias, dicionários especializados e de glossários, produzidos por autores e instituições internacionais, nacionais e regionais reconhecidos na área em questão, o que se revelou bastante produtivo e trouxe enormes vantagens para a nossa pesquisa. No Anexo II, Quadros 15

e 16 são descritos todos os textos que compõem os *corpora* utilizados no presente estudo.

5.3.2 Apreciação dos textos

Logo após a fase de recolha, procedemos a uma leitura mais atenta dos textos de especialidade pré-seleccionados relativos ao domínio e sub-domínio propostos neste estudo. Esta tarefa possibilitou, não só, a apreensão de alguns conceitos de uso mais corrente por parte dos especialistas, como também a percepção do sistema conceptual de especialidade em causa. Entretanto, foi inevitável lidar com problemas de vocabulário e de estilo, bem como de idiomatismos e com a necessidade de adquirir um conhecimento mais aprofundado nesta área técnica de especialização.

De referir, ainda, que alguns dos textos não representam a totalidade da obra, constituindo apenas longos excertos, cuidadosamente seleccionados. A exclusão de elementos provocadores de “ruído” no processo de extracção terminológica, como bibliografias, imagens (fotografias e desenhos), quadros (dados estatísticos), fórmulas matemáticas e respectivas legendas, para além de não colocarem em risco a coesão e coerência dos textos que integraram os *corpora*, contribuem para um melhor funcionamento do *Corpógrafo* e consequente fiabilidade dos resultados finais apresentados. A esse propósito Habert *et al.* (2000, in Berland, 2002), entendem que esses textos devem ser, antes de mais, suficientemente homogêneos para que o tratamento automático aplicado produza um resultado fiável.

Em relação ao texto de especialidade, Hoffmann (1988, in Silva, 2004: 2) tal como outros autores de abordagens mais contemporâneas, defende que pertence à linguagem natural, diferenciando-se desta através de alguns recursos linguísticos, bem como de alguns objectivos comunicativos. Já Silva (2004: 2) define linguagem especializada como uma sublinguagem pertencente, portanto, à linguagem natural, sendo composta por um conjunto de recursos linguísticos utilizados em âmbito comunicativo delimitado por uma especialidade para garantir a compreensão daqueles que nela trabalham.

Neste sentido, conforme é possível verificar através da listagem dos textos incluídos nos *corpora* (cf. Anexo II), todos os textos se incluem na categoria de textos de especialidade, variando apenas o grau de especialização (traduzido na prática através das diferentes situações comunicativas devidamente seleccionadas). Os textos incluídos nos *corpora* foram categorizados e agrupados conforme as diferentes situações

comunicativas seleccionadas. Neste estudo propomos os seguintes modelos comunicativos:

Textos científicos e semi-científicos

Correspondem a textos informativos, terminologicamente abundantes e geralmente elaborados por especialistas para especialistas. Este tipo de textos tem desenvolvido significativamente o aspecto ligado à investigação e à teorização da área de especialidade, pelo que o seu grau de conceptualização pode ser bastante elevado.

Textos técnicos e semi-técnicos

Correspondem a textos informativos, elaborados por especialistas e destinados a um público específico e, por vezes, não-especialista. Incluem textos elaborados em forma de relatórios, manuais e guias, razão pela qual estão mais ligados ao aspecto pragmático da área de especialidade.

Tanto os textos científicos e semi-científicos, como os técnicos e semi-técnicos, respeitam, aquando da sua elaboração, um conjunto de características relativamente padronizadas e detectáveis com relativa facilidade, tais como uma maior utilização de nomes compostos, de plurais e de verbos nominalizados (Ahmad *et al.* 1997: 726). Estes textos pressupõem a existência de características como a objectividade, especificidade, concisão, clareza e, sobretudo, precisão para que os conceitos nele veiculados se apresentem de forma devidamente estruturada e adequada ao seu público-alvo. Essa adequação reflecte, por sua vez, a existência de determinados elementos nos quais se incluem a terminologia, o estilo, a quantidade e, sobretudo, a qualidade da informação presente nos textos. Este conjunto de elementos e características têm como resultado uma *performance* comunicativa eficaz.

Enquadrados numa dada especialidade, o seu grau de especificidade é, por vezes, tão elevado que são poucos os pontos de contacto com textos do mesmo domínio, mas vocacionados para um público-alvo menos conhecedor. Geralmente, os níveis semi-científico e semi-técnico destes textos permitem, através da análise contextual, a compreensão dos conceitos presentes, mesmo por parte do público mais leigo.

Textos didácticos

As produções científica e técnica estão frequentemente ligadas à produção didáctica, que funciona como veículo eficaz de divulgação. O seu conteúdo inclui

terminologia específica da área de conhecimento abordada (como por exemplo as palavras-chave), definições, bem como o tipo de informação necessária à construção do sistema conceptual e da rede semântica. Estamos de acordo com a posição de Conceição (2003: s/ página), quando se refere aos textos didácticos típicos como “(...) *um género ou tipo textual, tende a apresentar, na sua maioria, repetições, retomadas explicativas e/ou reformulações (...)*”. É neste meio que é possível detectar elementos que conduzem a informação terminológica. O autor observa, igualmente, que são indicadores como “*termo, conceito, definição, sinónimo, expressão, também chamado, em sentido restrito, vulgarmente denominado, de conectores assim como, isto é, ou seja, por exemplo*” (Conceição, 2003: s/ página), entre muitos outros, que podem contribuir para a percepção e identificação de informação terminológica relevante. A importância dos textos didácticos é, igualmente, destacada por Stubbs (1996: 67) ao afirmar que os manuais, os Professores e o Ensino possuem autoridade reconhecida e requerida para este tipo de pesquisa.

Por todos estes motivos, demos especial relevo ao nível da sua representatividade nos *corpora* do nosso estudo.

Glossários

Em relação aos glossários, podemos referir que estão presentes nos *corpora* essencialmente com dois graus de especialização: os que se enquadram nos textos didácticos de vulgarização científica e os de semi-técnicos que servem de complemento a manuais e guias de acção no âmbito da Protecção Civil.

Finda esta tarefa de apreciação dos textos, seleccionámos em definitivo aqueles que passariam a constituir os dois *corpora*.

5.4. Etapa 4 – Organização do domínio e representação do mapa conceptual

Uma leitura mais atenta dos textos seleccionados em ambas as línguas para integrarem os *corpora* e de uma análise mais reflectida sobre a terminologia neles contida, permitiu uma ampliação da moldura conceptual no sentido de preparar e constituir as futuras Bases de Dados. Procedemos, assim, à organização e estruturação

do conhecimento e posteriormente à representação em diagrama do modelo conceptual – cf. Figura 1. O conhecimento inicial do sistema conceptual *Riscos Naturais: Cheias e Natural Hazards: Floods* foi adquirido e complementado por vezes em simultâneo com a execução das tarefas de pesquisa bibliográfica, recolha e apreciação de textos, conforme relatado na etapa anterior.

Na realidade, apesar da proveitosa introdução efectuada pelo especialista da área, houve necessidade de ultrapassar a carência de um conhecimento mais aprofundado no domínio específico escolhido para este estudo. Neste sentido, foi efectuada uma criteriosa pesquisa e recolha bibliográfica (conforme descreveremos na etapa seguinte). Quanto à organização do sistema conceptual de domínio ela está, de certa forma, patente na própria concepção dos *corpora* e permite uma visão bastante genérica do domínio; em cada língua, o *corpus* agrupa-se em duas classificações – *Riscos Naturais / Natural Hazards*, com especial predominância nos Riscos Hidro-meteorológicos e subdivide-se em *Cheias / Floods*.

Dado que os textos e as suas características são determinantes no processo de elaboração de um *corpus* de domínio especializado, foram incluídos textos referentes ao domínio geral dos *Riscos Naturais* (com especial incidência nos Riscos Hidro-meteorológicos) e ao sub-domínio das *Cheias*. Uma vez que num sistema conceptual os conceitos não se manifestam de forma isolada (pois estão ligados através de uma teia de relações), também as áreas do conhecimento humano não se apresentam hermeticamente estanques; é inevitável a existência de ligações entre elas, umas mais robustas, outras mais ténues, conforme os casos. Devido a esta interoperabilidade entre o domínio e o sub-domínio do nosso estudo, quer no plano conceptual, quer no plano linguístico, bem como na sua representação discursiva neste projecto, entendemos ser adequado proceder à fusão de ambos num único *corpus* (um para cada língua), com claro aproveitamento das respectivas sinergias. No nosso caso, esta opção permite, igualmente, o enriquecimento quantitativo e qualitativo, por exemplo, da lista de termos a extrair dos *corpora* textuais.

A organização do domínio permitiu-nos, pois, proceder à estruturação da rede de relações conceptuais e posteriormente à representação em diagrama do modelo conceptual, conforme representado na Figura 1:

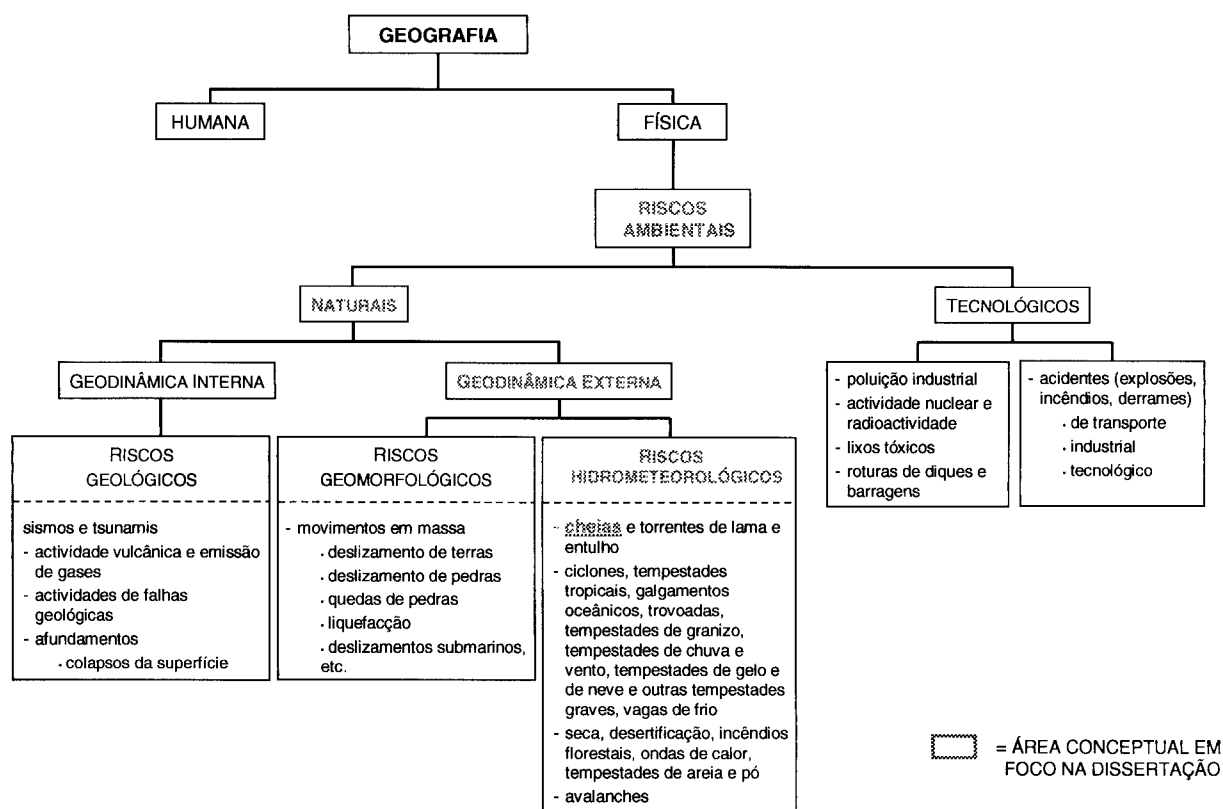


Figura 1 – Representação em diagrama do modelo conceitual

Na prática, a organização conceptual na qual assenta a estrutura dos *corpora* deste estudo é muito semelhante a esta representação. Existem, contudo, inúmeras ligações e relações colaterais que, por motivos também já mencionados, não justificavam a expansão do número das bases de dados do estudo, mas que inequivocamente contribuem para o aumento do número de informação terminológica referente a esta temática. São disso exemplo as três áreas ligadas às acções a desenvolver perante a existência de Riscos Naturais: Mitigação/Prevenção, Gestão da Crise e Recuperação.

Em síntese, foi essencialmente com base na transmissão de conhecimentos do especialista e na pesquisa e consulta dos textos desta área de conhecimento que nos foi possível criar um modelo conceptual, aqui representado e estruturado em forma de diagrama (*cf.* Figura 1). Consideramos este procedimento imprescindível não só para definição de hierarquias e dependências coerentes, para a formalização das relações e

interacções entre os diversos elementos que alicerçam ontologias, assim como para a representação e estruturação lógica do conhecimento adquirido.

Dado que o Projecto de Geografia – *Riscos Naturais* – é mais amplo e engloba outros sub-domínios para além das *Cheias* (sismos e *tsunamis*, movimentos de massa, secas, etc.), o diagrama proposto deixa abertura para o aprofundamento da granularidade. Importa ainda referir que a área conceptual em foco no nosso estudo se encontra destacada a verde.

5.5. Etapa 5 – Constituição e apreciação dos *corpora* e das Bases de Dados no Corpógrafo

Conforme a seguir descrito, nesta etapa procedemos à *construção dos corpora* textuais nas línguas portuguesa e inglesa e à constituição das respectivas Bases de Dados no Corpógrafo.

5.5.1. Constituição e apreciação dos *corpora*

A utilização de *corpora* em formato digital revolucionou os métodos da pesquisa terminológica, uma vez que, juntamente com o programa informático adequado, é possível extrair não só termos, mas também informação contextual, que ajuda a precisar e situar um termo no sistema conceptual a que pertence. Permite, ao mesmo tempo, dar a perceber o funcionamento de um dado termo num dado contexto. Nesse sentido, a aplicação de *corpora* na área dos *Riscos Naturais / Cheias* e *Natural Hazards / Floods* permitiu-nos a extracção de inúmeros elementos válidos, desde termos e definições a contextos reais e ilustrativos do uso da terminologia do domínio em estudo.

Os *corpora Riscos Naturais/Cheias* e *Natural Hazards/Floods* têm a dimensão total aproximada de 144.000 palavras para o corpus português e 143.000 palavras para o *corpus* em inglês. Cada um dos *corpora* é constituído por textos de discurso escrito respectivamente nas línguas portuguesa ou inglesa. Cronologicamente, a produção textual situa-se entre 1992 e 2004, com a excepção de um texto do *corpus* inglês, produzido em 1964 e, ainda, de alguns textos que não fornecem esta informação (em ambos os *corpora*).

Tendo presente a hipótese colocada neste estudo, não é nossa intenção apresentar exaustivamente nos dois *corpora* todos os textos mais representativos da área de especialidade. As características intrínsecas do estudo permitem que os *corpora* propostos sejam apenas representativos da área temática em questão. Conforme já aclarado, no caso de se dar seguimento ao futuro projecto terminográfico de construção de um glossário multilingue, então já se justificaria, e até se aconselharia, a inclusão de um maior número de textos, sobretudo se estes forem cronologicamente mais recentes.

De acordo com o referido na Etapa 3, foi dada especial importância à selecção dos textos considerando a nossa pretensão de elaborarmos dois *corpora* (um em português e ou outro em inglês) de pequena dimensão e com algum grau de refinamento. Neles procurámos incluir, dos domínios já identificados, textos científicos, técnicos, académicos e glossários. Desta forma, foram excluídos, por exemplo, textos genéricos e artigos de imprensa escrita não-especializada que não contêm terminologia aprofundada sobre os domínios, ou se a têm, utilizam-na por vezes inadvertidamente.

Os objectivos e as funções de um *corpus* são, de acordo com Austermühl (2001:127-8), inúmeras, estando directamente dependentes do interesse específico de quem o construiu e/ou de quem o vai utilizar. Este autor refere, como exemplo, que um arquivo de textos científicos pode ser, simultaneamente, ferramenta de pesquisa para um cientista e uma preciosa fonte terminológica para um tradutor. Tendo presente esta ideia, atribuímos grande relevância à qualidade dos textos a incluir nos *corpora*. Como tal, a maioria dos textos seleccionados foram retirados de livros e artigos académicos e técnicos de autoria de Professores universitários e de técnicos qualificados nas áreas ligadas aos *Riscos Naturais* e *Cheias*, conforme é possível verificar através do Anexo II, onde estão incluídos os Quadros 15 e 16, referentes às listas de textos dos *corpora* comparáveis. É neste contexto que Maia (2002b: 2) refere:

“The corpora required (... for this type of research) need to consist of texts that are not just well written, in the sense that they represent texts normally produced in a particular domain of knowledge: they need to use terms that are generally accepted in the community that works in that domain.”

(Maia , 2002b: 2)

5.5.2. Constituição e apreciação das Bases de Dados

Antes de procedermos à pesquisa de candidatos a informação terminológica, foi necessário construir as Bases de Dados no *Corpógrafo*. Foram criadas apenas duas

Bases de Dados, uma para cada par de línguas (na língua portuguesa, *Riscos Naturais / Cheias* e, na língua inglesa, *Natural Hazards / Floods*) para as quais foram remetidos os termos seleccionados por entre os inúmeros candidatos automaticamente propostos pelo *Corpógrafo*.

Estamos cientes que a tipologia dos *Riscos Naturais* e *Cheias* não se esgota no seu conceito mais geral. No entanto, achámos que uma maior granularidade implicaria *corpora* de maior dimensão, bastante superior a 144.000 palavras por língua. Acrescente-se que, apesar de este projecto concreto incidir sobre o sub-domínio das *Cheias*, foi deixada abertura suficiente no modelo de representação conceptual para mais tarde adquirir maior extensão e granularidade no caso de ser necessário incluir os restantes sub-domínios referentes aos *Riscos Naturais*.

5.6. Etapa 6 – Processo de extracção de informação terminológica a partir dos *corpora* para análise e apresentação dos resultados

Como é consensualmente aceite, um elemento isolado do seu contexto não revela muita informação. O significado de um elemento deverá ser, desta forma, interpretado em sintonia com os outros elementos que com ele ocorrem e que não correspondem necessariamente àqueles que se posicionam imediatamente antes ou depois do elemento em questão.

Tendo presente esta ideia, entendemos que existem genericamente duas formas de detecção e extracção de candidatos a unidades terminológicas, bem como de outros tipos de informação terminológica:

- a directa, através das análises morfológica, sintáctica e semântica, utilizando características intrínsecas do termo, e
- a indirecta, através da análise contextual, utilizando marcadores linguísticos exteriores ao termo.

No caso do nosso estudo, concentrámo-nos no método indirecto. Para tal, escolhemos como *expressões de pesquisa*, marcadores lexicais indiciadores de informação terminológica e *padrões suporte*, para actuarem nos *corpora* de especialidade previamente preparados de modo a renderem candidatos a termos, definições e outra informação terminológica relevante. Existem *expressões de busca*

que são mais óbvias do que outras, podendo corresponder a unidades monolexémicas ou multilexémicas. Na primeira situação equivalem, na esmagadora maioria, a verbos e, na segunda, a grupos verbais que, por sua vez podem constituir fraseologias, expressões fixas, coocorrências, etc.

É também possível chegar a informação terminológica diversa através de indicadores menos óbvios, como é o caso da pontuação: entre aspas [...], entre parênteses [...], dois pontos [:], entre vírgulas [, ...], etc.. No entanto, estes exemplos constituem um recurso pouco viável, sobretudo na pesquisa automática, já que apresentam propostas com bastante “ruído”.

Gostaríamos, ainda, de realçar o facto de este processo de obtenção de informação terminológica não invalidar a intervenção humana pois, por mais especializado que seja o discurso de uma área, é inevitável o uso de língua geral. E esta inevitabilidade sujeita-nos a fenómenos linguísticos como a polissemia, homonímia, sinonímia, de idiomatismos e até mesmo de alguns vícios da linguagem que podem dificultar o trabalho de pesquisa a pessoas menos avisadas. Para melhor ilustrar o que acabámos de referir, tomemos como exemplos os seguintes casos que temos como sintomáticos:

- no domínio dos *Riscos Naturais* há ainda muita confusão entre os termos *Risco* e *Perigo* / *Risk* e *Hazard*;
- no sub-domínio das *Cheias*, deparamo-nos, igualmente, com uma situação de confusão no uso de *Cheias* e *Inundações* / *Floods* e *Flooding*.

Apresentadas estas considerações iniciais, gostaríamos de realçar que a recolha de dados é uma das etapas mais significativas de um projecto de pesquisa terminológica. Neste sentido, é necessário seguir critérios bem definidos para que essa tarefa seja executada de forma precisa e proveitosa. No trabalho de identificação e extracção de unidades terminológicas, o terminólogo é, por vezes, confrontado com o dilema de considerar determinado lexema como pertencente ao domínio em análise ou à língua geral. Existem alguns procedimentos que ajudam o investigador a determinar se o candidato a termo o é realmente ou se deverá ser rejeitado como não-termo. Um desses procedimentos é a análise contextual, que através da observação dos elementos que envolvem o candidato a termo permite identificar o(s) conceito(s) directamente envolvido(s). Dos vários elementos que compõem os contextos de definição, explicativo e associativo, atribuímos especial destaque aos marcadores lexicais de definição que

melhor resultam no sistema conceptual dos *Riscos Naturais – Cheias / Natural Hazards - Floods*.

Conforme já referimos, a detecção e extracção automáticas ou semi-automáticas de termos e de informação terminológica contextual, podem ser potenciadas através da utilização de recursos linguísticos utilizados no discurso científico e técnico em geral.

Na prática, a extracção de informação terminológica a partir dos *corpora* constituídos no *Corpógrafo* processou-se da seguinte forma:

1º - nas *BD Terminológicas* do *Centro de Conhecimento*, seleccionámos a opção *Pesquisar Termos*;

2º - perante a lista de candidatos apresentada pelo *Corpógrafo*, escolhemos aqueles que julgámos corresponderem a termos e procedemos à sua inserção na *Base de Dados*;

3º - demos início à pesquisa de definições utilizando a opção *Usando Restrições Lexicais*;

4º - perante os exemplos apresentados automaticamente, anotámos os *Padrões Suporte* (que no nosso estudo designamos por *marcadores lexicais indiciadores de informação terminológica*) utilizados pelo *Corpógrafo* para estabelecer a ligação entre os candidatos a termo e definição, propondo estes dois elementos como candidatos às respectivas categorias. Cabe-nos, depois, guardar a definição que julgamos mais acertada, assim como também nos é dada a oportunidade de redigir a definição, se assim o entendermos. A partir desta altura é-nos oferecida a possibilidade de utilizarmos os *Padrões Suporte* (por ele identificados automaticamente) noutra área do *Corpógrafo*. Assim, na *Pesquisa*,

5º - optámos por *Concordância Frase* e testámos na *expressão de pesquisa* alguns marcadores de definição (*Padrões Suporte*), propostos pelo próprio *Corpógrafo*. A informação é apresentada em destaque (negrito) e devidamente contextualizada. Consoante os marcadores introduzidos, varia a informação terminológica obtida a partir dos nossos *corpora*.

Este conjunto de procedimentos realça a extrema importância e dedicação que devemos atribuir aos *corpora*, especialmente na sua fase de construção, mas sobretudo a significativa utilidade dos marcadores de informação terminológica para obtenção de elementos terminológicos menos óbvios, mas que, ao serem apresentados no seu devido contexto, se tornam mais visíveis.

A detecção e recolha destes indicadores foram efectuadas utilizando a área de trabalho do *Centro de Conhecimento do Corpógrafo*, cujas funções já descrevemos. Concretamente, procedemos à realização de operações de pesquisa terminológica (candidatos a termos) e pesquisa de informação contextual destinada às respectivas definições terminológicas, relações conceptuais, entre outros. Para tal, introduzimos como *Expressão de Pesquisa na Concordância Frase*, unidades terminológicas certificadamente pertencentes à área em estudo e, a partir dos resultados obtidos, recolhemos o máximo possível de informação terminológica referente a esses termos para obtermos pistas ou ligações para outros. O *Corpógrafo* possibilita que a pesquisa e extracção se processem de uma forma automática e manual, permitindo assim a intervenção do terminólogo e/ou do especialista em qualquer ponto da operação. Como em todas as ferramentas informáticas vocacionadas para o estudo linguístico, são apresentados resultados em conformidade com os parâmetros escolhidos pelo utilizador, não havendo lugar a qualquer análise de resultados, razão pela qual é ao investigador que cabe a tarefa de interpretação dos dados. A respectiva análise recai sobre os elementos sob investigação, podendo incidir sobre praticamente todo o tipo de informação linguística, nomeadamente no que se refere aos aspectos lexicográficos, morfológicos, sintácticos, semânticos e pragmáticos.

Tendo efectuado a descrição da metodologia utilizada, passamos seguidamente a descrever e a analisar os principais resultados obtidos.

6. Apresentação e análise descritiva dos resultados

Estamos convictos que apenas através do estudo empírico, com a utilização e descrição de dados autênticos, é possível confirmar hipóteses, elaborar teorias credíveis e consolidar outras já apresentadas e defendidas. Foi nesse sentido que nos propusemos ensaiar, nos *corpora* especialmente construídos para este estudo, alguns marcadores lexicais indiciadores de informação terminológica, assim como alguns *padrões suporte*. Era nossa intenção que esta investigação nos permitisse avaliar, não só a exequibilidade, mas também o contributo de ambos (depois de devidamente ajustadas à realidade do *Corpógrafo*), numa pesquisa terminológica com características idênticas à que é proposta neste estudo.

Antes de passarmos à apresentação de alguns exemplos de extracção terminológica, queremos realçar a vantagem que o *Corpógrafo* oferece na utilização dos símbolos [.*] (ponto asterisco) a seguir à raiz da expressão de pesquisa. Este recurso, válido para ambos os sistemas linguísticos deste estudo, permite aceder de imediato às formas variáveis de género, número e categorias gramaticais (nomes, verbos, advérbios, etc.). Por exemplo, a expressão de pesquisa [*signific.**] permite ao *Corpógrafo* apresentar resultados como *significar(em)*, *significante*, *significado*, *significativo*, etc. É assim possível obter várias categorias gramaticais: nomes, verbos, adjectivos, advérbios, conectores, etc. Este sistema permite, ainda, apresentar concordâncias, expressões ou fórmulas lexicais como é o caso de *definida como* ou *é constituída por*. Os seus correspondentes em língua inglesa, *lexical bundles* ou *multi-word sequences* (Biber *et al.*, 2004: s/ página), assumem formas como: *can be described as* ou *is composed of*.

Apesar do sucesso apriorístico desta prática que nos levou, aliás, a empreender este estudo, nem sempre os resultados apresentados pelo *Corpógrafo* se revelam úteis à pesquisa terminológica. Tomemos como exemplo o caso de [*definid.**].

Expressão de pesquisa: *definid.**

Corpus pesquisado: Riscos Naturais / Cheias (156746 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 151 distribuídas por 29 ficheiros

Frequência da totalidade das instâncias da concordância: 0.09 %

que para além de propostas válidas como:

10	-1	De acordo com a nossa definição , <u>uma bacia hidrográfica é um sistema fechado tridimensional (fala-se também de meio receptor) constituído por um determinado volume subterrâneo e uma superfície de intercepção , percorrida por uma rede hidrográfica concentrando os escoamentos superficiais .</u>	
----	----	--	---

propõe também propostas que não comportam informação terminológica útil, como:

23	-1	Em certas condições , podem já ser utilizados fundos comunitários para medidas preventivas em determinadas áreas e para certos tipos de riscos definidos .	
29	-1	Esta região identifica-se com a zona geográfica de características bem definidas , denominada de RIBATEJO .	

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo *Corpógrafo*.

Dado a finalidade deste estudo não ser a análise exaustiva e sistemática dos marcadores lexicais de definição e dos *padrões suporte*, não nos propomos analisar todos aqueles que se encontram listados quer no Capítulo 4, quer no Anexo I. O que propomos é observar a validade de alguns marcadores lexicais de definição e *padrões suporte* em contextos de ocorrência na tarefa de detecção e obtenção de informação terminologicamente interessante a partir do *Corpógrafo*. No seguimento do referido por Dubuc (1992:45-7, *cf.* ponto 3.3.1.), a análise contextual, é por nós entendida como todo o processo de investigação, recolha, selecção, análise e descrição de toda a informação terminológica contida em contextos de ocorrência, relativos a um conceito e respectiva(s) unidade(s) terminológica(s).

Posto isto, passamos em seguida à apresentação e análise mais detalhadas dos resultados obtidos. Para facilidade de apresentação dos resultados, separamos as quatro possibilidades de pesquisa e extracção de informação terminológica, bem como o ensaio dessa mesma informação.

6.1. A primeira possibilidade

A primeira fase (a obtenção de marcadores lexicais indiciadores de informação terminológica, com especial incidência nos marcadores de definição) realizou-se na sua totalidade no *Corpógrafo V2* (disponível até Setembro 2005). A aquisição destes marcadores (cujos quadros apresentamos no ANEXO I), processou-se de forma manual, dado ter sido a partir da opção *Contextos e Concordâncias* que o *Corpógrafo* nos possibilitou a obtenção dos marcadores enquanto ensaiávamos a viabilidade de alguns candidatos a termos na área de especialidade em estudo. Apresentamos e analisamos,

em seguida, dois exemplos extraídos do *Corpógrafo* V2 representativos do que acabámos de referir:

Pesquisa de contexto para o candidato a termo: **ponta de cheia**

Corpus pesquisado: Riscos Naturais / Cheias (158344 átomos)

Ocorrências encontradas: 17

Ocorrências por milhão no corpus (OCM): 107.360

Listando resultados...

#	concordância no contexto de uma frase	info
2	Estes , dispondo-se praticamente paralelos entre si , com bacias fortemente longitudinais e pequenas distâncias entre as suas embocaduras , <u>apresentam</u> , sobretudo nos sectores de jusante , uma forte inclinação que , aliada à sua fraca capacidade regularizadora , ocasiona escoamentos de valor elevado , ocorrendo a ponta de cheia poucas horas após a precipitação ter cessado .	
9	Síntese Conclusiva Embora tendo como função primordial a redução de volumes de escoamento e de caudais de ponta de cheia para jusante , as soluções de controlo na origem <u>constituem</u> elementos preciosos de requalificação e valorização do tecido urbano , possibilitando a inserção de mais valias paisagísticas , que <u>contribuem para</u> a interligação da Sociedade com o meio que a envolve .	
16	O elevado valor que as pontas de cheia atingem no troço português do Douro <u>resulta da</u> combinação de vários factores como a intensidade de precipitação e a forma da bacia hidrográfica , pouco alongada em relação à dos seus afluentes .	

Pesquisa de contexto para o candidato a termo: **volume of flow**

Corpus pesquisado: Natural Hazards / Floods (160383 átomos)

Ocorrências encontradas: 8

Ocorrências por milhão no corpus (OCM): 49.880

Listando resultados...

#	concordância no contexto de uma frase	info
1	MEASURABLE FEATURES OF FLOOD Floods <u>may be measured</u> as to height , area inundated , peak discharge , and volume of flow .	
2	The height of a flood <u>is of interest to</u> those planning to build structures along or across streams ; the area inundated <u>is of interest to</u> those planning to occupy in any manner the floodplains adjacent to a stream ; the peak discharge <u>is of interest to</u> those designing spillways , bridges , culverts , and flood channels ; and the volume of flow <u>is of interest to</u> those designing storage works for irrigation , water supply , and flood control .	
3	The volume of flow <u>as commonly used is measured</u> in acre-feet or in cubic feet per second-days .	

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo *Corpógrafo*.

Após a introdução da expressão de pesquisa *ponta de cheia e volume of flow*, detectámos diversos marcadores lexicais que, previsivelmente, poderiam evidenciar a presença de informação terminológica relevante. Foi, assim, possível extrair marcadores como *apresentar, ocasionar, ocorrer, constituir, atingir, possibilitar, contribuir, resultar e [to] be measured, [to] be of interest, [to] be obtained*.

Depois de recolhido um número considerável deste tipo de marcadores confirmámos a sua exequibilidade como elementos reveladores da presença de informação terminológica para além do termo. Julgamos importante referir que, o *Corpógrafo V2* permitia a pesquisa de definições apenas de forma manual, ao contrário do *Corpógrafo V3*, que oferece a possibilidade directa de pesquisar definições de forma automática.

Pesquisa sobre palavras: "consiste.*"

Corpus: Riscos Naturais / Cheias (145061 átomos)

Resultados: 9

Foram encontradas 9 concordâncias para a pesquisa: "consiste.*"

Frequência da totalidade das instâncias da concordância: 0.00 %

#	contexto anterior	instância	contexto posterior	info
1	e filosofia da abordagem tradicional (<u>designada</u> na terminologia francesa por tout à l'égout)	consiste	em recolher tão depressa quanto possível as águas pluviais, transportá-las por baixo do solo em	
5	<u>conceito de</u> bacia hidrográfica em termos de balanço <u>significa</u> fazer uma abordagem chamada sistémica: isto	consiste	em <u>considerá-la como</u> um sistema perfeitamente delimitado e em avaliar os volumes dos fluxos hídricos	
6	<u>intervenção na gestão das crises em tempo</u> quase real. O Sistema Mundial de Vigilância Meteorológica	consiste	na recolha e permuta mundial permanente da informação meteorológica e no processamento dessa informação em	

Pesquisa sobre palavras: "consist.*"

Corpus: Natural Hazards / Floods (145917 átomos)

Resultados: 36

Foram encontradas 36 concordâncias para a pesquisa: "consist.*"

Frequência da totalidade das instâncias da concordância: 0.02 %

#	contexto anterior	instância	contexto posterior	info
28	formations is disturbed by plowing or overgrazing. THE WORK OF STREAMS The work of streams	consists	of three closely related activities - erosion, transportation, and deposition. Stream erosion be progressive removal	
31	action to avoid or reduce their risk and prepare for effective response. Early warning systems	consist	of three elements (i) forecasting and prediction of impending events, (ii) processing and dissemination of	
34	mineral material from the floor sides of the channel, whether bedrock or regolith. Stream transportation	consists	of movement of the eroded particles dragged over the stream bed, suspended in the body	

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo *Corpógrafo*.

A expressão de pesquisa “*consist.**”[PT] (verbo consistir) e “*consist.**”[EN] (verbo [to] consist) produziu alguns resultados objectivamente mais interessantes ao nível terminológico, sobretudo na língua inglesa com um maior número de exemplos apresentados: apenas nove em português em oposição aos trinta e seis em inglês.

6.2. A segunda possibilidade

A segunda possibilidade surgiu como consequência da substituição da versão do *Corpógrafo*. Assim, nesta segunda fase, foi possível obter resultados de forma totalmente automática e visualizar os *padrões suporte* nos quais se baseou a busca automática em muito semelhantes aos marcadores lexicais de definição por nós obtidos manualmente, conforme descrito na primeira possibilidade. Para a extracção de informação terminológica, o *Corpógrafo V3* tem à nossa disposição a opção *Centro de Conhecimento*, seguida de outras que nos conduzem à pesquisa de definições. Em seguida, escolhemos como técnica de pesquisa a utilização de restrições lexicais. Como exemplo meramente ilustrativo do que acabámos de referir, atentemos ao caso dos termos *cheia* e *flood*:

Cheia

Pesquisa efectuada sobre o corpus Riscos Naturais / Cheias.
Definições propostas [56]:

Reguengo do Alviela, 1989 A **cheia** é um fenómeno natural extremo e temporário provocada por uma só chuvada torrencial ou por um longo período de chuvas copiosas, ou ainda por nevões ou pelo derretimento de calotes de gelo nalgumas partes do globo.

Padrão de suporte: (^|^o|^a|^um|^uma|o|a|um|uma)cheia é (uma|um|o|a)

Ficheiro de origem: C-01-PTPT.doc (SNBPC - Riscos Naturais: Cheias)

Introdução Designa-se por **cheia** a inundação temporária causada pela água de um rio, ribeiro, curso de água, oceano, lago ou outro corpo de água sobre terrenos adjacentes, podendo afectar o uso do solo e as restantes actividades humanas (D.R.C.G., 1969).

Padrão de suporte: (^|^entende-se|^designa-se|^denomina-se|entende-se|designa-se|

denomina-se)(por de como) cheia
Ficheiro de origem: C-11-PTPT.doc (Contribuição das Ciências Sociais para o Estudo dos Impactes das Cheias)

As cheias têm causado periódicas destruições das áreas mais baixas das antigas e modernas cidades .
Padrão de suporte: (^ os ^as os as)cheias (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuam funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
Ficheiro de origem: C-09-PTPT.doc (Riscos associados com cheias fluviais.)

flood

Pesquisa efectuada sobre o corpus Natural Hazards / Floods.

Definições propostas [229]:

Flood frequencies , such as the " 100-year flood , " are determined by plotting a graph of the size of all known floods for an area and determining how often floods of a particular size occur .
Padrão de suporte: (^ ^called ^named called named)flood
Ficheiro de origem: F-06-ENUSA.doc (Basics of flooding and flood protection.)

Very large floods happen very seldom The size , or magnitude , of a flood is described by a term called <u>recurrence interval</u> .
Padrão de suporte: (^ ^the ^a ^an the a an)flood is (initiated protected secreted calculated considered spread defined released produced described regarded viewed)
Ficheiro de origem: F-08-ENUSA.doc (Floods and flood plains)

At this time the flood takes on the characteristics of a flash flood , with the added danger of ice flows that , when driven by the energy of the floodwave , can inflict serious damage on <u>structures</u> .
Padrão de suporte: (^ ^the ^a ^an ^each the a an each)flood (results releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings consists corresponds leads depends takes receives derives plays)
Ficheiro de origem: F-10-ENUSA.doc (Significant Floods in the United States During the 20th Century - USGS Measures a Century of Floods.)

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo *Corpógrafo*.

Conforme podemos verificar através dos resultados obtidos, os *padrões suporte*, nos quais se baseia a detecção de definições, são apresentados individualmente em cada uma das propostas a definições para um determinado termo. É, assim, possível aferir em que *padrão suporte* o *Corpógrafo V3* se baseou para apresentar aquela definição.

Comparativamente à primeira possibilidade, trata-se de uma forma mais directa e célere de obter informação terminológica. Tal é possível pois, em vez de seleccionarmos como expressão de pesquisa indiciadora de informação terminológica um verbo ou grupo verbal, seleccionamos termos de maior significado da área de conhecimento e obtemos propostas interessantes como as que acabámos de apresentar.

6.3. A terceira possibilidade

A terceira possibilidade reflecte, por assim dizer, o resultado da complementaridade dos elementos obtidos nas duas primeiras fases: os marcadores lexicais indiciadores de informação terminológica e os *padrões suporte*, os quais nos propusemos testar numa área distinta do *Corpógrafo V3*: a da *Pesquisa – Concordância Frase*. Esta área permite visualizar, sem restrição no número de palavras (apenas limitado a uma frase), todo o contexto no qual se insere a expressão de pesquisa, seja ela constituída por uma ou mais palavras (por exemplo, *cheias*; *velocidade de subida do nível da água*; *entende-se*; *chamado*; *etc.*). Uma outra vantagem da *Concordância Frase* é o facto de podermos inserir na *expressão de pesquisa* o elemento linguístico que quisermos, enquanto em *Pesquisar Definição*, a única escolha possível é o termo que queremos ver definido.

Embora, por vezes, a sua visualização não seja imediata, constatámos através de uma análise mais atenta que é possível localizar elementos terminologicamente relevantes para a pesquisa que propomos na área dos *Riscos Naturais*: *Cheias*.

O primeiro exemplo que propusemos testar não consta da lista dos *padrões suporte*, embora constitua um exemplo típico de marcador lexical de informação terminológica para obtenção de termos que integram a hierarquia da área conceptual ou

que fazem parte da rede semântica do termo. Tomemos, desta forma, como exemplo os casos das combinatórias lexicais de *tais como* e *ou seja*:

tais como

- Teste no *Corpógrafo V2*:

Pesquisa sobre palavras: "tais" "como"

Corpus: Riscos Naturais / Cheias (145061 átomos)

Resultados: 3

Foram encontradas 15 concordâncias para a pesquisa: "tais" "como"

Frequência da totalidade das instâncias da concordância: 0.01 %

#	contexto anterior	Instância	contexto posterior	info
7	acções nas zonas inundáveis A tradicional estratégia de modificar as cheias baseia-se nas medidas estruturais	tais como	a construção de barragens e a criação de albufeiras, construção de diques, muros de defesa, alterações	
11	difícil manutenção. Também as frequentes estruturas implantadas nos canais activos e nas planícies de inundaçã	tais como	pilares de pontes ou viadutos, muros e edifícios podem restringir fortemente o fluxo de cheia,	
15	zonas inundáveis devem ser repartidas por diferentes instituições com ligações a várias actividades técnicas ou sectoriais,	tais como	a hidrologia, a hidráulica, a engenharia civil, o planeamento, a geologia, a pedologia, o ordenamento do solo,	

- Teste no *Corpógrafo V3*:

Expressão de pesquisa: tais como

Corpus pesquisado: Riscos Naturais / Cheias (156746 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 18 distribuídas por 10 ficheiros

Frequência da totalidade das instâncias da concordância: 0.01 %

#	f	Frase onde ocorre a concordância	Info
4	-1	Os ventos fortes , as marés de tempestade e as cheias são fenómenos associados aos sistemas de tempestades , tais como os ciclones tropicais , depressões e sistemas frontais , que se estendem por centenas de quilómetros e cuja duração varia entre algumas horas e vários dias .	
6	-1	Os processos naturais e os processos geológicos em particular , tais como aqueles que produzem sismos , erupções vulcânicas , deslizamentos de terras e avalanches de lama , quedas de blocos , assoreamentos , inundações , erosão costeira , .	
12	-1	A caracterização das cheias nas bacias hidrográficas permite definir o hidrograma de cheia e as propriedades do escoamento , tais como as profundidades , as velocidades , as dimensões do material de fundo aluvionar , o caudal sólido , etc. Esta caracterização é essencial para a definição do controlo das cheias , de defesa contra inundações e de regularização fluvial dos leitos aluvionares e ainda para o conhecimento do habitat nas margens e no leito dos rios .	

ou seja**- Teste no Corpógrafo V2:**

Pesquisa sobre palavras: "ou" "seja"

Corpus: Riscos Naturais / Cheias (145061 átomos)

Resultados: 3

Foram encontradas 3 concordâncias para a pesquisa: "ou" "seja"

Frequência da totalidade das instâncias da concordância: 0.00 %

#	contexto anterior	Instância	contexto posterior	info
1	novas criadas, ao nível do Biota a adaptação tende a ocorrer em termos de evolução	ou seja	numa escala temporal dilatada, enquanto que o mais frequente para o curto espaço de tempo,	
2	com vista a se equacionar qual poderá ser o efeito da destruição do respectivo habitat,	ou seja	com vista a se saber se será previsível o abandono da zona pelas espécies aí	
3	encostas da serra, verificou-se que as mesmas não sofreram grandes danos devido à erosão hídrica,	ou seja	não se encontraram vestígios de desmoronamentos ou deslizamentos, pelo que as mesmas mantêm a sua	

- Teste no Corpógrafo V3:

Expressão de pesquisa: ou seja

Corpus pesquisado: Riscos Naturais / Cheias (156746 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 22 distribuídas por 14 ficheiros

Frequência da totalidade das instâncias da concordância: 0.01 %

#	f	Frase onde ocorre a concordância	Info
18	-1	Assim , a avaliação das situações associadas às cheias centram-se naquilo que Lazarus e Folkman (1984) <u>designam por " primary appraisal' ou seja , na resposta à questão seguinte : " Até que ponto é negativo aquilo que me está a acontecer ?</u>	
21	-1	Ou seja , a apresentação dos resultados dos estudos de ondas de inundação fica ao critério do consultor que os realiza , envolvendo a decisão sobre diversos aspectos, nomeadamente , a escala das cartas a utilizar , os valores característicos a incluir .	
22	-1	Se alguma estatística descritiva poderia ser feita (e tomando em conta também as cheias de Lisboa de 1967 e 1983) , é a de que as cheias extremas ocorrem com maior probabilidade ao fim-de-semana , e de noite 2 - AS CHEIAS DE OUTUBRO DE 1997 No presente capítulo só serão referidas <u>aquelas cheias mais extraordinárias , ou seja , apesar da definição crua de cheia fazer corresponder esse fenómeno a toda e qualquer subida dos níveis no leito de um rio , não será assim tratada neste artigo , como é evidente .</u>	

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo Corpógrafo.

Estas amostras permitem-nos concluir que existem marcadores lexicais, detectados a partir dos termos (em contexto) propostos como candidatos pelo

Corpógrafo V2 que apresentam resultados positivos na nova versão do *Corpógrafo*, pelo que a actualidade dos marcadores é uma evidência.

Por outro lado, existem marcadores lexicais detectados no *Corpógrafo V2* que são igualmente apresentados pelo *Corpógrafo V3* como *padrões suporte*, de que são exemplo, *denominad(o/os/a/as)*, *denominad(o/os/a/as)*, *corresponde(m)* e *resulta(m)*.

6.4. A quarta possibilidade

No decorrer do nosso estudo empírico (utilizando *Corpógrafo V3*), foi-nos proporcionada uma quarta possibilidade: a utilização da *Concordância Frase* como auxílio à pesquisa terminológica no seguimento da declarada ausência de resultados em *Pesquisar Definições – Usando restrições lexicais*.

A constatação que retirámos do nosso estudo é que, apesar de o *Corpógrafo* listar determinados termos para que possam ser seleccionados como tal, certo é que nem sempre os textos onde esses termos são detectados comportam as respectivas definições. Esse resultado negativo é expresso pela *Pesquisar Definições – Usando restrições lexicais*, de duas formas:

- ou refere que não existem definições propostas, como é o caso de:

regime de caudais

Pesquisa efectuada sobre o corpus Riscos Naturais / Cheias.

Definições propostas [0]:

- ou apresenta uma proposta de definição, que na realidade não o é, como neste caso concreto:

Pesquisar Definições – Usando restrições lexicais:

caudal de cheia

Pesquisa efectuada sobre o corpus Riscos Naturais / Cheias.

Definições propostas [1]:

Os **caudais de cheia** em Coimbra eram da ordem dos 2500 m³/s , sendo amortecidos para 1200 m³/s através dos dois aproveitamentos referidos anteriormente .

Padrão de suporte: (^|desencadeia|^causa|^provoca|desencadeia|causa|provoca)(os|as) caudais de cheia

Ficheiro de origem: C-10-PTPT.doc (O Rio Mondego. O Ambiente Fluvial e a sua Ecologia.)

Contudo, se inserirmos o termo na *Concordância Frase*, o *Corpógrafo* apresenta diversos dados terminológicos associados a este termo:

Concordância Frase:

Expressão de pesquisa: caudal de cheia

Corpus pesquisado: Riscos Naturais / Cheias (156746 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 9 distribuídas por 6 ficheiros

Frequência da totalidade das instâncias da concordância: 0.00 %

#	f	Frase onde ocorre a concordância	Info
1	-1	Com efeito , o Rio Douro em Portugal tem <u>margens muito íngremes e uma bacia pouco alongada em relação à dos seus afluentes e com um escasso recobrimento vegetal</u> , sendo estas as principais causas da intensidade <u>das cheias</u> deste rio , assim como do elevado quociente entre o seu caudal de cheia e o seu <u>caudal médio anual</u> ; verifica-se que o quociente desses dois valores <u>atingem na bacia</u> do Douro números que não são frequentes nos rios europeus .	
2	-1	De acordo com este autor , as <u>grandezas hidrográficas (comprimento do rio , área da bacia hidrográfica)</u> , hidrológicas (<u>caudal médio anual</u> , caudal de cheia média anual , <u>relação entre estes dois caudais</u> , duração das cheias) e morfológicas (<u>profundidade e largura do leito principal</u> , <u>inclinação média do perfil longitudinal</u>) são relacionadas com o número de ordem do rio .	
4	-1	Pelo contrário , <u>um rio de ordem XV</u> poderá apresentar um comprimento da ordem dos 3900 km , um <u>caudal médio anual</u> de 8500 m3/s , um caudal de cheia média anual de 40000m3/s , um <u>leito principal</u> com uma largura de 1000 m , e uma <u>inclinação longitudinal</u> de 0,002% .	
6	-1	Antes da <u>regularização fluvial</u> o caudal de cheia centenária em Coimbra foi calculado em 3663 m3/s (D.G.R.A.H. , 1987) .	
7	-1	1) , já em contexto <u>de enfraquecimento do caudal de cheia a velocidade máxima (em vazante) à superfície</u> foi de 2,0 m/s e <u>no fundo</u> cerca de 1,5 m/s (<u>o que permite arrastar grãos de dimensão inferior a 3 mm</u>) ; a velocidade incrementava ligeiramente <u>para montante</u> .	
8	-1	De acordo com RZHANITSYN (1960) a <u>hidrografia (comprimento do rio desde a nascente e área da bacia)</u> , a hidrologia (<u>caudal médio</u> , caudal de cheia médio , <u>a relação entre ambos e a duração das cheias</u>) e a morfologia (<u>profundidade na estiagem</u> , <u>largura média do canal e inclinação longitudinal</u>) podem ser relacionadas com o número de ordem do rio .	
9	-1	<u>Bacias de Retenção/Infiltração</u> estas estruturas podem apresentar duas funções : <u>retenção sem infiltração</u> , onde apenas é amortecido o caudal pluvial ; <u>retenção e infiltração</u> , sendo retido o caudal de cheia , que é posteriormente infiltrado no solo .	

Nesta situação compensou recorrer à *Concordância Frase*, que sempre apresenta o contexto de ocorrência do termo e, não raras as vezes, apresenta igualmente outro tipo de informação terminológica, como indícios ou orientações que nos permitem a pesquisa desse termo noutras direcções.

O que acabámos de referir espelha-se no exemplo do termo *cheia(s)*, que em *Pesquisar Definições – Usando restrições lexicais* (pesquisa efectuada sobre o corpus

Riscos Naturais: Cheias) propõe 56 definições. O mesmo termo inserido na *Concordância Frase* rende 423 concordâncias, as quais incluem muita informação terminologicamente relevante para a caracterização do termo e posterior execução da respectiva Ficha Terminológica. A título meramente demonstrativo recolhemos três amostras de informação que não são facultadas pela opção *Pesquisar Definições*:

Cheia

2	-1	<u>Tudo isso estará na origem das grandes quantidades de materiais rochosos e de detritos vegetais que chegam ao rio principal , que os transportará e os depositará em tempo de cheia .</u>	
45	-1	<u>Pelo contrário , em bacias de grandes dimensões , a ponta de cheia , e as inerentes inundações , demoram mais tempo a instalar-se , permitindo um aviso mais atempado às populações .</u>	
48	-1	<u>O laminar de caudais nas barragens hidroeléctricas existentes permite atenuar os riscos de inundação nas duas primeiras localidades , mas o efeito da orografia do terreno (serra da Peneda) suscita um acréscimo de precipitação que se traduz na formação de elevados caudais de cheia , nem sempre passíveis de serem armazenados nas albufeiras .</u>	

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo *Corpógrafo*.

Demonstrámos, assim, que é possível encontrar na *Concordância Frase* informação terminológica relevante para uma pesquisa que não consta do conjunto de propostas fornecidas por *Pesquisar Definição*.

No próximo quadro encontram-se representados alguns exemplos de marcadores lexicais indiciadores de informação terminológica utilizados na detecção de informação terminológica em contexto de ocorrência na área *Pesquisa – Concordância Frase* do *Corpógrafo V3*.

Língua portuguesa			Língua inglesa		
Expressões de Pesquisa no <i>Corpógrafo V3</i>	Variações das Expressões de Pesquisa	Nº de Concordâncias	Expressões de Pesquisa no <i>Corpógrafo V3</i>	Variações das Expressões de Pesquisa	Nº de Concordâncias
definid.*	definida(s)/o(s) definir(em)(-se) definição	151	defin.*	definition(s) define(s) defined	99
caracterizad.*	caracterizada(s)/o(s) caracterizar caracterização	79	characteriz.*	charcterized	5
	caracterizam(-se)		characteris.*	charcterised	9
chamad.*	chamada(s)/o(s) chamamos chamar(-se) chamava(m) chama(-se)	36	called	called	43
designad.*	designada(s)/o(s) designa designadamente	24	name.*	name namely	11
designad.* por	designada(s)/o(s) por	6	designat.*	designate(s) designated designating designation	12
significa	significa ⁽¹⁾	6	means	means	53
significam	significam ⁽¹⁾	2			
constitu.*	constitui constituído(s)/a(s) constituem constituiria constituir(-se) constituintes	120	constitut.*	constitute(s)	23
conceito.*	conceito(s)	59	concept.*	concept(s) conceptualization conceptualizes conceptualizing conceptual	25
termo	termo	89	term.*	term(s) termed	120
o termo	o termo	9	the term	the term	16

Figura 2 – Resultados de pesquisa de alguns marcadores lexicais indiciadores de informação terminológica.

- (1) Apesar da expressão *significa.** ter rendido 120 ocorrências, optámos por ensaiar o singular do plural (*significa* e *significam*) em separado, dado a taxa de ruído ter atingido os 95%.

Não queremos deixar de chamar a atenção para os resultados obtidos com os marcadores lexicais de definição *o termo* e *the term*. Nesta situação pusemos em prática, adoptando à realidade do *Corpógrafo V3*, um dos padrões sintácticos de restrição enunciados por Pearson (1998: 128-134), denominado “*flagged term*”, como critério de refinamento para detecção e extracção de informação terminológica: a utilização do artigo definido. Observemos, pois, os resultados obtidos nas duas línguas em estudo:

Expressão de pesquisa: **o termo**

Corpus pesquisado: Riscos Naturais / Cheias (156746 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 9 distribuídas por 6 ficheiros

Frequência da totalidade das instâncias da concordância: 0.00 %

#	f	Frase onde ocorre a concordância	Info
1	-1	SÍNTESE E CONCLUSÕES FINAIS Falámos de risco no sentido pré-científico , Desde que surge a ideia de criar uma ciência do risco ou do perigo (cindinicologia talvez fosse o termo indicado) , analisam-se os muitos casos concretos de acontecimentos danosos para o homem e seus bens e verifica-se que uma coisa é o risco em sentido restrito , distante , outra é o perigo , próximo ; outra , ainda , é a manifestação do risco , ultrapassando o controle humano , ou seja , a crise .	
2	-1	REBELO (1999) preferem reservar o termo perigo para ser utilizado no contexto da sequência Risco Perigo Crise , estabelecida de acordo com trabalhos patrocinados pela UNESCO (1987 e 1989) e , particularmente , do trabalho teórico desenvolvido por L .	
3	-1	Neste caso o termo Hazard teria um significado muito próximo do de risco , devendo ser este o termo apropriado em termos de tradução .	
4	-1	A mais recente definição francesa especializada era a de Roche (1986) : o termo bacia hidrográfica comporta simultaneamente a noção topográfica de zona limitada por uma linha de partilha de águas e a de superfície de intercepção da precipitação .	
5	-1	Quem desenvolve modelos usando intervalos de tempo discretos , geralmente regulares , tem o hábito de procurar simplificar o funcionamento de uma bacia hidrográfica , por um lado , redistribuindo , em cada intervalo de tempo , o sinal de entrada nas suas diferentes formas de saída e , por outro , propagando os diferentes sinais de saída de acordo com as funções do tempo no interior do sistema . É frequente encontrarem-se , para estas duas operações , as expressões de funções de produção e de transferência . O termo redistribuição é mais exacto que o termo produção e é utilizado nesta ficha .	
6	-1	A parcela A parcela é o local privilegiado para a observação experimental dos mecanismos locais do ciclo hidrológico . Utiliza-se frequentemente o termo parcela para representar o espaço delimitado , de área raramente superior a algumas dezenas de m2 , no qual se efectuam as experimentações . É nesta escala que são avaliados os parâmetros das principais leis físicas da hidrologia relativas à precipitação , à infiltração e à evapotranspiração . A este nível de descrição não existe uma verdadeira transferência , neste nível só se estuda , em geral , a redistribuição da precipitação .	
7	-1	Riscos tecnológicos O termo riscos tecnológicos abrange uma vasta gama de riscos potenciais .	
8	-1	O termo " consequências das cheias " refere-se já aos impactes desse desastre no bem-estar de um determinado grupo social : diferentes grupos humanos sofrerão diferentes consequências das cheias , mesmo que se esteja a referir a um mesmo desastre .	
9	-1	Utilizamos o termo " danos " para referir essas perdas potenciais que , se forem evitadas , irão ser tratadas como benefícios na análise de custos e benefícios .	

Expressão de pesquisa: the term

Corpus pesquisado: Natural Hazards / Floods (157536 átomos : 46 ficheiros)

Nº de concordâncias obtidas: 16 distribuídas por 10 ficheiros

Frequência da totalidade das instâncias da concordância: 0.01 %

#	f	Frase onde ocorre a concordância	Info
3	-1	The term " risk management " describes a decision-making process that involves defining need , recognizing the options that are available and acceptable , and choosing an appropriate alleviation strategy .	10
4	-1	We can define the term flood as the condition that exists when the discharge of a river cannot be accommodated within its normal channel .	10
5	-1	The term lake includes a wide range of water bodies .	10
6	-1	As we saw in Chapter 16 , the term alluvium is used to describe any stream-laid sediment deposit .	10
7	-1	The term na-tech (for natural-technological) is increasingly being used to describe these complex disasters for which there is a recognisable and significant human contribution .	10
8	-1	Thus , despite the convenience of the term , truly 'natural' hazards do not exist .	10
9	-1	The term environmental hazard has the advantage of including a wide variety of hazard types ranging from 'natural' (geophysical) events , through 'technological' (man-made) events to 'social' (human behaviour) events (Figure 1.7) .	10
13	-1	In his examination of the ecology of floods , Hillman highlights that the term 'flood' usually has a negative , hazardous or disastrous connotation .	10
14	-1	'Flood disaster' derives most of its meaning from 'disaster' , which is the term often given to the harmful event that occasionally arises as a consequence of hazard .	10

Nota: As formatações a negrito e sublinhado são da nossa responsabilidade, não correspondendo, portanto, à formatação original apresentada pelo *Corpógrafo*.

Da breve análise sobre a qualidade das ocorrências que acabámos de apresentar podemos concluir que é possível trabalhar e evidenciar no *Corpógrafo V3* algumas teorias apresentadas no sub-capítulo 3.3. (*Extracção de informação terminológica relevante: termos, definições, contextos de uso, sistemas conceptuais e redes semânticas*), daí resultando evidentes benefícios para a pesquisa terminológica.

A exploração desta quarta e última possibilidade permite-nos verificar que a associação das possibilidades de pesquisa oferecidas pelos marcadores lexicais indiciadores de informação terminológica às possibilidades trazidas pelos *padrões suporte* é vantajosa, representando um contributo indispensável numa pesquisa terminológica como a proposta pelo Departamento de Geografia para os *Riscos Naturais: Cheias*. Por um lado, os *padrões suporte* significam um avanço importante na

pesquisa terminológica automática ou semi-automática, ao oferecerem, com rapidez e segurança, grande parte da informação necessária ao preenchimento da Ficha Terminológica. No entanto, conforme evidenciado pelos resultados do nosso estudo, os *padrões suporte* não substituem por inteiro as possibilidades de pesquisa que os marcadores lexicais indiciadores de informação terminológica oferecem. Neste capítulo demos exemplos reveladores dessa complementaridade no trabalho de pesquisa terminológica obtida com os *padrões suporte*, com informação distinta, obtida através dos marcadores utilizados como *expressão de pesquisa* na *Concordância Frase*.

Ao longo do Capítulo 6 apresentámos um conjunto de possibilidades de detecção e extracção de informação terminológica a partir do *Corpógrafo*, com especial destaque para a Versão 3 (*Corpógrafo V3*), por via das *Pesquisas de Termos e de Definições*, bem como das *Concordâncias*. Tendo analisado os dados extraídos passamos, em seguida, a uma síntese de conclusões e futuras propostas de trabalho alcançadas ao longo do presente estudo.

Conclusões

Conforme afirmámos no início do trabalho, procurámos com este estudo chamar a atenção para uma forma distinta de pesquisa e extracção de informação terminológica. Para tal testámos e avaliámos o comportamento dos marcadores lexicais indiciadores de informação terminológica e dos *padrões suporte* no *Corpógrafo*.

Antes de termos dado início ao estudo empírico foi necessário determinar as bases teóricas nas quais se fundamentaria o nosso estudo. Nesse sentido, podemos afirmar que este trabalho envolveu conhecimentos de três grandes áreas de estudo: Terminologia, Linguística de *Corpus* e Geografia (*Riscos Naturais: Cheias*), o que evidencia que a inter e multidisciplinaridade é cada vez mais necessária e determinante, sobretudo no estudo de uma língua de especialidade. O aproveitamento das sinergias de todas as áreas envolvidas fez com que os intervenientes neste trabalho partilhassem de uma visão integrada com base num delineamento conceptual determinante para a dinâmica do conhecimento e funcionamento da língua de especialidade à qual se refere este projecto.

A aquisição de terminologia baseada nos *corpora*, construídos para o efeito, aumentou a qualidade, não apenas da pesquisa terminológica, mas também do seu produto ao abrir um novo caminho à investigação empírica através da pesquisa exaustiva de mais informação terminológica, pela diminuição de erros e pela possibilidade de fornecer mais informação de contexto. A preferência dada a material linguístico autêntico no domínio dos *Riscos Naturais* e sub-domínio das *Cheias* contribuiu de forma determinante para o sucesso da extracção terminológica (automática e semi-automática) através da nova ferramenta informática - *Corpógrafo* - ao serviço da Linguística e de qualquer área de conhecimento de especialidade com representação linguística.

Da mesma forma, o destaque que atribuímos ao contexto não se justifica apenas pelo facto de ter uma função de atestação do uso de um termo num determinado domínio conceptual. O contexto tem como função primeira contribuir para a compreensão do termo e consequentemente do conceito que ele linguisticamente representa. Os elementos de um contexto mais elaborado contribuem, inclusivamente, para uma maior clareza na representação do sistema conceptual e da rede semântica do domínio, ao fornecerem importantes pistas para a percepção das relações com outros

termos presentes no texto. No nosso estudo, os contextos representaram uma inestimável fonte de informação terminológica.

Foi a partir dos contextos apresentados pelo *Corpógrafo V2* que extraímos e testámos os marcadores lexicais indiciadores de informação terminológica (através da opção *Concordância e Colocações*). Os referidos contextos foram mais tarde utilizados, juntamente com os *padrões suporte* apresentados pelo *Corpógrafo V3*, nas nossas pesquisas. Tanto os marcadores lexicais, como os *padrões suporte* foram identificados e analisados antes de serem submetidos ao *Corpógrafo* de forma a testar o seu contributo na pesquisa e extracção de informação terminológica.

Para alcançarmos os objectivos propostos para o nosso estudo, estabelecemos uma metodologia composta por seis etapas, que nos conduziram ao estudo empírico dos elementos até então delineados apenas teoricamente.

Devido à substituição do *Corpógrafo V2*, pelo *Corpógrafo V3*, vimo-nos forçados, a dada altura do nosso estudo, a proceder a uma reformulação da nossa proposta, já que no início esta se baseava exclusivamente nas opções disponíveis no *Corpógrafo V2*. Assim, aproveitado esta nova oportunidade oferecida pela sua versão mais recente, quisemos testar nos nossos *corpora* de especialidade apenas uma parte do processo de uma pesquisa automática e semi-automática. Nesta perspectiva, podemos concluir que, tanto os marcadores lexicais indiciadores de informação terminológica, como os *padrões suporte* constituem uma forma eficaz para detectar e extrair informação terminológica relevante, tendo apresentado, a esse nível, resultados bastante proveitosos.

Na descrição do processo de extracção e na apresentação dos resultados, ao estabelecermos um paralelo entre a extracção manual e a extracção automática e semi-automática da informação terminológica, não menosprezámos, de forma alguma, todo o caminho percorrido nesta área; muito pelo contrário, foi a partir dos estudos e das propostas dos cinco autores por nós escolhidos, que baseamos o nosso estudo no *Corpógrafo V2*. Com o surgimento da nova versão, julgámos interessante aproveitar o trabalho já desenvolvido para estabelecer um paralelo com a nova forma de extracção de informação proposta pelo *Corpógrafo V3*.

A referida substituição das versões do *Corpógrafo* determinou uma alteração na proposta inicial do nosso estudo. Assim, se no *Corpógrafo V2*, a intenção era testar os marcadores lexicais indiciadores de informação terminológica directamente nos *corpora* (dado que nesta versão não era necessária a constituição de Bases de Dados), de forma a

obter outro tipo de informação para além de unidades terminológicas; o mesmo não sucedeu com os testes efectuados no *Corpógrafo V3*. Esta nova versão, mais sofisticada, permite progredir com maior celeridade no trabalho de pesquisa, desde que o primeiro se constitua uma Base de Dados na qual são guardadas as unidades terminológicas entretanto seleccionadas.

Os principais objectivos deste estudo foram, assim reformulados tendo em consideração a nova proposta do nosso estudo e podem ser sintetizados em dois pontos:

- observar o comportamento e avaliar o grau de eficácia dos marcadores lexicais de definição e *padrões suporte* no *Corpógrafo* na detecção e extracção de informação terminológica útil;
- efectuar uma análise descritiva e comparativa entre os dados obtidos através dos marcadores lexicais indiciadores de informação terminológica e os *padrões suporte*.

Tendo atingido os objectivos básicos delineados para este estudo, é-nos possível retirar algumas conclusões, assim como apontar algumas ideias para trabalhos futuros ligados a esta temática.

Nesta perspectiva, de acordo com as nossas percepções, concluímos que:

1. A Terminologia desempenha um papel cada vez mais importante na transmissão de conhecimento no complexo enredo da *Era da Informação*, que é simultaneamente de globalização e especialização.
2. A pesquisa terminológica tem assinalado uma importante evolução nestes últimos anos com o progresso nos estudos da Linguística de *Corpus*, dos *corpora* textuais e das línguas de especialidade, em simultâneo com o surgimento de novas ferramentas de trabalho (nomeadamente ferramentas informáticas), que permitem aprofundamento dos estudos, já que alguns programas realçam, a partir dos *corpora*, aspectos das línguas de especialidade dificilmente detectáveis de forma manual.
3. O *Corpógrafo* provou ser uma ferramenta útil e eficiente quanto ao seu objectivo de detectar e extrair informação terminológica a partir de *corpora*, funcionando bem nos dois sistemas linguísticos testados no nosso estudo (português e inglês).

4. A pesquisa terminológica com o intuito de extracção de informação pode ser efectuada de diversas formas. A nossa opção recaiu na forma indirecta, conduzindo-nos ao estudo de determinadas fórmulas linguísticas: designadas genericamente de descritores e/ou marcadores pelos cinco autores escolhidos como referência para este estudo e por nós referidos como marcadores lexicais indiciadores de informação terminológica e, pelo *Corpógrafo V3*, como *padrões suporte*. Apesar das *nuances* que as distinguem, o seu objectivo é comum: a detecção e extracção de informação terminológica.
5. Pelos resultados apresentados através do *Corpógrafo* (nas versões V2 e V3), tanto os marcadores lexicais indiciadores de informação terminológica, como os *padrões suporte* constituem uma base válida para a detecção e a extracção de informação terminológica.
6. A produtividade dos marcadores lexicais e dos *padrões suporte* é variável, uma vez que nem todos apresentam os mesmos resultados. O domínio dos *Riscos Naturais: Cheias* não se apresenta na sua essência rigidamente estruturado e hierarquizado. Por este motivo, os marcadores lexicais e *padrões suporte* de estruturação conceptual nem sempre produziram os resultados inicialmente aguardados.
7. Os marcadores mais produtivos são os de origem verbal (verbos e grupos verbais), comparativamente a nomes, adjectivos e advérbios. Verificámos, igualmente, bons resultados aquando da utilização de alguns marcadores lexicais, nomeadamente de reformulação discursiva.
8. Através do ensaio de exemplos de marcadores lexicais indiciadores de informação terminológica demonstrámos a possibilidade de aplicar no *Corpógrafo V3* algumas teorias apresentadas relativamente a processos de detecção e extracção de informação terminológica relevante.
9. Desta forma, através dos resultados alcançados nos testes no *Corpógrafo*, a extracção terminológica ganha algum grau de profundidade quando associamos as possibilidades de pesquisa oferecidas pelos marcadores lexicais indiciadores de informação terminológica às possibilidades que os *padrões suporte*. Consideramos, assim vantajosa a utilização destes recursos lexicais, representando mesmo um contributo indispensável numa pesquisa terminológica.

Tendo em consideração os conceitos de *termo* e *definição terminológica*, propusemos, neste trabalho uma nova abordagem metodológica analítico-descritiva utilizando, para esse efeito, o *Corpógrafo*. Podemos asseverar que, neste estudo, o trabalho terminológico foi assumidamente descritivo, lançando as bases para um futuro trabalho mais amplo e igualmente descritivo, embora com algumas características normativas, necessariamente presentes na construção de um glossário de *Riscos Naturais*. Quando referimos normativo, fazemo-lo, não no sentido de impor a nossa visão sobre a terminologia desta área de especialidade (dado não representarmos o IPQ – Instituto Português da Qualidade), mas no sentido de elaborar um glossário assumindo os requisitos de um trabalho terminológico e terminográfico de forma a sugerir o uso mais adequado da terminologia que o irá compor.

Nos tempos que correm é necessário vencer algumas ideias pré-concebidas (felizmente cada vez em menor número) de que os programas informáticos são incapazes de efectuar escolhas acertadas e apresentar resultados credíveis para uma pesquisa terminológica.

A utilização do *Corpógrafo* (sobretudo na sua mais recente versão, a V3) é exemplo de uma nova perspetivação da pesquisa terminológica, não querendo isto dizer que é melhor ou pior que qualquer outra abordagem; simplesmente concluímos que no nosso estudo provou ser uma ferramenta de sucesso irrefutável. Neste sentido, apenas nos resta referir que o estudo procurou demonstrar, numa perspectiva linguística e comunicativa, a viabilidade e produtividade de reconhecimento na detecção e extracção de informação terminológica. A detecção e extracção automáticas ou semi-automáticas de termos e de informação contextual, incluindo definições, contextos de uso, sistemas conceptuais e redes semânticas, através de programas informáticos é potenciada através de recursos linguísticos, como os marcadores lexicais indiciadores de informação terminológica e os *padrões suporte*, utilizados no discurso de especialidade.

Como trabalho futuro, pretendemos continuar a utilizar o *Corpógrafo* como forma de detecção e extracção de informação terminológica para a construção de um Glossário de *Riscos Naturais* em parceria com o Departamento de Geografia da Faculdade de Letras da Universidade do Porto. Nesse sentido, trabalharemos com o intuito de explorar e aprofundar ao máximo todas as capacidades do *Corpógrafo*, para que, definitivamente, passe a constituir uma das mais vantajosas ferramentas de pesquisa terminológica que os investigadores têm ao seu dispor.

Bibliografia e fontes

- AHMAD, K., ROGERS, M. (1997). "Corpus Linguistics and Terminology Extraction." In *Handbook of Terminology Management. Basic Aspects of Terminology Management*. Compilado por Sue Ellen Wright e Gerhard Budin. Philadelphia/Amsterdam: John Benjamins Publishing Company. Volume 2, pp.725-760.
- ALSINA, V., ESTOPÀ, R. (1996). "Las profesiones y los usuarios de la terminología". In *Terminómetro. La Terminología en España*. União Latina. Número Especial: pp.85-87.
- AUSTERMÜHL, F. (2001). *Electronic Tools for Translation – Translation Practices Explained*. Manchester: St. Jerome Publishing. Capítulo 8, pp.124-133
- BERDET, E. F. (1999). "Mujer y diccionario: lo femenino en los ejemplos lexicográficos". Artigo apresentado no XX Congresso de AESLA (Asociación Española de Lingüística Aplicada), realizado em Logroño (Espanha), em Março de 1999. Disponível em <http://www.hispanista.com.br/revista/artigo48esp.htm>, [acedido em 2005.01.13].
- BERLAND, S., GRABAR, N. (2002). *Assistance automatique pour l'homogénéisation d'un corpus Web de spécialité*. In JADT 2002 : 6^{èmes} Journées Internationales d'Analyse Statistique des Données Textuelles. Disponível em http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/PDF-2002/berland_grabar.pdf, [acedido em 2005.03.31].
- BERLAND, S. (2000). *Constitution de corpus à partir du Web pour l'acquisition terminologique : une expérience*. Chapitre 1 : La terminologie textuelle. Disponível em <http://www-poleia.lip6.fr/~slodzian/sberland/Chapitre1.html>, [acedido em 2005.03.31].
- BIBER, D., CONRAD, S., CORTES, V. (2004). *If you look at ...: Lexical Bundles in University Teaching and Textbooks*. Disponível em <http://applied.oxfordjournals.org/cgi/content/abstract/25/3/371>, [acedido em 2005.01.31].
- BIBER, D., CONRAD, S., REPPEN, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge. Cambridge University Press.

- BLANCHON, E. (1997). "Point de vue sur la définition". In META, Vol. XLII, nº 1. Montreal: Les Presses de l'Université de Montreal. 168-173. Disponível em <http://www.erudit.org/revue/meta/1997/v42/n1/002090ar.pdf>, [acedido em 2005.06.03].
- CABRÉ, M. T. (2003). "Theories of terminology. Their description, prescription and explanation." In *Terminology: International journal of theoretical and applied issues in specialized communication*. John Benjamins Publishing Company: Volume 9, number 2, pp.163-199.
- CABRÉ, M. T. (1999). *Terminology: theories, methods and applications*. Volume 1. Ed. by Juan C. Sager. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- CABRÉ, M. T. (1998). "Importancia de la Terminologia en la Fijación de la Lengua". In *Terminologia: questões teóricas, métodos e projectos*. Cursos da Arrábida. Comissão Nacional para as Comemorações dos Descobrimentos Portugueses. Mateus, M. L., Correia, M. (Coordenação). Mem-Martins: Publicações Europa-América. Nº 4, pp.19-44.
- CABRÉ, M. T. (1995). "La terminologia, tendencias y aplicaciones." In *Ciencia da Informação*. Volume 24, Numero 3. Disponível em: <http://www.ibict.br/cienciadainformacao/viewissue.php?id=33>, [acedido em 2004.10.22].
- CASTILLO, R.A. (1998). "Ideas Sobre el Trabajo Terminográfico". In *Terminologia: questões teóricas, métodos e projectos*. Cursos da Arrábida. Comissão Nacional para as Comemorações dos Descobrimentos Portugueses. Mateus, M.L., Correia, M. (Coordenação). Mem-Martins: Publicações Europa-América. Nº 4, pp.97-119.
- CHUNG, T.M. (2003). "A corpus comparison approach for terminology extraction." In *Terminology: International journal of theoretical and applied issues in specialized communication*. Philadelphia/Amsterdam: John Benjamins Publishing Company. Volume 9, number 2, pp.221-245.
- COLUSSI, L. (2002). *A Reescritura da Informação Científica em Textos de Popularização da Ciência*. Dissertação de Mestrado. Santa Maria (Brasil): Universidade Federal de Santa Maria. Disponível em: <http://w3.ufsm.br/desireemroth/dissertacoes/lucolussi.pdf>, [acedido em 2005.04.11].

- GRUBER, T. R. (1993). *A translation approach to portable ontologies*. *Knowledge Acquisition*, 5 (2):199-220. Disponível em: <http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>. [acedido em 2003.09.20].
- ISO/FDIS 1087-1:2000, (E/F). *Terminology work – Vocabulary / Travaux terminologiques – Vocabulaire*. ISO 2000.
- JOHANSSON, S. (1991). "Times change, and so do corpora". In *English Corpus Linguistics. Studies in honour of Jan Svartvik*. Karin Aijmer e Bengt Altenberg (Eds). London and New York: Longman, pp.305-314.
- KAGEURA, K., DAILLE, B., NAKAGAWA, H., CHIEN, L.-F. (2004). "Introduction. Recent trends in computational terminology." In *Terminology: International journal of theoretical and applied issues in specialized communication*. John Benjamins Publishing Company: Volume 10, number 1, pp.1-21.
- KEMERLING, G. (2001), *Definition and Meaning*. Disponível em <http://www.philosophypages.com/lg/e05.htm>, [acedido em 2004.12.03].
- KRIEGER, M. G. (2000). *Terminology Revisited*. DELTA [on-line] 2000, Vol.16, nº 2, pp.209-228. Disponível em: <http://www.scielo.br/pdf/delta/v16n2/a01v16n2.pdf>, [acedido em 2005.02.20].
- LARA, M.L.G. (2004). "Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária." In *Ciência da Informação*. Brasília: v.33, n.2, pp.91-96, Maio/Agosto.2004. Disponível em: <http://www.pibict.br/cienciainformacao/include/getdoc.php?id=1089&article=304&mode=pdf> [acedido em 2004.12.15].
- LEECH, G. (1991). "The state of the art in corpus linguistics". In *English Corpus Linguistics. Studies in honour of Jan Svartvik*. Karin Aijmer e Bengt Altenberg (Eds). London and New York: Longman, pp.8-29.
- L'HOMME, M-C., HEID, U. e SAGER, J.C. (2003). "Terminology during the past decade (1994-2004). An Editorial Statement." In *Terminology: International journal of theoretical and applied issues in specialized communication*. John Benjamins Publishing Company: Volume 9, number 2, pp.151-161.

- I'HOMME, M-C. (2002). *What can verbs and adjectives tell us about terms?* Disponível em: <http://www.ling.umontreal.ca/lhomme/docs/tke2002.pdf>, [acedido em 2004.03.31].
- McENERY, T., WILSON, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- MAIA, B. (2003a). "Ontologies". In Lewandowska-Tomaszczyk, and P. J. Melia (Eds). *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang GmbH, pp.60-66.
- MAIA, B. (2003b). "Using corpora for terminology extraction: Pedagogical and computational approaches". In Lewandowska-Tomaszczyk, and P. J. Melia (Eds). *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang GmbH, pp.147-162.
- MAIA B. (2003c). "The pedagogical and linguistic research implication of GC to on-line parallel and comparable corpora". In *CP3A 2003: Corpora Paralelos, Aplicações e Algoritmos Associados*. Disponível em URL: http://poloclup.linguateca.pt/docs/cp3a/cp3a_pedag_art.pdf, [Acedido em 2004.01.03].
- MAIA B., SARMENTO L. (2003d) "Gestor de Corpora - Um ambiente Web para Linguística Computacional". In *CP3A 2003: Corpora Paralelos, Aplicações e Algoritmos Associados*. Disponível em: http://poloclup.linguateca.pt/docs/cp3a/cp3a_gc_art.pdf, [acedido em 2004.01.03].
- MAIA, B. (2002a). *What are comparable corpora?*, pp.27-34. Disponível em: <http://www.coli.uni-sb.de/muco03/maia.pdf>, [acedido em 2003.10.10].
- MAIA, B. (2002b). "Corpora for Terminology Extraction – the Differing Perspectives and Objectives of Researchers, Teachers and Language Services Providers". In *LREC Workshop #8: Language Resources for Translation Work and Research*. Disponível em: <http://www.ifi.unizh.ch/cl/yuste/postworkshop/repository/bmaia.pdf>, [acedido em 2004.01.03].
- MAIA, B. (1997). "Making corpora – a learning process". [Comunicação apresentada em Bertinoro *Corpus use and learning to translate*, em 14-15 Novembro de 1997]. Disponível em: <http://www.sslmit.unibo.it/cultpaps/maia.htm>, [acedido em 2004.01.03].

- MATEUS, M.L., CORREIA, M. (Coordenação). (1998). *Terminologia: questões teóricas, métodos e projectos*. In Cursos da Arrábida. Comissão Nacional para as Comemorações dos Descobrimentos Portugueses. Mem-Martins: Publicações Europa-América. Nº 4.
- MELLET, S. (2002). “Corpus et recherches linguistiques”. In *Corpus et recherches linguistiques*. Nº 1. Disponível em : <http://revel.unice.fr/corpus/document.html?id=7>, [acedido em 2004.05.25].
- NASCIMENTO, M. F. B. (2003). “O papel dos corpora especializados na criação de bases terminológicas”. Disponível em: http://www.clul.ul.pt/equipa/homenagem_mateus_2003_nascimento.pdf, [acedido em 2004.03.12].
- PAVEL, S., NOLET, D. (2002). *Manual de Terminologia*. Canada: Translation Bureau. Terminology and Standardization Directorate. (Tradução de Enilde Faulstich).
- PEARSON, J. (1998). *Terms in Context. Studies in Corpus Linguistics*. Volume 1. Amsterdam / Philadelphia: John Benjamins B.V.
- POINTER FINAL REPORT. *Proposals for an Operational Infrastructure for terminology in Europe*. BJL Consult s.a./n.v. (coordination). Bélgica. Disponível em: <http://www.computing.surrey.ac.uk/ai/pointer/report/>, [acedido em 2005.03.28].
- REY, A. (1995). *Essays on Terminology*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- RONDEAU, G. (1984). *Introduction à la Terminologie*. 2^{ème} edition. Québec : Gaëtan Morin Éditeur.
- SAGER, J.C. (2000). *Essays on Definition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- SAGER, J.C. (1990). *A practical course in terminology processing*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

- SARDINHA, T.B. (2000). *Linguística de Corpus: Histórico e Problemática*. D.E.L.T.A., vol.16, Nº 2, pp.323-367. Disponível em <http://scielo.br/pdf/delta/v16n2/a05v16n2.pdf> [acedido em 2003.09.15].
- SARDINHA, T.B. (1999). *Representatividade de Corpus*. Disponível em: http://www2.lael.pucsp.br/~tony/1999representativeness_usp.pdf, [acedido em 2004.09.25].
- SARMENTO, L., MAIA, B. & SANTOS, D. "The Corpógrafo - a Web-based environment for corpora research". In *Proceedings of LREC 2004*. Lisboa, Portugal. Disponível em: <http://poloclup.linguatca.pt/docs/cg/SarmientoMaiaSantosLREC2004.pdf>, [acedido em 2004.10.22].
- SILVA, S. (2004). "Reconhecimento de termos e produção de definições no direito administrativo: um estudo sob a ótica da lingüística do texto especializado e da lingüística de corpus." In *Revista Virtual de Estudos da Linguagem – ReVEL*. Ano 2, n. 3. Disponível em: http://planeta.terra.com.br/educacao/revel/edicoes/num_3/revel_3%20ling_corpus_direito_adm.pdf, [acedido em 2005.03.31].
- SONNEVELD, H. B., LOENING, K. L. (1993). *Terminology: applications in interdisciplinary communication*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- STUBBS, M. (1966). *Text and Corpus Analysis. Computer-assisted Studies of Language and Culture*. Oxford: Blackwell Publishers.
- TEMMERMAN, R. (2000). *Towards new ways of terminology description: the sociocognitive approach*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- TEUBERT, W. (1997). *Translation: ontologies for multilingual applications?* [Comunicação apresentada em Bertinoro "Corpus use learning to translate", a 14-15 Novembro de 1997]. Disponível em: <http://www.sslmit.unibo.it/cultpaps/paps.htm>. [acedido em 2003.09.03].
- UNESCO-OMM (s/data). *GLOSSÁRIO INTERNACIONAL DE HIDROLOGIA*. Pierre Hubert (org.). Disponível em: <http://www.cig.ensmp.fr/~hubert/glu/aglo.htm>, [acedido em 2005.08.20].

- VILELA, M. (1995). *Léxico e Gramática*. Ensino da Língua Portuguesa: Léxico, Dicionário, Gramática. Coimbra: Livraria Almedina.
- WALINSKI, J. (2003). "Automatic acquisition of terminology from comparable corpora". In Lewandowska-Tomaszczyk, and P. J. Melia (Eds). *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang GmbH, pp.69-88.
- WRIGHT, S. E., BUDIN, G. (1997). *Handbook of Terminology Management*. Volume 1. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- WÜSTER, E. (2003). "The Wording of the World presented graphically and terminologically." In *Terminology. International journal of theoretical and applied issues in specialized communication*. John Benjamins Publishing Company. Volume 9, number 2, pp.269-297.
- YZAGUIRRE, L. (1996). "Ingenería lingüística y terminología". In *Terminómetro. La Terminología en España*. União Latina. Número Especial: pp.69-71.

ANEXOS

ANEXO I

Tal como referimos na nota de esclarecimento em 4.1.3. (*Os marcadores lexicais indiciadores de informação terminológica e o Corpógrafo V2*), os elementos que se encontram dentro dos parênteses rectos representam as várias possibilidades combinatórias com os restantes elementos que compõem o que designamos por marcador linguístico.

MARCADORES LINGÜÍSTICOS INDICIADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA PORTUGUESA - SINGULAR]	
1.	... - "termo", ...
2.	... (termo)...
3.	... (ou termo)...
4.	termo -
5.	termo - é [o a um uma]
6.	termo (ou _ termo _)
7.	termo (sinónimo de _ termo _)
8.	termo (também _ termo _)
9.	termo, [tal como o termo indica, como o próprio nome indica,]
10.	[a o] termo [apresenta constitui contém corresponde define deriva tem]
11.	termo é [um conjunto] constituído por
12.	termo é parte de
13.	termo é uma espécie de
14.	termo pode definir-se como
15.	[o a] termo é [o a um uma]
16.	[o a] termo é constituído/a [por pela]
17.	[o a] termo é formado/a
18.	[o a] termo não é (outro termo), mas sim
19.	[o a] termo recebe [do(s) da(s)]
20.	[o a] denominado/a termo, caracterizado/a por
21.	a noção de termo
22.	a termo é definida como
23.	[aquele aquela] a que se chama de termo
24.	[conhecido conhecida] como termo
25.	considerar termo como sendo [o a um/a]
26.	Considera-se termo [ao aquele/a]
27.	designa-se [de por] termo
28.	designa-se [de por] termo [ao à]
29.	distinção entre termo e termo
30.	é denominado termo
31.	é [designado designada] [de por] termo
32.	entenda-se por termo como [o a um/a]
33.	este(a) termo
34.	(fala-se também de termo)
35.	[o a] termo - recebe [do/a]
36.	[o a] termo [apresenta contém corresponde deriva é]
37.	[o a] termo é constituído [pelo/a por]
38.	[o a] termo é definido como
39.	[o a] termo é formado
40.	[o a] termo é [o a]

41.	[o a] <i>termo</i> é [um/a]
42.	[o a] <i>termo</i> não é <i>termo</i> , mas sim
43.	[o a] conceito - <i>termo</i> - é considerado
44.	[o a] conceito de <i>termo</i> tem um significado
45.	[o a] conceito entendido no sentido mais restrito (amplo/abrangente, ...) <i>termo</i> significa
46.	[o a] denominado <i>termo</i> , caracterizado por
47.	[o a] <i>termo termo</i> encerra
48.	outra designação - <i>termo</i>
49.	[pelo/a] [chamado/a] <i>termo</i>
50.	poderá falar-se de <i>termo</i> que
51.	[um/a] <i>termo</i> é [um/a]

Quadro 7 – *Marcadores linguísticos indiciadores de informação terminológica* [Língua Portuguesa – Singular]

MARCADORES LINGÜÍSTICOS INDICADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA PORTUGUESA – PLURAL]	
1.	<i>termo</i> - como o próprio nome indica,...
2.	<i>termo</i> [constituem definem são têm]
3.	<i>termo</i> podem definir-se como
4.	<i>termo</i> são [parte de uma espécie de]
5.	<i>termo</i> , caracterizados/as pelos/as
6.	Aqueles/as a que se chamam de <i>termo</i>
7.	[os as] <i>termo</i> [apresentam contêm correspondem derivam recebem]
8.	[os as] <i>termo</i> não são (<i>outro termo</i>), mas sim
9.	[os as] <i>termo</i> recebem [do(s) da(s)]
10.	[os as] <i>termo</i> são
11.	[os as] <i>termo</i> são [as os]
12.	[os as] <i>termo</i> são constituídos/as [por pelos/as]
13.	[os as] <i>termo</i> são definidos/as como
14.	[os as] <i>termo</i> são formados/as
15.	[os as] denominados/as <i>termo</i> ,
16.	conhecidos/as como <i>termo</i>
17.	Consideram-se <i>termo</i> [os as aqueles/as]
18.	designam-se [de por] <i>termo</i>
19.	Designa-se por <i>termo</i> [aos às]
20.	distinção entre <i>termo</i> e <i>termo</i>
21.	[Entenda-se Entende-se] por <i>termo</i> [os as]
22.	[Entenda-se Entende-se] <i>termo</i> como [os as uns umas]
23.	estes/as <i>termo</i>
24.	os <i>termo</i> não são (<i>outro termo</i>), mas sim
25.	pelos/as chamados/as <i>termo</i>
26.	chamados/as <i>termo</i>
27.	(fala-se também de <i>termo</i>)
28.	que são denominados/as <i>termo</i>
29.	são designados/as [de por] _ <i>termo</i>
30.	são de <i>termo</i>

Quadro 8 – Marcadores linguísticos indiciadores de informação terminológica [Língua Portuguesa – Plural]

MARCADORES LINGÜÍSTICOS INDICADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA INGLESA – SINGULAR]	
1.	, or <i>term</i>
2.	, <i>term</i>
3.	, <i>term</i> plays [a an]
4.	, the <i>term</i> ,
5.	... (or <i>term</i>) ...
6.	... <i>term</i> ...
7.	[a an the] <i>term</i> [has refers to represents results from serves as tells us [about] functions as conducts consists of contains controls has occurs [when] plays]
8.	[a an the] <i>term</i> is [a an the]
9.	[is can be] called [a an the] <i>term</i>
10.	[is] known as <i>term</i>
11.	a [simple] definition of <i>term</i>
12.	forming what we call <i>term</i>
13.	like (<i>outro termo</i>), [the] <i>term</i>
14.	<i>term</i> - [a an the brings takes this is]
15.	<i>term</i> - ...
16.	<i>term</i> - is [a an the]
17.	<i>term</i> - this is
18.	<i>term</i> (or <i>term</i>) ...
19.	<i>term</i> [is can be] [defined as determined by [also] known as initiated located in made in the made [up] of named as released used to initiated part of composed of constituted by part of that part where]
20.	<i>term</i> is
21.	<i>term</i> is a [kind of system of]
22.	<i>term</i> is the main
23.	<i>term</i> is the way in which
24.	<i>term</i> the manner in which
25.	<i>term</i> , as this <i>term</i> describes
26.	<i>term</i> , or
27.	<i>term</i> , or <i>term</i> ,
28.	<i>term</i> , which is
29.	the <i>term</i>
30.	This <i>term</i> is termed <i>term</i> .
31.	We [consider speak of] <i>term</i> as

Quadro 9 – Marcadores linguísticos indicadores de informação terminológica [Língua Inglesa – Singular]

MARCADORES LINGÜÍSTICOS INDICIADORES DE INFORMAÇÃO TERMINOLÓGICA [LÍNGUA INGLESA – PLURAL]	
1.	, <i>term</i> play [a an]
2.	, or <i>term</i>
3.	, <i>term</i>
4.	, the <i>term</i> ,
5.	... (or <i>term</i>) ...
6.	... <i>term</i> ...
7.	[are can be] called [a na the] <i>term</i>
8.	[are] known as <i>term</i>
9.	a [simple] definition of <i>term</i>
10.	forming what we call <i>term</i>
11.	like (<i>outro termo</i>), the <i>term</i>
12.	<i>term</i> - [a an the bring take these are]
13.	<i>term</i> - ...
14.	<i>term</i> - are [a an the]
15.	<i>term</i> - these are
16.	<i>term</i> (or <i>term</i>) ...
17.	<i>term</i> , which are
18.	<i>term</i> [are can be] [defined as determined by [also] known as located in made in the made [up] of named as released used to initiated part of composed of constituted by those parts where]
19.	<i>term</i> [have refer to represent result from serve as tell us [about] function as conduct consist of contain control have occur [when] play]
20.	<i>term</i> are
21.	<i>term</i> are [a an the]
22.	<i>term</i> are a [kind of system of]
23.	<i>term</i> are the main
24.	<i>term</i> are the way in which
25.	<i>term</i> the manner in which
26.	<i>term</i> , as this <i>term</i> describes
27.	<i>term</i> , or
28.	<i>term</i> , or <i>term</i> ,
29.	<i>term</i> , which are
30.	These <i>term</i> are termed <i>term</i> .
31.	We [consider speak of] <i>term</i> as

Quadro 10 – Marcadores linguísticos indiciadores de informação terminológica [Língua Inglesa – Plural]

PADRÕES SUPORTE UTILIZADOS PELO Corpógrafo V3 [LÍNGUA PORTUGUESA - SINGULAR]	
1.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (uma um o a)
2.	(^o ^a ^um ^uma o a um uma)\w+ __TERMO__ é (uma um) (dos das) (variad numeros)(os as)
3.	(^ o ^a ^um ^uma o a um uma) __TERMO__ - é (uma um o a)
4.	(^entende-se ^designa-se ^denomina-se entende-se designa-se denomina-se)(por de como) __TERMO__
5.	(^ é é)(chamad denominad designad)(o a) __TERMO__
6.	conhecid(o a) como __TERMO__
7.	(^ o ^a ^um ^uma o a um uma) __TERMO__ (recebe resulta provém) (do da dos das de)
8.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (constituíd determinad compost)(o a) (por pelo pela pelos pelas de)
9.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (produzid formad segregad)(o a) (por pelo pela pelos pelas de no na nos nas)
10.	(^ denomina-se ^designa-se ^chama-se denomina-se designa-se chama-se) __TERMO__
11.	(^é é)(designad chamad denominad)(o a) (por de) __TERMO__
12.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (o a) principal
13.	(^ o ^a ^um ^uma o a um uma) __TERMO__ : *
14.	(^cada cada) __TERMO__ (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
15.	(^cada cada)\w+ de __TERMO__ (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
16.	(^ o ^a ^um ^uma o a um uma) __TERMO__ (do da)\w+ (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
17.	(^ o ^a ^um ^uma o a um uma) __TERMO__ depende (de da do das dos)
18.	(^é é)(o a) __TERMO__ \.
19.	(^que que) se (designa denomina entende chama) (por de) __TERMO__
20.	- __TERMO__ -
21.	(^as as)(funções características propriedades especificidades estruturas divisões) (do da) __TERMO__ são
22.	(^quando quando) ocorre (o a uma um) __TERMO__
23.	(^desencadeia ^causa ^provoca ^origina desencadeia causa provoca origina)(o a um uma) __TERMO__
24.	(^ o ^a ^um ^uma o a um uma) __TERMO__ está (protegid alojad localizad instalad situad)(o a)

25.	(^estruturas ^funções ^características ^propriedades ^especificidades estruturas funções características propriedades especificidades)(dominantes principais básicas) (do da) __TERMO__
26.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ pode , \w+ , ser (aproveitad utilizad usad)(o a) (para no na)
27.	(^por por)definição , (o a) __TERMO__ (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
28.	(^ ^o ^a ^um ^uma o a um uma)\w+ (forma constitui) (o a) __TERMO__
29.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ (conduz induz provoca causa leva) (o a ao à)
30.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ que (ocorre acontece) quando
31.	(^dá-se ^ocorre dá-se ocorre)por __TERMO__
32.	(^ ^no ^na no na) __TERMO__ existem
33.	(^utiliza-se ^usa-se utiliza-se usa-se)o termo __TERMO__ para designar
34.	(^as as)(funções características propriedades especificidades estruturas divisões) (dominantes principais básicas) (do da) __TERMO__ :
35.	__TERMO__ - (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
36.	__TERMO__ - (zona área)
37.	__TERMO__ - uma (zona área)
38.	__TERMO__ - também (denominad designad chamad conhecid considerad)(o a)
39.	(^constitui ^forma ^compõe constitui forma compõe)(o a) __TERMO__
40.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ deve ser (vist interpretad entendid compreendid)(o a) como
41.	(^denominamos ^designamos denominamos designamos) __TERMO__ quando
42.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
43.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ (actua age) sobre
44.	__TERMO__ , que não é mais que
45.	(^corresponde corresponde)(ao à) __TERMO__
46.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ é (transformad convertid)(o a) em
47.	, (chamad denominad designad)(o a) __TERMO__ ,
48.	(designad denominad chamad)(o a) __TERMO__ , que
49.	(designad denominad chamad)(o a) __TERMO__
50.	(^que que) se (denomina designa) de __TERMO__
51.	__TERMO__ , que (banha compreende contém engloba possui liga) (tod vári múltipl variadíssim)(as os)

52.	(^ o ^a ^um ^uma o a um uma) __TERMO__ , (constituíd formad determinad compost segregad produzid induzid libertad hidrolisad)(o a) (por pelas pelos pelo pela)
53.	(^ ^como como) __TERMO__ considera-se (o a)
54.	__TERMO__ é , por definição ,
55.	(^é é)(um uma) \w+ designad(a)o por __TERMO__
56.	(^ o ^a ^um ^uma o a um uma) __TERMO__ está encarregue de
57.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (definid descri vist entendid caracterizad conhecid)(a)o como
58.	(^ o ^a ^um ^uma o a um uma) __TERMO__ , que é (o a um uma)
59.	(^ o ^a ^um ^uma o a um uma) __TERMO__ que é (constituíd formad compost iniciad utilizad aplicad libertad segregad calculad feit elaborad executad realizad produzid induzid causad provocad determinad caracterizad descri defini d vist entendid considerad)(o a)
60.	(^ o ^a ^um ^uma o a um uma) __TERMO__ expressa-se em
61.	(^ o ^a ^um ^uma ^ao ^à o a um uma ao ^à) __TERMO__ que (conduz acontece regula actua causa origina ocorre se dá determina corresponde contem contém existe suporta deriva traz recebe liga limita defin e desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz se deve exprime permite faz suporta segrega transmite depende hidrolisa) (o a)
62.	(^o o)(termo conceito) __TERMO__ que (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz deve-se exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
63.	(^ o ^a ^um ^uma o a um uma) __TERMO__ .*, também (conduz acontece regula actua causa origina ocorre se dá determina corresponde contem contém existe suporta deriva traz recebe liga limita defin e desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz se deve exprime permite faz suporta segrega transmite hidrolisa) (o a ao ^à)
64.	(^ no ^na no na) __TERMO__ , (um uma o a)
65.	__TERMO__ , (ou seja isto é) , (um uma o a)
66.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (característic típica)(a)o (de do da)
67.	(^ o ^a ^um ^uma o a um uma) __TERMO__ pode ser
68.	(^ o ^a ^um ^uma o a um uma) __TERMO__ de (um uma) \w+ define-se como (o a um uma)
69.	(^ o ^a ^um ^uma o a um uma) __TERMO__ pode ser (constituíd formad compost iniciad utilizad aplicad libertad segregad calculad feit elaborad executad realizad produzid induzid causad provocad determinad caracterizad descri defini d vist entendid considerad)(o a)
70.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é igual (a ao ^à)
71.	(^ o ^a ^um ^uma o a um uma) __TERMO__ \(.*) (ou é) (um uma o a)
72.	(^ o ^a ^um ^uma o a um uma) __TERMO__ , (isto é ou seja) ,
73.	(^ ^por por) __TERMO__ entende-se
74.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (o a) principal
75.	aquilo que (chamamos designamos denominamos) de __TERMO__
76.	(^é é)conhecid(o a) como __TERMO__
77.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é feit(o a) (de no na em)

78.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é \w+ \w+ (do da dos das de que)
79.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é \w+ (do da dos das)
80.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é onde (o a os as)
81.	(^ o ^a ^um ^uma um uma o a) __TERMO__ tem (muit vári múltipl variad inúmer)(os as)
82.	(^ o ^a ^um ^uma um uma o a) __TERMO__ tem (alguns algumas)
83.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (a o) (forma modo maneira) como
84.	(^ o ^a ^um ^uma o a um uma) __TERMO__ consiste em
85.	(^ o ^a ^um ^uma o a um uma) __TERMO__ que (conduz acontece regula actua causa origina ocorre se dá determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz se deve exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
86.	(^ o ^a ^um ^uma o a um uma) __TERMO__ (situa-se encontra-se localiza-se) (no na em)
87.	(^ o ^a ^um ^uma o a um uma) __TERMO__ (liga-se designa-se estende-se torna-se origina-se dá-se compõe-se)
88.	(^cabe cabe)(ao à) __TERMO__
89.	(^ o ^a ^um ^uma o a um uma) __TERMO__ é (constituíd formad compost iniciad utilizad aplicad libertad segregad calculad feit elaborad executad realizad produzid induzid causad provocad determinad caracterizad descri definid vist entendid considerad)(o a)
90.	(^ o ^a ^um ^uma o a um uma) __TERMO__ : (um uma o a)
91.	(^ no ^na ^em no na em) __TERMO__ (o a um uma)
92.	, (o a) __TERMO__ \.
93.	__TERMO__ - isto é
94.	__TERMO__ - (um uma o a)
95.	(^ o ^a ^um ^uma o a um uma) __TERMO__ serve (de para)
96.	(^ o ^a ^um ^uma o a um uma) __TERMO__ tem sido (caracterizad vist descri considerad entendid referid)(o a)
97.	__TERMO__ , (contendo possuindo compreendendo)
98.	(^ o ^a ^um ^uma o a um uma) __TERMO__ tem a (propriedade característica) de
99.	(^utiliza-se ^usa-se utiliza-se usa-se) __TERMO__ para designar
100.	__TERMO__ , que (conduz acontece regula actua causa origina ocorre se dá determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz se deve exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
101.	(^ o ^a ^um ^uma o a um uma) __TERMO__ toma (designações denominações)
102.	(^é ^é)produzid(o a) por __TERMO__
103.	(^é ^é)(o a) __TERMO__ que
104.	(^ o ^a ^um ^uma o a um uma) __TERMO__ não é .* , mas sim
105.	(^ o ^a ^um ^uma o a um uma) __TERMO__ dest(as es) \w+ (tem possui contém)
106.	(^diz-se diz-se)que (o a) \w+ está (no na) __TERMO__

107.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ ou \w+ , que (conduz acontece regula actua causa origina ocorre se dá determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla engloba descreve induz se deve exprime resulta consiste cresce circula permite faz suporta segrega transmite depende hidrolisa)
108.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ pode .* ser (constituíd formad compost iniciad utilizad aplicad libertad segregad calculad feit elaborad executad realizad produzid induzid causad provocad determinad caracterizad descri definid vist entendid considerad)(o a)
109.	(^uma uma)(definição explicação) (simples básica) de __TERMO__
110.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ .* , (bem assim) como \w+ é capaz de (produzir fazer estabelecer)
111.	(^a a) essência (do da) __TERMO__ \ (.* \) (consiste está)
112.	(^dá dá) pelo nome de __TERMO__
113.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ que através de
114.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ (o a) qual é (o a) responsável (pelo pela pelos pelas)
115.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ ou \w+ \w+ dá e mantém
116.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ não envolve apenas
117.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ não é (um uma)
118.	remete-nos para a definição de __TERMO__
119.	(^uma uma) das (características estruturas funções particularidades) (essenciais principais básicas fundamentais) (de do da) __TERMO__ é (a o)
120.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ \ (que é .* \)
121.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ \ (uma .* \)
122.	(^do do) __TERMO__ que é (o a) (sistema função estrutura)
123.	(^objectivo objectivo) d(e o a) __TERMO__
124.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ , também referid(o a) como
125.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ e (o a) .* afectam (o a)
126.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ (frequentemente muitas vezes) surge
127.	(^quer quer) (na no) __TERMO__ quer (no na) .* existem
128.	(^quer quer) (na no) \w+ quer (no na) __TERMO__ existem
129.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ \ (ou seja .* \)
130.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ , como .* , é responsável (por pelo pela pelos pelas)
131.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ caracteriza-se (por pelo pela)
132.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ , (a o) qual
133.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ , por definição ,
134.	__TERMO__ - .* , (conduz acontece regula actua causa origina ocorre dá-se determina corresponde contem contém existe suporta deriva traz recebe liga limita define desempenha representa domina apresenta produz possibilita possui tem funciona controla)
135.	(^ ^o ^a ^um ^uma o a um uma) __TERMO__ , (um uma o a) .* , é (usad utilizad)(o a)

Quadro 11 – *Padrões Suporte* utilizados pelo *Corpógrafo V3* [Língua Portuguesa – Singular]

PADRÕES SUPORTE UTILIZADOS PELO Corpógrafo V3 [LÍNGUA PORTUGUESA - PLURAL]	
1.	(^ ^os ^as os as) __TERMO__ são \w+ (especializad específic típic característic revestid)(os as)
2.	(^ ^as ^os as os) __TERMO__ são (segregad formad constituíd determinad definid descrit conhecid vist entendid caracterizad denominad chamad designad)(as os)
3.	(^designa-se ^entende-se designa-se entende-se)por __TERMO__
4.	(^que que)são (segregad formad constituíd determinad definid descrit conhecid vist entendid caracterizad denominad chamad designad)(as os) __TERMO__
5.	(^ ^os ^as os as) __TERMO__ (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuam funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
6.	(^denominam-se ^designam-se ^chamam-se denominam-se designam-se chamam-se)(de por) __TERMO__
7.	(^que que)se (designam denominam chamam) (por de) __TERMO__
8.	(^que que)se (denomina designa chama) (de por) __TERMO__
9.	(^ ^os ^as os as) __TERMO__ não são .*, mas sim
10.	__TERMO__ :.*
11.	(^ ^os ^as os as) __TERMO__ (originam-se designam-se ligam-se dão-se estendem-se)
12.	(^existem existem)\w+ (classes tipos) de __TERMO__
13.	(^outro outro)tipo de \w+, (as os) __TERMO__,
14.	(^chamad ^designad ^denominad chamad designad denominad)(os as) __TERMO__
15.	__TERMO__ ou \w+, que (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuam funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
16.	(^ ^os ^as os as) __TERMO__, .*, que
17.	(^ ^os ^as os as) __TERMO__ apresentam as seguintes (características propriedades funcionalidades particularidades)
18.	(^são são) (segregad formad constituíd determinad definid descrit conhecid vist entendid caracterizad denominad chamad designad produzid)(as os) por __TERMO__
19.	(^são são)(as os) __TERMO__ que (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuam funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam se dão originam causam acontecem trazem resultam segregam englobam descrevem induzem se devem exprimem consistem crescem circulam permitem fazem transmitem dependem)
20.	(^ ^os ^as os as) __TERMO__, (bem assim) como .* são capazes de (produzir fazer estabelecer)
21.	(^cabe cabe) (às aos) __TERMO__

22.	(^nos ^nas nos nas)__TERMO__ , (os as) .* (ocorrem originam causam acontecem fazem induzem resultam actuum)
23.	(^ ^nos ^nas nos nas)__TERMO__ (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
24.	(^ ^os ^as os as)__TERMO__ mais (abundantes comuns importantes) são
25.	(^ ^os ^as os as)__TERMO__ , (contendo possuindo compreendendo apresentando actuando tendo englobando consistindo)
26.	(^a a)maioria (das dos) __TERMO__ (deriva apresenta contém corresponde produz determina existe limita recebe possui controla regula possibilita actua funciona ocorre liga define desempenha representa domina conduz tem suporta dá-se origina causa acontece traz resulta segrega engloba descreve induz deve-se exprime consiste cresce circula permite faz transmite depende)
27.	(^ ^como ^por como por)__TERMO__ (considera-se entende-se)
28.	(^ ^os ^as os as)__TERMO__ têm a (propriedade característica função funcionalidade) de
29.	(^utiliza-se ^usa-se utiliza-se usa-se)o (termo conceito) __TERMO__ para designar
30.	(^ ^os ^as os as)__TERMO__ têm (muit múlti variad inúmer vári)(os as)
31.	(^ ^os ^as os as)__TERMO__ têm (alguns algumas dois duas um uma)
32.	(^ ^os ^as os as)__TERMO__ ou \w+ - (os as) que (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam se dão originam causam acontecem trazem resultam segregam englobam descrevem induzem se devem exprimem consistem crescem circulam permitem fazem transmitem dependem)
33.	(^ ^os ^as os as)__TERMO__ (do da) .* (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
34.	(^ ^os ^as os as)__TERMO__ , que (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam se dão originam causam acontecem trazem resultam segregam englobam descrevem induzem se devem exprimem consistem crescem circulam permitem fazem transmitem dependem)
35.	(^ ^os ^as os as)__TERMO__ que tomam designações
36.	(^em em)__TERMO__ (a o)
37.	(^ ^os ^as os as)__TERMO__ , isto é ,
38.	(^uma uma)definição (simples básica comum) de __TERMO__
39.	(^ ^os ^as os as)__TERMO__ são (aquelas as) partes (de do da dos das que)
40.	(^ ^os ^as os as)__TERMO__ (são fazem) parte (de do da)
41.	(^ ^os ^as os as)__TERMO__ servem (de para)

42.	, (os as) __TERMO__ , (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
43.	(^ os as os as) __TERMO__ podem ser (entendid caracterizad descrit considerad vist definid produzid causad feit executad realiz calculad)(os as)
44.	(^ os as os as) __TERMO__ têm sido (entendid caracterizad descrit considerad vist definid)(os as)
45.	(^ os as os as) __TERMO__ que (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam se dão originam causam acontecem trazem resultam segregam englobam descrevem induzem se devem exprimem consistem crescem circulam permitem fazem transmitem dependem)
46.	(^ os as os as) __TERMO__ (compõem-se tornam-se)
47.	(^ os as os as) __TERMO__ são (compost constituíd formad)(os as) (por de pelo pela pelos pelas)
48.	(^ os as os as) __TERMO__ são (usad utilizad aplicad transformad)(os as)
49.	(^ os as os as) __TERMO__ estão encarregues de
50.	(^ os as os as) __TERMO__ são responsáveis (pelo pela pelos pelas)
51.	(^ os as os as) __TERMO__ estão (localizad situad alojad protegid)(os as)
52.	(^ os as ^estes ^estas os as estes estas) __TERMO__ (situam-se localizam-se encontram-se)
53.	(^ os as os as) __TERMO__ que (ocorrem acontecem se dão) quando
54.	(^quando quando)(ocorrem acontecem se dão) __TERMO__
55.	(^dá-se ^ocorre dá-se ocorre)por __TERMO__
56.	- __TERMO__ -
57.	(^as as)(funções características especificidades funcionalidades) (dos das) __TERMO__ são
58.	(^desencadeia ^causa ^provoca desencadeia causa provoca)(os as) __TERMO__
59.	(^as as)(divisões estruturas funções partes unidades) (dominantes principais básicas fundamentais) (dos das) __TERMO__
60.	(^ os as os as) __TERMO__ podem , também , ser (utilizad usad)(os as) (para no na)
61.	(^por por)definição , (os as) __TERMO__ (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
62.	(^os as os as).* (formam constituem) (os as) __TERMO__
63.	(^ nos nas nos nas) __TERMO__ existem
64.	(^as as)(divisões estruturas funções partes unidades) (dominantes principais básicas fundamentais) (dos das) __TERMO__ :
65.	(^o o)(termo conceito) __TERMO__ que
66.	(^ os as os as) __TERMO__ - (zona área)
67.	(^ os as os as) __TERMO__ - uma (zona área)

68.	(^ os as os as)__TERMO__ - também (designad chamad conhecid denominad)(os as)
69.	(^ os as os as)__TERMO__ devem ser (interpretad definid vist)(os as) como
70.	. * - __TERMO__ \.
71.	(^denominamos designamos denominamos designamos)__TERMO__ quando
72.	(^entende-se define-se entende-se define-se)como __TERMO__
73.	(^ os as os as)__TERMO__ , que não são mais que
74.	(^correspondem correspondem)(aos às) __TERMO__
75.	. * , (chamad denominad designad)(os as) __TERMO__ ,
76.	(^ os as os as)__TERMO__ , que (compreendem banham incluem) tod(as os)
77.	(^ os as os as)__TERMO__ , (formad constituíd determinad compost segregad)(os as)
78.	(^ os as os as)__TERMO__ são , por definição ,
79.	(^ os as os as)__TERMO__ dest(as es) \w+ (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
80.	(^ os as os as)__TERMO__ expressam-se em
81.	(^ os as os as)__TERMO__ são (característic típic)(as os) (de dos das do da)
82.	(^ os as os as)__TERMO__ podem ser
83.	(^ os as os as)__TERMO__ de (um uma) \w+ definem-se como (o a)
84.	(^ os as os as)__TERMO__ . * são (determinad constituí formad compost segregad)(as os) pel(os as)
85.	(^ os as os as)__TERMO__ , que são
86.	(^ os as os as)__TERMO__ \ (. * \) são
87.	(^ os as os as) . * designam-se por __TERMO__
88.	(^ os as os as)__TERMO__ ao considerar como tal
89.	(^ os as os as)__TERMO__ são (os as) principais
90.	(^aquilo aquilo)que (designamos chamamos denominamos) de __TERMO__
91.	(^ os as os as)__TERMO__ são (feit iniciad libertad)(os as)
92.	(^uma uma)definição de __TERMO__
93.	(^ os as os as)__TERMO__ podem \w+ ser (definid caracterizad chamad)(os as) como
94.	(^ os as os as)__TERMO__ : (os as)
95.	, (os as) __TERMO__ \.
96.	(^ os as os as)__TERMO__ - (derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuum funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
97.	(^ os as os as)__TERMO__ - isto é
98.	(^ os as os as)__TERMO__ - (os as)
99.	(^ os as os as)__TERMO__ \ (. * \) podem \w+ ser definid(os as)
100.	(^ os as os as)__TERMO__ , \ (. * \) que são
101.	(^a a)essência (dos das) __TERMO__ \ (. * \) consiste

102.	(^ os as os as)__TERMO__ (agem actuem) sobre
103.	(^são são)(denominad chamad designad)(os as) __TERMO__
104.	(^dão dão)pelo nome de __TERMO__
105.	(^ os as os as)__TERMO__ que através de
106.	(^ os as os as)__TERMO__ (os as) quais são (os as) responsáveis (pelo pela pelos pelas)
107.	(^ os as os as)__TERMO__ não são (\w+){2,3} , mas
108.	remete-nos para a definição de __TERMO__
109.	(^uma uma)das (características funções funcionalidades especificidades estruturas partes) (essenciais principais fundamentais básicas primordiais) (dos das) __TERMO__ é (a o)
110.	(^ os as os as)__TERMO__ \ (que são .* \)
111.	(^ os as os as)__TERMO__ é o termo .* que define
112.	(^ os as os as)__TERMO__ \ (.* \) \.
113.	(^ dos ^das dos das)__TERMO__ que são
114.	(^objectivo objectivo)(dos das) __TERMO__
115.	(^ os as os as)__TERMO__ , também referid(os as) como
116.	(^ os as os as)__TERMO__ \ (ou seja .* \)
117.	__TERMO__ caracterizam-se (por pelo pela)
118.	(^ os as os as)__TERMO__ e (o a os as) .* afectam (o a os as)
119.	(^ os as os as)__TERMO__ (frequentemente muitas vezes) surgem
120.	(^quer quer)(nos nas) __TERMO__ quer (nos nas) .* existem
121.	(^quer quer)(nos nas) \w+ quer (nos nas) __TERMO__ existem
122.	(^ os as os as)__TERMO__ , por exemplo , (são derivam apresentam contêm correspondem produzem determinam existem limitam recebem possuem controlam regulam possibilitam actuem funcionam ocorrem ligam definem desempenham representam dominam conduzem têm suportam dão-se originam causam acontecem trazem resultam segregam englobam descrevem induzem devem-se exprimem consistem crescem circulam permitem fazem transmitem dependem)
123.	(^ os as os as)__TERMO__ dividem-se , .* , em
124.	(^são são)(os as) __TERMO__ (os as) principais responsáveis (por pelo pela pelas pelos)
125.	(^ os as os as)__TERMO__ são (igualmente também) responsáveis (por pelo pela pelas pelos)
126.	(^ os as os as)__TERMO__ se encontram agrupad(os as) (nos nas)
127.	(^são são) (as os) __TERMO__ ou .* que
128.	(^ os as os as)__TERMO__ são (os as) \w+ (dominantes principais)
129.	(^a a) maior parte (dos das) __TERMO__ apresentam
130.	(^nos nos) \w+ , (as os) __TERMO__ são produzid(os as)

Quadro 12 – *Padrões Suporte* utilizados pelo *Corpógrafo V3* [Língua Portuguesa – Plural]

PADRÕES SUPORTE UTILIZADOS PELO CORPÓGRAFO V3 [LÍNGUA INGLESA - SINGULAR]	
1.	(?<!of)(^ the ^a ^an the a an)__TERMO__ is (a an the)
2.	(^called ^named called named)__TERMO__
3.	(^is is)called __TERMO__
4.	(^is is)called (a an the) __TERMO__
5.	(^is is)known as __TERMO__
6.	(^known known)as __TERMO__
7.	(^ the ^a ^an the a an)__TERMO__ is made of
8.	(^ the ^a ^an the a an)__TERMO__ is made up of
9.	(^ the ^a ^an the a an)__TERMO__ is the main
10.	(^ the ^a ^an the a an)__TERMO__ is (that the) part of
11.	(^ the ^a ^an the a an)__TERMO__ is (part composed) of
12.	(^ the ^a ^an the a an)__TERMO__ is where
13.	(^ the ^a ^an the a an)__TERMO__ (serves functions works) as
14.	(^ the ^a ^an the a an)__TERMO__ \ (.* \) is (a an the)
15.	(^ the ^a ^an the a an)__TERMO__ has (many several a an all one two various the)
16.	(^ the ^a ^an the a an)__TERMO__ is (initiated protected secreted calculated considered spread defined released produced described regarded viewed)
17.	(^ the ^a ^an the a an)__TERMO__ tells about
18.	(^ the ^a ^an the a an)__TERMO__ can be (characterised characterized defined described regarded considered viewed used transformed changed calculated initiated secreted protected divided)
19.	(^ the ^a ^an the a an)__TERMO__ is the way in which
20.	(^ the ^a ^an ^each the a an each)__TERMO__ (results releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings consists corresponds leads depends takes receives derives plays)
21.	(^a a)(simple basic common) definition of __TERMO__
22.	(^a a)definition of __TERMO__
23.	(^ the ^a ^an the a an)__TERMO__ can \w+ be (characterised characterized defined described regarded considered viewed)
24.	(^ the ^a ^an the a an)__TERMO__ is used to
25.	(^ the ^a ^an the a an)__TERMO__ is (located expressed made) in
26.	(^like like).* , the __TERMO__
27.	(^like like).* , __TERMO__
28.	, the __TERMO__ , (results releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings consists corresponds leads depends takes receives derives plays)
29.	__TERMO__ : (a an the)
30.	(^at at)the __TERMO__ (the a an)
31.	(^name name)as the __TERMO__
32.	, (the a an) __TERMO__ \.

33.	__TERMO__ - (releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives plays)
34.	__TERMO__ - this is
35.	__TERMO__ - (a an the)
36.	__TERMO__ - is (a an the)
37.	__TERMO__ - set of
38.	__TERMO__ , which is (a an the)
39.	(^we we)define as __TERMO__
40.	__TERMO__ \ (. * \) can \w+ be (defined described regarded considered viewed)
41.	__TERMO__ \ (. * \) , which is (a an the)
42.	(^by by) __TERMO__ we understand
43.	(^which ^what which what)is called __TERMO__
44.	(^the ^a ^an the a an) __TERMO__ is (formed constituted) by
45.	(^the ^a ^an the a an) __TERMO__ \ (. * \) \.
46.	(^the ^a ^an the a an) __TERMO__ is not . * , but
47.	(^each each)\w+ of __TERMO__ (releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives plays)
48.	(^the ^a ^an the a an) __TERMO__ of . * (releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives plays)
49.	(^that that)is (called named) __TERMO__
50.	- __TERMO__ -
51.	(^when when)(a an the) __TERMO__ (occurs takes place)
52.	(^causes causes)(a an the) __TERMO__
53.	(^dominant ^main ^basic ^important dominant main basic important)(structures parts divisions) of __TERMO__
54.	(^the ^a ^an the a an) __TERMO__ can also be (used utilized) (to in)
55.	(^by by)definition, (the a an) __TERMO__ (releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives plays)
56.	(^the the)\w+ (forms constitutes composes) the __TERMO__
57.	(^the ^a ^an the a an) __TERMO__ that (occurs takes place) when
58.	(^occurs occurs)by __TERMO__
59.	(^in in) __TERMO__ there (are is)
60.	(^in in)(the a an) __TERMO__ there (are is)
61.	(^we we)use the term __TERMO__ to (describe define mean denote refer)
62.	(^the the)(basic main dominant) (divisions parts structures) of __TERMO__ :
63.	(^the ^a ^an the a an) __TERMO__ - (releases prevents holds dominates causes originates represents controls contains conducts defines spreads influences determines becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives plays)
64.	(^the ^a ^an the a an) __TERMO__ - (zone area)

65.	(^the a an the a an) __TERMO__ - (a an the) (zone area)
66.	(^the a an the a an) __TERMO__ - also (called known named)
67.	(^the a an the a an) __TERMO__ (should could) be (understood seen regarded viewed defined) as
68.	- __TERMO__ \.
69.	(^the a an the a an) __TERMO__ that is nothing more
70.	(^that that) __TERMO__ occurs
71.	(^the a an the a an) __TERMO__ is transformed into
72.	, called __TERMO__ ,
73.	(^the a an the a an) __TERMO__ that (bathes occupies absorbes) all
74.	(^the a an the a an) __TERMO__ , (formed constituted determined) by
75.	(^the the).* is called __TERMO__
76.	(^we we)consider __TERMO__ (a an the)
77.	(^the a an the a an) __TERMO__ is by definition
78.	(^is is)a structure called __TERMO__
79.	(^the a an the a an) __TERMO__ is in charge of
80.	(^the a an the a an) __TERMO__ of these \w+ (has possesses)
81.	(^the a an the a an) __TERMO__ (that which also) (prevents is holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives releases plays)
82.	(^the a an the a an) __TERMO__ .*, also (prevents is holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives releases plays)
83.	(^during during)(the a an) __TERMO__
84.	(^we we)say that the \w+ is in __TERMO__
85.	(^in in)(the a an) __TERMO__ , (a an)
86.	(^the the)term __TERMO__ that
87.	(^the a an the a an) __TERMO__ , that is (a an)
88.	(^the a an the a an) __TERMO__ is (characteristic typical) from
89.	(^the a an the a an) __TERMO__ of (a an) \w+ is defined as
90.	(^the a an the a an) __TERMO__ is the same as
91.	(^as as) __TERMO__ we (consider understand)
92.	(^the a an the a an) __TERMO__ (presents has) the following (traits features characteristics functions)
93.	(^the a an the a an) __TERMO__ , which (prevents is holds dominates causes originates represents controls contains conducts defines spreads influences determines produces becomes presents occurs describes regulates acts enables equals brings results consists corresponds leads depends takes receives derives releases plays)
94.	(^is ^are is are)(produced made) by __TERMO__
95.	(^which which)are (called named) __TERMO__
96.	(^is is)the __TERMO__ (that which)
97.	(^in in)the __TERMO__ (the a an) .* (occurs takes place)
98.	(^the the)(most common commonest) __TERMO__

99.	(^the ^a ^an the a an)__TERMO__ is responsible for
100.	(^we we)use __TERMO__ to (designate refer denote describe)
101.	(^the ^a ^an the a an)__TERMO__ is used to (designate refer denote describe)
102.	(^the ^a ^an the a an)__TERMO__ , that is ,
103.	(^the ^a ^an the a an)__TERMO__ , i.e\.
104.	(^the ^a ^an the a an)__TERMO__ , (the an a) \w+ (which that)
105.	(^the ^a ^an the a an)__TERMO__ (may can) also be \w+ed by
106.	(^the ^a ^an the a an)__TERMO__ \ (the one that .* \)
107.	(^definition definition): __TERMO__ \ (.* \) is (a an the)
108.	(^the the)cause of __TERMO__ is
109.	(^the ^a ^an the a an)__TERMO__ such as
110.	(^there there)are (groups sets types classes categories) of __TERMO__
111.	(^groups ^sets ^types ^classes ^categories groups sets types classes categories)of __TERMO__
112.	(^groups ^sets ^types ^classes ^categories groups sets types classes categories)of __TERMO__ :
113.	(^the ^a ^an the a an)__TERMO__ \ (the .* \)
114.	(^area ^location ^zone area location zone)called __TERMO__
115.	(^common common)__TERMO__ \w+ are :
116.	\ (called the __TERMO__ \)

Quadro 13 – *Padrões Suporte* utilizados pelo *Corpógrafo V3* [Língua Inglesa – Singular]

PADRÕES SUPORTE UTILIZADOS PELO CORPÓGRAFO V3 [LÍNGUA INGLESA – PLURAL]	
1.	(^ the the)__TERMO__ are (generally usually frequently) very
2.	(?<!of)(^ the the)__TERMO__ are (an a the)
3.	(?<!in)(^ the the)__TERMO__ are (an a the)
4.	(^ the the)__TERMO__ are similar to
5.	(^forming ^constituting ^composing forming constituting composing)what we call __TERMO__
6.	(^are are)called (the a an) __TERMO__
7.	(^are are)known as __TERMO__
8.	(^ the the)__TERMO__ are made of
9.	(^ the the)__TERMO__ are made up of
10.	(^ the the)__TERMO__ are made in the
11.	(^ the the)__TERMO__ are (those the) parts of
12.	(^ the the)__TERMO__ are (part composed) of
13.	(^ the the)__TERMO__ is where
14.	(^ the the)__TERMO__ (serve function work) as
15.	(^ the the)__TERMO__ \ (. * \) \.
16.	(^ the the)__TERMO__ \ (. * \) or
17.	(^ the the)__TERMO__ \ (. * \) are
18.	###(^ the the)__TERMO__ have (traz muito ruído)
19.	(^ the the)__TERMO__ (are can be should be may be) (initiated protected secreted calculated considered spread surrounded constituted formed composed characterised characterized defined described regarded viewed used located expressed situated placed protected bathed released produced understood seen divided)
20.	(^ the the)__TERMO__ tell about
21.	(^ the the)__TERMO__ (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
22.	(^ the the)__TERMO__ (can may) w+ be (initiated protected secreted calculated considered spread surrounded constituted formed composed characterised characterized defined described regarded viewed used located expressed situated placed protected bathed released produced understood seen divided)
23.	(^a a)(simple basic common typical) definition of __TERMO__
24.	(^a a)definition of __TERMO__
25.	(^like like).* , the __TERMO__
26.	(^like like).* , __TERMO__
27.	, the __TERMO__ , (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
28.	(^ the the)__TERMO__ : the
29.	(^ the the)__TERMO__ - these are

30.	(^ the the)__TERMO__ - (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
31.	(^ the the)__TERMO__ - the
32.	(^ the the)__TERMO__ - a (class type) of
33.	(^ the the)__TERMO__ - are the
34.	(^ the the)__TERMO__ -
35.	(^ the the)__TERMO__ have been \w+ (initiated protected secreted calculated considered spread surrounded constituted formed composed characterised characterized defined described regarded viewed used located expressed situated placed protected bathed released produced understood seen divided)
36.	(^ the the)__TERMO__ are the (main major principal)
37.	(^ the the)__TERMO__ are the way in which
38.	(^ the the)__TERMO__ \ (.* \) (can may) \w+ be (defined viewed considered described regarded)
39.	(^ the the)__TERMO__ \ (.* \) , which are
40.	, the __TERMO__ \.
41.	, or __TERMO__
42.	(^as as)__TERMO__ , we understand
43.	(^we we)define as __TERMO__
44.	(^ the the)__TERMO__ are not .* , but
45.	(^ the the)__TERMO__ (present have) the \w+ (traits features characteristics structures)
46.	(^ the the)__TERMO__ , which (are prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
47.	(^are are) (initiated protected secreted calculated considered spread surrounded constituted formed composed characterised characterized defined described regarded viewed used located expressed situated placed protected bathed released produced) by __TERMO__
48.	(^there there)are (groups sets types classes categories) of __TERMO__
49.	(^another another)type of .* , the __TERMO__ ,
50.	(^in in)the __TERMO__ the .* (occur take place)
51.	(^the the)(most common commonest) __TERMO__
52.	(^the the)majority of the __TERMO__ (has have)
53.	(^ the the)__TERMO__ are responsible for
54.	(^as ^by as by)__TERMO__ we (consider understand)
55.	(^we we)use __TERMO__ to (designate describe define mean denote refer)
56.	(^we we)use the term __TERMO__ to (designate describe define mean denote refer)
57.	(^ the the)__TERMO__ is used to (designate describe define mean denote refer)
58.	(^ the the)__TERMO__ \ (.* \) , which are
59.	(^ the the)__TERMO__ , that is,
60.	(^which ^that which that)are (called named) __TERMO__

61.	(^each each)\w+ of __TERMO__ (prevents originates contains holds controls conducts takes brings produces represents defines spreads influences determines becomes presents dominates describes occurs regulates acts enables equals results causes consists corresponds leads receives derives functions depends changes)
62.	(^ the the) __TERMO__ of \w+ (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
63.	(^are are)the __TERMO__
64.	- __TERMO__ -
65.	(^when when) __TERMO__ occur
66.	(^cause cause) __TERMO__
67.	(^dominant ^main ^basic ^major dominant main basic major) (divisions parts structures elements) of __TERMO__
68.	(^by by)definition , the __TERMO__ (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
69.	(^the the)\w+ (forms composes constitutes) the __TERMO__
70.	(^ the the) __TERMO__ that (occur take place) when
71.	(^occur ^take place occur take place)by __TERMO__
72.	(^in in) __TERMO__ there (are is)
73.	(^the the) (basic main major dominant) (divisions parts structures elements) of __TERMO__
74.	(^ the the) __TERMO__ - (zone area)
75.	(^ the the) __TERMO__ (a an) (zone area)
76.	(^ the the) __TERMO__ - \w+ (called named known)
77.	(^the the).* (constitute compose form) the __TERMO__
78.	- __TERMO__ \.
79.	(^the the) __TERMO__ that are nothing more
80.	(^correspond correspond)to __TERMO__
81.	(^ the the) __TERMO__ are (transformed changed modified divided) into
82.	(^ the the) __TERMO__ that (bathe occupy absorb) all
83.	(^ the the) __TERMO__ , (formed constituted determined) by
84.	(^we we)consider __TERMO__
85.	(^ the the) __TERMO__ are , by definition
86.	(^are are)(a an) (structure element) called __TERMO__
87.	(^ the the) __TERMO__ are in charge of
88.	(^ the the) __TERMO__ of these \w+ have
89.	(^ the the) __TERMO__ (that which) are
90.	(^ the the) __TERMO__ (that also which) (prevent release originate contain hold control conduct take bring produce represent define spread influence determine become present dominate describe occur regulate act enable equal play result cause consist correspond lead receive derive function depend change)
91.	(^during during)the __TERMO__
92.	(^we we)say that the \w+ is in __TERMO__
93.	(^in in)the __TERMO__ , (a an)
94.	(^in in) __TERMO__ , (a an)

95.	(^the the)term __TERMO__ (that which)
96.	(^the the) __TERMO__ , which are
97.	(^the the) __TERMO__ are (characteristic typical common basic)
98.	(^the the) __TERMO__ of a \w+ are defined as
99.	(^the the) __TERMO__ \ (.* \) are
100.	(^the the) __TERMO__ , i\ . e\.
101.	i\ . e\ . , * __TERMO__
102.	i\ . e\ . * __TERMO__
103.	(^the the) __TERMO__ (have comprise comprehend possess contain) (many several various one two a an all some)
104.	(^areas ^locations areas locations)called __TERMO__
105.	(^forms ^variations forms variations)of __TERMO__ are called
106.	(^forms ^variations forms variations)of .* are called __TERMO__
107.	(^the the)causes of __TERMO__ are
108.	(^common common) __TERMO__ are :
109.	(^the the) __TERMO__ such as
110.	(^known known)as __TERMO__
111.	(^the the) __TERMO__ \ (the ones that .* \)
112.	(^the the) __TERMO__ require
113.	(^for for)the __TERMO__ , (a an) w+ must have
114.	(^definition definition) : __TERMO__ \ (.* \) are
115.	(^the the) __TERMO__ such as
116.	(^the the) __TERMO__ are diagnosed using
117.	(^there there)are (groups sets types classes categories) of __TERMO__
118.	(^groups ^sets ^types ^classes ^categories groups sets types classes categories)of __TERMO__
119.	(^groups ^sets ^types ^classes ^categories groups sets types classes categories)of __TERMO__ :
120.	(^the the) __TERMO__ \ (the .* \)
121.	(^the the) __TERMO__ \ (.* \) \.
122.	\ (called the __TERMO__ \)

Quadro 14 – *Padrões Suporte* utilizados pelo *Corpógrafo V3* [Língua Inglesa – Plural]

ANEXO II

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
Domínio: RISCOS NATURAIS						54.093
RN-01-PTPT	"Teoria do Risco e Geografia"	Fernando Rebelo, Professor Catedrático; Universidade de Coimbra, Portugal	<i>Riscos Naturais e Acção Antrópica</i> . 2001. Coimbra: Imprensa da Universidade de Coimbra, Portugal: 239-258	PTPT	Texto semi-científico / didáctico: publicação de especialidade (texto digitalizado - excerto)	6.006
RN-02-PTPT	"Considerações sobre riscos naturais num espaço de transição. Exercícios cartográficos numa área a Sul de Coimbra"	Lúcio Cunha, Centro de Estudos Geográficos. Faculdade de Letras da Universidade de Coimbra, e Luca Dimuccio, Bolseiro da Fundação para a Ciência e Tecnologia, Portugal.	<i>Territorium - Revista de Geografia Física Aplicada no Ordenamento do Território e Gestão de Riscos Naturais</i> : 9 (2002).	PTPT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	2.894
RN-03-PTPT	"Riscos Hidrológicos"	Fernando Rebelo, Professor Catedrático; Universidade de Coimbra, Portugal	<i>Riscos Naturais e Acção Antrópica</i> . (2001) Coimbra: Imprensa da Universidade de Coimbra, Portugal: 165-235.	PTPT	Texto semi-científico / didáctico: publicação de especialidade (texto digitalizado - excerto)	3.315
RN-04-PTPT	"Glossário - Protecção Civil da Região Autónoma dos Açores"	SRPCBA - Serviço Regional de Protecção Civil e Bombeiros dos Açores, Portugal.	Internet: < http://www.srpcba.pt/glossario.htm > (s/ data), (acedido em 2004.08.10).	PTPT	Glossário: informação de especialidade / divulgação ao público (texto em linha - excerto)	2.498
RN-05-PTPT	"Aspectos gerais sobre a bacia hidrográfica"	LNEC - Laboratório Nacional de Engenharia Civil, Portugal	Internet: < http://www.dha.lnec.pt/nti/portugues/estudos/ProjectoPlanet/bvP.htm >, (s/ data), (acedido em 2004.02.20).	PTPT	Texto técnico: informação de especialidade (texto em linha - excerto)	3.912
RN-06-PTPT	"A aplicabilidade da Geografia Física aos estados ambientais salientada através de exemplos portugueses recentes"	Fernando Rebelo, Professor Catedrático; Universidade de Coimbra, Portugal	<i>Territorium - Revista de Geografia Física Aplicada no Ordenamento do Território e Gestão de Riscos Naturais</i> : 6 (1999) - pp.61-63. (Comunicação apresentada ao VIII Colóquio Ibérico de Geografia. Lisboa.1999).	PTPT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	1.889

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
RN-07-PTPT	"Dia Internacional para a Redução dos Desastres Naturais. Nota para a Comunicação Social."	Instituto de Meteorologia, Portugal.	Internet: < http://www.meteo.pt/dia_mundial_reducao_desastres_naturais_08102003.htm >, de 2003.10.07, (acedido em 2004.02.26).	PTPT	Texto semi-técnico: divulgação ao público (texto em linha).	2.218
RN-08-PTPT	"Plano Municipal de Emergência"	Câmara Municipal do Porto - Serviço Municipal de Protecção Civil, Portugal.	Proposta de aprovação do Plano Municipal de Emergência (1999.06.23):14-28	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto digitalizado - excerto)	2.045
RN-09-PTPT	"Disciplina de Geologia Ambiental. Licenciatura em Engenharia do Ambiente. 3º Ano - 2º Semestre 2002/2003"	Óscar M. C. Ferreira e Ana Matias, Universidade do Algarve, Portugal.	Internet: < http://www.ualg.pt/fcma/dact/ficheiros/fich_OF/Geologia%20Ambiental%20S%20Oscar.doc >, (2000.Março), (acedido em 2003.11.21).	PTPT	Texto semi-científico: informação de especialidade (texto em linha)	1.185
RN-10-PTPT	"A importância da integração dos Riscos Geológicos no planeamento, no ordenamento do território e na protecção do meio ambiente: conceitos e algumas ideias"	Luis Manuel Fernandes Simões - Professor Adjunto - Director do Curso de Engenharia do Ambiente da ESTV - Escola Superior de Tecnologia de Viseu, Portugal.	Internet: < http://www.ipv.pt/millennium/ect7_lmfs.htm >, (s/ data), (acedido em 2004.08.10).	PTPT	Texto científico / didáctico: informação de especialidade (texto em linha)	3.585
RN-11-PTPT	"As catástrofes naturais e o sector segurador em 2002. Estudo da Sigma – Swiss Re"	Associação Portuguesa de Seguradores, Portugal.	Internet: < http://www.apseguradores.pt/boletim/APSNOTICIAS_JUL_SET_2003/ONLINE/Catastrofes.htm >, 2003.Julho e Setembro, (acedido em 2004.08.10)	PTPT	Texto semi-técnico: relatório / informação de especialidade / divulgação ao público (texto em linha)	761

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
RN-12-PTPT	"Documento de trabalho sobre protecção civil: maior sensibilização do público e segurança face a riscos naturais ou induzidos pelo homem."	Comissão Europeia, Direcção-Geral Ambiente, Direcção B - Qualidade do Ambiente e Recursos Naturais, ENV.B4 - Protecção Civil e Acidentes Ambientais, Bruxelas, 05.02.2003, Versão 3b, Portugal.	Internet: < http://europa.eu.int/comm/environment/civil/pr ote/pdfdocs/outline3_pt.pdf >, de 2003.02.05, (acedido em 2004.09.07)	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha)	2.355
RN-13-PTPT	"A prevenção das catástrofes naturais, o ordenamento e gestão do território no contexto do desenvolvimento sustentável."	Luís A. Mendes Victor, Professor Doutor	Comunicação: PEL, Plano Estratégico de Lagos – Fórum Lagos 2020 – 31-01-2004 / Internet: < http://www.cm-lagos.pt/PEL/Documentos/PEL%20Conf3%20M endes_Victor%20Catastrofes.doc >, de 2004.01.31, (acedido em 2004.09.07).	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha)	1.378
RN-14-PTPT	"Plano Distrital de Emergência"	Governo Civil de Santarém, Portugal	Internet: < http://www.gov-civil-santarem.pt/projecto.htm#10 >, de 2003.05.27, (acedido em 2004.09.08).	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha - excerto)	7.618
RN-15-PTPT	"Retrospectiva Anual das Catástrofes Naturais"	Associação Portuguesa de Seguradores, Portugal.	Internet: < http://www.apseguradores.pt/boletim/ONLINE 105_NOVEMBRO_2002/catastrofes.htm >, de 2002.11.05, (acedido em 2004.09.09).	PTPT	Texto semi-técnico: relatório / informação de especialidade / divulgação ao público(texto em linha)	626
RN-16-PTPT	"Plano Regional da Política do Ambiente. Opções Estratégicas e Medidas de Intervenção. Componentes e Factores Ambientais. Riscos Naturais e Induzidos pelo Homem."	DRAmb - Direcção Regional do Ambiente, Portugal	Internet: < http://www.sra.pt/dramb/plano_ambiente/opco es_estrategicas/op_files/riscosnatu.htm >, (s/data), (acedido em 2004.09.09).	PTPT	Texto técnico: relatório / informação de especialidade / divulgação ao público (texto em linha)	981

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
RN-17-PTPT	"Situações de Risco e Protecção Civil"	INAG - Instituto da Água, Portugal	Internet: < http://www.inag.pt/inag2004/port/a_intervencao/planeamento/pna/pdf_pna_v1/v1_c2_t09.pdf >, (s/data), (acedido em 2004.09.24).	PTPT	Texto semi-técnico: relatório / informação de especialidade / divulgação ao público (texto em linha)	10.827
Sub - domínio: CHEIAS						90.125
C-01-PTPT	"SNBPC - Riscos Naturais: Cheias"	SNBPC - Serviço Nacional de Bombeiros e Protecção Civil, Portugal.	Internet: < http://www.snbpc.pt/snpc2000/riscos/naturais/cheias/frames/r2_cheias.htm >, de 2000, (acedido em 2003.09.21).	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha).	4.072
C-02-PTPT	"O risco das inundações e a sua gestão. Uma visão nacional e uma visão europeia".	João S. Rocha, Eng. Civil, Investigador Coordenador, Laboratório Nacional de Engenharia Civil, Lisboa, Portugal	Internet: < http://www.aprh/congressoagua98/files/com/004.pdf >, (s/data), (acedido em 2004.08.03).	PTPT	Texto técnico: relatório / informação de especialidade (texto em linha)	6.618
C-03-PTPT	"Relatório da visita efectuada às Ribeiras de Odelouca e de Monchique em Dezembro de 1997, após as cheias de Outubro"	Ana Telhado, Engenheira do INAG - Instituto da Água, Portugal	Internet: < http://www.inag.pt/inag2004/port/a_intervencao/d_hidrico/pdf/a_impacte_amb/cheias_odelouc_a_1997.pdf >, de 1998.02.06, (acedido em 2004.08.20).	PTPT	Texto técnico: relatório / informação de especialidade (texto em linha)	3.440
C-04-PTPT	"Relatório da 2ª visita efectuada às ribeiras de Odelouca, de Monchique, do Monchicão e dos Carvalhos, após as cheias de Outubro de 1997"	INAG - Instituto da Água, Portugal	Internet: < http://www.inag.pt/inag2002/port/a_intervencao/d_hidrico/pdf/a_impacte_amb/cheias_odelouc_a_visita2.pdf >, de 2002, (acedido em 2003.09.21).	PTPT	Texto técnico: relatório / informação de especialidade (texto em linha)	1.683

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
C-05-PTPT	"Risco e crise nas inundações rápidas em espaço urbano. Alguns exemplos portugueses analisados a diferentes escalas"	Fernando Rebelo, Instituto de Estudos Geográficos. Faculdade de Letras. Universidade de Coimbra, Portugal.	<i>Territorium</i> - Revista de Geografia Física Aplicada no Ordenamento do Território e Gestão de Riscos Naturais, 4. Minerva, Coimbra.1997:29-47.	PTPT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	9.377
C-06-PTPT	"As cheias na área urbana do Porto. Risco, percepção e ajustamentos"	Edite Velhas, Instituto de Geografia. Faculdade de Letras. Universidade do Porto, Portugal.	<i>Territorium</i> - Revista de Geografia Física Aplicada no Ordenamento do Território e Gestão de Riscos Naturais, 4. Minerva, Coimbra.1997:49-62.	PTPT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	5.690
C-07-PTPT	"As cheias do Rio Tâmega. O caso da área urbana de Amarante"	António de Sousa Pedrosa, Instituto de Geografia. Faculdade de Letras da Universidade do Porto e Francisco da Silva Costa, Departamento de Geografia e Planeamento, Universidade do Minho, Portugal.	<i>Territorium</i> - Revista de Geografia Física Aplicada no Ordenamento do Território e Gestão de Riscos Naturais, 6. Minerva, Coimbra, 1999:49-60.	PTPT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	3.249
C-08-PTPT	"Inundações do Rio Douro: dados históricos e hidrológicos"	Cristina Aires e Diamantino Insua Pereira, Departamento de Ciências da Terra, Universidade do Minho & Teresa Mira Azevedo, Departamento de Geologia, Universidade de Lisboa, Portugal.	Internet: < http://www.lettras.up.pt/geograf/APEQ/p11.htm >, de 2000.Outubro, (acedido em 2003.09.17).	PTPT	Texto científico / didáctico: informação de especialidade (texto em linha)	762
C-09-PTPT	"Riscos associados com cheias fluviais"	P. Proença Cunha, Centro de Geociências da Universidade de Coimbra; Departamento de Ciências da Terra da Universidade de Coimbra, Portugal.	APG - Associação Portuguesa de Geólogos: Seminário "Riscos Geológicos", 29-30 Setembro, 2003 / Internet: < http://www.apgeologos.pt/files/Cheias%20Fluviais.pdf >, (acedido em 2004.02.17).	PTPT	Texto científico / didáctico: informação de especialidade (texto em linha)	2.568

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
C-10-PTPT	"O Rio Mondego. O Ambiente Fluvial e a sua Ecologia"	João S. Rocha, Engenheiro Civil, Investigador Coordenador, Departamento de Hidráulica e Ambiente, Laboratório Nacional de Engenharia Civil, Lisboa, Portugal.	Internet: < http://www.dha.lnec.pt/nre/portugues/funcionarios/papers/jrocha/criomon7.pdf > (s/data), (acedido em 2004.02.17).	PTPT	Texto técnico: informação de especialidade (texto em linha)	5.139
C-11-PTPT	"Contribuição das Ciências Sociais para o Estudo dos Impactes das Cheias"	Maria Luísa Lima e Luís Miguel Faisca; Laboratório Nacional de Engenharia Civil, Lisboa, Portugal.	<i>Contribuição das Ciências Sociais para o Estudo dos Impactes das Cheias</i> . 1992. Lisboa.	PTPT	Texto científico: publicação de especialidade (texto digitalizado)	19.598
C-12-PTPT	"Domínio Hídrico. Conceitos e Normas."	INAG - Instituto da Água, Portugal	Internet: < http://www.inag.pt/inag2002/port/a_intervencao/d_hidrico/pdf/dom_hidr_conceitos.pdf >, (s/data), (acedido em 2004.08.10)	PTPT	Texto técnico: informação de especialidade (texto em linha)	2.234
C-13-PTPT	"Chuvas torrenciais causaram enormes inundações na Europa" (Swiss Re Profile "Torrential rains cause major flooding across Europe" - Updated 21.08.2002)	Associação Portuguesa de Seguradores, Portugal.	Internet: < http://www.apseguradores.pt/boletim/ONLINE102_JULHOAGOSTO/VOLTA.htm >, de 2002.08.21 , (acedido em 2004.08.10).	PTPT	Texto semi-técnico: relatório / informação de especialidade / divulgação ao público (texto em linha)	1.118
C-14-PTPT	"Que faz a Europa? Tempestades e Inundações"	Portal da União Europeia	Internet: < http://europa.eu.int/comm/research/leaflets/disasters/pt/storms.html >, (s/data), (acedido em 2004.08.10)	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha).	299

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
C-15-PTPT	"As soluções de controlo na origem como estratégia de gestão integrada de águas pluviais em meio urbano."	Ana Fonseca Galvão, Eng ^a . Ambiente, IST; Pedro Dinis Valério, Eng ^o . Ambiente, IST; José Saldanha Matos, Prof. Auxiliar, IST; Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal.	Internet: < http://alfa.ist.utl.pt/~jjdd/ControloOrigem.doc >, (s/data), (acedido em 2004.09.13).	PTPT	Texto científico / didáctico: informação de especialidade (texto em linha)	3.681
C-16-PTPT	"Plano Regional da Política do Ambiente. Caracterização Base. Caracterização Geral do Estado do Ambiente. Riscos Naturais e Induzidos pelo Homem."	DRAmb - Direcção Regional do Ambiente, Portugal.	Internet: < http://www.sra.pt/dramb/plano_ambiente/prpa_caracterizacao/caract_files/riscos2.htm >, (s/data), (acedido em 2004.09.03).	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha)	2.254
C-17-PTPT	"Segurança de Barragens - Ondas de Inundação. Aplicação da Legislação Portuguesa."	Vítor Ribeiro, Engenheiro Civil (FEUP), HIDRORUMO - Projecto e Gestão, Grupo EDP, Portugal.	Internet: < http://www.aprh.pt/congressoagua98/files/com/107.pdf >, (s/data), (acedido em 2004.09.07).	PTPT	Texto científico / didáctico: informação de especialidade (texto em linha)	4.325
C-18-PTPT	"Planos de Bacia Hidrográfica do Rio Sado - Volume II – Enquadramento – Anexo I - Glossário (Rev. [nº] – data [1999-12-30])"	Direcção Regional do Ambiente e do Ordenamento do Território do Alentejo, Ministério do Ambiente e do Ordenamento do Território, Portugal	Internet: < http://www.dram-a.pt/pbh/pbhsado/fase1_volume2_anexo1.pdf >, 1999.12.30, (acedido em 2004.09.08).	PTPT	Texto técnico: informação de especialidade (texto em linha)	5.070
C-19-PTPT	"Glossário Internacional de Hidrologia"	UNESCO - United Nations Educational, Scientific and Cultural Organization, UNEP - United Nations Environmental Program	Internet: < http://www.igeo.pt/IGEO/portugues/Novidades_eventos/Dia%20do%20Ambiente%202003/Biblionline.htm >, (s/data), (acedido em 2004.09.30).	PTPT	Glossário: informação de especialidade / divulgação ao público (texto em linha - excerto)	2.216

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
C-20-PTPT	"Qual o Grau de Excepcionalidade das Cheias ocorridas no Início do Ano Hidrológico de 1997/98?"	SNIRH - Sistema Nacional de Informação de Recursos Hídricos. Rui Rodrigues, Chefe da Divisão de Recursos Superficiais, INAG; Cláudia Brandão e Teresa Álvares, Técnicas Superiores da Direcção de Serviços de Recursos Hídricos, INAG, Portugal.	Documento elaborado para apresentação no 4º Congresso da Água. Lisboa , Março de 1998 / Internet: < http://snirh.inag.pt/snirh/download/cheias97_98.pdf >, (acedido em 2004.09.13).	PTPT	Texto científico: informação de especialidade (texto em linha)	2.634
C-21-PTPT	"Principais actividades do LNEC em áreas relevantes para a segurança. Segurança contra inundações."	Laboratório Nacional de Engenharia Civil, Portugal.	Internet: < http://www-ext.lnec.pt/LNEC/news_imagens/inundacoesInternat.PDF >, (s/data), (acedido em 2004.09.13).	PTPT	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha).	310
C-22-PTPT	"As cheias e a Gestão de Bacias Hidrográficas"	Rodrigo Maia, Professor Auxiliar; A. Álvares Ribeiro, Professor Catedrático; Faculdade de Engenharia da Universidade do Porto, Portugal	APRH - Associação Portuguesa de Recursos Hídricos: Internet: < http://www.aprh.pt/congressoagua98/files/com/093.pdf >, (s/data), (acedido em 2004.08.03)	PTPT	Texto científico / didáctico: informação de especialidade / divulgação ao público (texto em linha)	3.788
						144218

Quadro 15 – Descrição do corpus em língua portuguesa

LEGENDA:

C = Cheias

PTPT = Português de Portugal

RN = Riscos Naturais

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
Domínio:	NATURAL	HAZARDS				56.744
NH-01-ENUSA	"Definitions: Natural Hazards and Disasters"	Graham A. Tobin e Burrell E. Montz	<i>Natural Hazards. Explanation and Integration.</i> 1997. New York: The Guilford Press: 5-7.	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	773
NH-02-ENUSA	"What a Risk?"	Graham A. Tobin e Burrell E. Montz	<i>Natural Hazards. Explanation and Integration.</i> 1997. New York: The Guilford Press: 281-283.	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	798
NH-03-ENUSA	"Hazards: Risk, Impact, Vulnerability "	Graham A. Tobin e Burrell E. Montz	<i>Natural Hazards. Explanation and Integration.</i> 1997. New York: The Guilford Press: 29-33.	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	1.441
NH-04-ENUSA	"The components of Natural Hazard"	Graham A. Tobin e Burrell E. Montz	<i>Natural Hazards. Explanation and Integration.</i> 1997: New York: The Guilford Press: 7-29	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	6.728
NH-05-ENUSA	"Risk Management"	Graham A. Tobin e Burrell E. Montz	<i>Natural Hazards. Explanation and Integration.</i> 1997. New York: The Guilford Press: 306-308	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	405
NH-06-ENINT	"Terminolog y: basic terms of disaster risk reduction"	UN - United Nations / ISDR - International Strategy for Disaster Reduction	Internet: < http://www.unisdr.org/unisdr/Annex%201%20Terminology.pdf > (s/data), (acedido em 2003.07.12).	ENINT	Glossário: informação de especialidade / divulgação ao público (texto em linha - excerto)	2.619
NH-07-ENUSA	"Classificatio n of waters"	Frank. J. Trelease, Faculdade de Direito, Wyoming University, U.S.A.	<i>Handbook of Applied Hydrology.</i> 1964. USA: McGraw-Hill, Inc.: 25:2-5	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	2.421

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
NH-08-ENCAN	"Risk Assessment, Emergency Preparedness and Response to Hazards: The Case of the 1997 Red River Valley Flood, Canada"	C. Emdad Haque, Department of Geography, Brandon University, Brandon, Manitoba, Canada	<i>Natural Hazards</i> . 2000. New York: The Guilford Press. Volume 21: 225-245.	ENCAN	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	1.394
NH-09-ENINT	"World Map of Natural Hazards - A Global View of the Distribution and Intensity of Significant Exposures"	G. Berz, W. Kron, T. Loster, E. Rauch, J. Schimetschek, J. Schmieder, A. Siebert, A. Smolka and A. Wirtz, Munich Reinsurance Company, Munich, Germany	<i>Natural Hazards</i> . 2001. New York: The Guilford Press. Volume 23: 443-465.	ENINT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	8.985
NH-10-ENUSA	"The cycling of water on the continents"	Alan H. Strahler, Bonston University and Arthur N. Strahler, Columbia University (Emeritus), U.S.A.	<i>Physical Geography: science and systems of the human environment</i> . 2001. New York, London, Sydney. John Wiley & Sons: Chapter 16	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	9.489
NH-11-ENUSA	"Fluvial Processes and Landforms"	Alan H. Strahler, Bonston University and Arthur N. Strahler, Columbia University (Emeritus), U.S.A.	<i>Physical Geography: science and systems of the human environment</i> . 2001. New York, London, Sydney: John Wiley & Sons: Chapter 17	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	8.055

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
NH-12-ENUSA	"What are Natural Hazards?"	Bill MacGuire, Benfield Greig Hazard Research Centre; Ian Mason, Department of Space and Climate Physics; Christopher Kilburn, Benfield Greig Hazard Research Centre; University College London, United Kingdom.	<i>Natural Hazards and Environmental Change</i> . 2002. New York: Oxford University Press: Chapter 1: 1-2.	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	532
NH-13-ENUSA	"A natural hazards primer"	Bill MacGuire, Benfield Greig Hazard Research Centre; Ian Mason, Department of Space and Climate Physics; Christopher Kilburn, Benfield Greig Hazard Research Centre; University College London, United Kingdom.	<i>Natural Hazards and Environmental Change</i> . 2002. New York: Oxford University Press. Chapter 1: 10-14.	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	2.870
NH-14-ENUK	"Hazard in the environment"	Keith Smith, Emeritus Professor of Environmental Science, University of Stirling, Stirling, United Kingdom	<i>Environmental Hazards. Assessing Risk and Reducing Disaster</i> . 2001. 3 rd edition. New York: Routledge. Part I - Chapter 1: 6-25.	ENUK	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	5.619
NH-15-ENUSA	"Social Aspects of Disasters"	Charles H. V. Erbert, Professor of Physical Geography, University at Buffalo, State University of New York, U.S.A.	<i>Disasters. An Analysis of Natural and Human-Induced Hazards</i> . 2000. 4 th Edition. Dubuque, Iowa: Kendall/Hunt Publishing Company. Chapter 16: 199-204	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	3.615
NH-16-ENUSA	"Expanding the Horizons of Disaster Research."	Russell Dynes, Professor Emeritus, Disaster Research Center, University of Delaware, U.S.A.	<i>Natural Hazards Observer</i> , Vol. XXVIII No. 4 March 2004, Internet: < http://www.colorado.edu/hazards/o/mar04/mar04a.html >, (acedido em 2004.09.30)	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	1.000

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
Sub-	domínio:	FLOODS				86.369
F-01-ENUSA	"Floods - What Is A Flood?"	FEMA - Federal Emergency Management Agency	Internet: < http://www.fema.gov/hazards/floods/ >, de 2003.06.06 (acedido em 2003.07.12.	ENUSA	Texto semi-técnico: publicação de especialidade / divulgação ao público (texto em linha - excerto)	349
F-02-ENINT	"Types of Floods"	O. N. Dhar e S. Nandargi, Instituto Indiano de Meteorologia Tropical, Pune, India	<i>Natural Hazards</i> . 2003. New York: The Guilfor Press. Volume 28, nº 1:10-13.	ENINT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	1.273
F-03-ENINT	"Floods and Vulnerability: Need to Rethink Flood Management"	A. Dixit, Fundação para a Conservação da Água do Nepal, Kathmandu, Nepal	<i>Natural Hazards</i> . 2003. New York: The Guilfor Press. Volume 28, nº 1:155-177.	ENINT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	9.702
F-04-ENUSA	"Hydrological events: Floods"	Graham A. Tobin e Burrell E. Montz	<i>Natural Hazards. Explanation and Integration</i> . 1997. New York: The Guilfor Press. 62-68.	ENUSA	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	1.451
F-05-EN AUS	"Community safety - Floods"	Australian Government, Attorney-General Department, EMA - Emergency Management Australia	Internet: < http://ema.gov.au/ema/emaInternet.nsf/AllDocs/ RW10A4B850AA8E5BD1CA256C5A0020745A ?OpenDocument >, de 2003.05.30 (acedido em 2004.02.17).	EN AUS	Texto semi-técnico: publicação de especialidade / divulgação ao público (texto em linha)	1.583
F-06-ENUSA	"Basics of flooding and flood protection"	FMA - Floodplain Management Association	Internet: < http://floodplain.org/flood_basics.htm >, de 1992.Julho, (acedido em 2004.02.17).	ENUSA	Texto semi-técnico: publicação de especialidade / divulgação ao público (texto em linha)	565

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
F-07-ENUSA	"A Preparedness Guide"	U.S. Department of Commerce. National Oceanic and Atmospheric Administration National Weather Service - July 1992. NOAA, FEMA, and The American Red Cross.	Internet: < http://www.nws.noaa.gov/om/brochures/ffbro.htm > (s/ data), (acedido em 2004.02.17).	ENUSA	Texto semi-técnico: publicação de especialidade / divulgação ao público (texto em linha)	2.162
F-08-ENUSA	"Floods and flood plains"	USGS - United States Geological Survey, USA.	Internet: < http://water.usgs.gov/pubs/FS/OFR93-641/ >, de 1993, (acedido em 2004.02.17).	ENUSA	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha - excerto)	761
F-09-ENINT	"Regional Cooperation in Flood Management in the Ganges-Brahmaputra-Meghna Region: Bangladesh Perspective"	Q.K. Ahmad e A. U. Ahmed, Bangladesh Unnayan Paridash, New Market, Dahka, Bangladesh	<i>Natural Hazards</i> . 2003. New York: The Guilford Press. Volume 28, nº 1:181-198.	ENINT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	5.469
F-10-ENUSA	"Significant Floods in the United States During the 20th Century - USGS Measures a Century of Floods"	Charles A. Perry, U.S. Geological Survey Water Resources in Kansas, USA	Internet: < http://ks.water.usgs.gov/Kansas/pubs/fact-sheets/fs.024-00.html >, de 2000.Março, (acedido em 2004.02.18).	ENUSA	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha - excerto)	1.523

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
F-11-ENUSA	"Critical characteristics of flood hazard"	Gilbert F. White, Chicago University, USA	<i>Handbook of Applied Hydrology</i> . USA: McGraw-Hill, Inc. 1964 / Secção 25-V:118-119	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	671
F-12-ENUSA	"Measurable features of floods"	Tate Dalrymple, U.S. Geological Survey, USA	<i>Handbook of Applied Hydrology</i> . USA: McGraw-Hill, Inc. 1964 / Secção 25-I:2-5	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	1.759
F-13-ENUK	"8307 Environmental Geology - Flooding in Asia"	Oxford Brooks University, United Kingdom	Internet: < http://www.brokes.ac.uk/geology/8307/floods2.html >, baseado em artigos da publicação especializada New Scientist, (13 April 1991:34-41; 28 March 1992:17), (acedido em 2004.02.18).	ENEN	Texto semi-científico / didático: publicação de especialidade (texto digitalizado - excerto)	992
F-14-ENINT	"The Disastrous Flood of 1998 and Long Term Mitigation Strategies for Dhaka City"	I. M. Faisal, Dep. Estudos Ambientais, Universidade Norte Sul; M.R. Kabir, Dep. Engenharia de Recursos Hídricos, Universidade de Engenharia e Tecnologia; A. Nishat, IUCN; Dhaka, Bangladesh	<i>Natural Hazards</i> . 2003. New York: The Guilford Press. Volume 28, nº 1: 85-99	ENINT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	2.148
F-15-ENCAN	"Three Recent Extreme Floods in Bangladesh: A Hydro-Meteorological Analysis"	M.M.Q.Mirza, Instituto para os Estudos Ambientais, Universidade de Toronto, Canada	<i>Natural Hazards</i> . 2003. New York: The Guilford Press. Volume 28, nº 1:35-64	ENCAN	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	2.628

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
F-16-ENCAN	"The 1996 flooding in the Saguenay Valley, Quebec – lessons from the disaster"	Gregory R. Brooks, Geological Survey of Canada	Internet: < http://www.cseg.ca/conferences/2000/296.PDF >, 2000, (acedido em 2004.02.18).	ENCAN	Texto semi-técnico: informação de especialidade / divulgação ao público (texto em linha)	422
F-17-ENINT	"Flood Events in the Rhine Basin: Genesis, Influences and Mitigation"	Markus Disse e Heinz Engel, German Federal Institute of Hydrology, Koblenz, Germany	<i>Natural Hazards</i> . 2001. Volume 23, ners 2-3: 271-290.	ENINT	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	4.962
F-18-ENAU	"Flooding in Australia: a review of events in 1998"	Stephen W. Yeo, Department of Physical Geography, Macquaire University, Australia	<i>Natural Hazards</i> . 2002. New York: The Guilfor Press. Volume 25: 177-191	ENAU	Texto científico: publicação periódica de especialidade (texto digitalizado - excerto)	4.820
F-19-ENUSA	"Glossary"	Alan H. Strahler e Arthur N. Strahler	<i>Physical Geography: science and systems of the human environment</i> . 2002. New York, London, Sydney. John Wiley & Sons: Glossary.	ENUSA	Glossário / didático: informação de especialidade (texto em linha - excerto)	2.729
F-20-ENINT	"International Glossary of Hydrology"	UNESCO - United Nations Educational, Scientific and Cultural Organization, UNEP - United Nations Environmental Program	Internet: < http://www.cig.ensmp.fr/~hubert/glu/HINDEN.HTM >, (s/ data), (acedido em 2004.09.30).	ENINT	Glossário: informação de especialidade / divulgação ao público (texto em linha - excerto)	2.017
F-21-ENUSA	"Glossary of Flooding Terms"	FCD - Flood Control District of Maricopa County	Internet: < http://www.fcd.maricopa.gov/Flooding/Glossary.asp > (acedido em 2004.03.21)	ENUSA	Glossário: informação de especialidade / divulgação ao público (texto em linha - excerto)	2.639

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
F-22-ENINT	"Flood Control in Tokyo, Osaka and Nagoya"	Masahiko Oya, Emeritus Professor of Waseda University, and Technical Advisor of Nikken Consultants Inc., Tokyo, Japan	<i>Applied Geomorphology for Mitigation of Natural Hazards</i> . 2001. Dordrecht / Boston / London: Kluwer Academic Publishers. Volume 15: 111-119.	ENINT	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	1.751
F-23-ENUK	"Hydrological Hazards. Floods. Flood Hazards."	Keith Smith, Emeritus Professor of Environmental Science, University of Stirling, Stirling, United Kingdom	<i>Environmental Hazards. Assessing Risk and Reducing Disaster</i> . 3 rd edition. New York: Routledge. 2001: Part I - Chapter 11: 259-274/278-288.	ENUK	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	8.262
F-24-ENUSA	"Floods and other weather related hazards in a changing climate"	Bill MacGuire, Benfield Greig Hazard Research Centre; Ian Mason, Department of Space and Climate Physics; Christopher Kilburn, Benfield Greig Hazard Research Centre; University College London, London, United Kingdom.	<i>Natural Hazards and Environmental Change</i> . 2002. New York: Oxford University Press. Chapter 3: 53-60	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	2.385
F-25-ENUSA	"River and Urban Floods"	Charles H. V. Erbert, Professor of Physical Geography, University at Buffalo, State University of New York, USA	<i>Disasters. An Analysis of Natural and Human-Induced Hazards</i> . 2000. 4 th Edition. Dubuque, Iowa: Kendall/Hunt Publishing Company. Chapter 5: 53-61	ENUSA	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	4.290
F-26-ENUK	"Introduction to Floods and Flood Management"	Dennis J. Parker	<i>Floods</i> . 2000. Edited by D.J. Parker. London: Routledge Hazards and Disasters Series. Volume I. Part I: 22-36	ENUK	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	6.566

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
F-27-ENCAN	"Effects of Flow Regulation on Hydrologic patterns of a large inland delta."	Robert Leconte, Department of Construction Engineering, École de Technologie Supérieure, Montréal, Québec, Canada; Alain Pietroniro and Terry D. Prowse, Aquatic Ecosystems Impact Branch, National Water Research Institute, Saskatoon, Saskatchewan, Canada; Daniel L. Peters, Aquatic Ecosystems Impact Branch, National Water Research Institute, Saskatoon, Saskatchewan, Canada and Watershed Ecosystem Graduate Program, Trent University, Peterborough, Ontario, Canada	<i>Regulated Rivers Research & Management</i> . January-February 2001. Chichester: John Wiley & Sons, Ltd. Volume 17 Number 1: 51-65	ENCAN	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	4.583
F-28-ENINT	"Effects of flood interruption on species richness, diversity and floristic composition of woody regeneration in the uper Rhine alluvial hardwood forest."	Anne-Frédérique Deiller, Jean-Michel N. Walter and Michele Trémolieres, Centre d'Eludes el de Recherches Eco-Géographiques (CEREG, CNRS-ULP), Institut de Botanique, Strasbourg, France	<i>Regulated Rivers Research & Management</i> . July-October 2001. Chichester: John Wiley & Sons, Ltd. Volume 17 Number 4-5: 393-405,	ENINT	Texto científico / didático: publicação de especialidade (texto digitalizado - excerto)	5.220

CÓDIGO	TÍTULO	AUTORIA	FONTE	LÍNGUA E VARIANTE	DESCRIÇÃO DO TEXTO	Nº PALAVRAS
F-29-ENCAN	"Natural Hazards in Canada. Atmospheric, Hydrologic and Geophysical Hazards. Hydrologic Hazards. Floods."	Brun, S.E., Etkin, D., Law, D.G., Wallace, L., and White, R.	<i>Coping with Natural Hazards in Canada: Scientific, Government and Insurance Industry Perspectives.</i> (1997). Internet: < http://www.utoronto.ca/env/nh/pt2ch2-3-2.htm >, (acedido em 2004.06.17)	ENCAN	Texto semi-técnico: publicação de especialidade / divulgação ao público (texto em linha)	1.048
F-30-ENUSA	"Flood Control"	Microsoft ® Encarta ® Online Encyclopedia 2003	Internet: < http://encarta.msn.com > © 1997-2003 Microsoft Corporation, (acedido em 2004.06.17).	ENUSA	Texto semi-científico / didáctico: publicação de especialidade (texto em linha)	1.639
						143.113

Quadro 16 - Descrição do *corpus* em língua portuguesa

LEGENDA:

ENAU = English - Austrália

ENCAN = English - Canada

ENINT = English - International

ENUK = English - United Kingdom

ENUSA = English - United States of America

F = Floods

NH = Natural Hazards