

"O papel dos corpora especializados na criação de bases terminológicas"

Maria Fernanda Bacelar do Nascimento

Centro de Linguística da Universidade de Lisboa

O enquadramento em que actualmente ocorre a investigação e o desenvolvimento científico que têm como objecto o estudo das *línguas para fins específicos* (também chamadas *sublínguas*, *línguas especializadas* ou *línguas científicas*) assenta em princípios teóricos e metodológicos que põem em evidência a importância do uso de *corpora* linguísticos nesta área; assim, ganham incremento a construção destes recursos linguísticos – quer gerais, quer especializados – e a construção de ferramentas informáticas adequadas à extracção de conhecimentos relevantes de carácter terminológico.

O valor destes recursos linguísticos e informáticos tem sido destacado na identificação e caracterização das unidades lexicais que, em função do seu uso em determinados contextos e situações, detêm o estatuto de termo ou de não-termo, pois "el carácter de término no se da *per se*, sino en función del uso de una unidad léxica en un contexto expresivo y situacional determinado." CABRÉ (1999: 124). Daqui decorre, naturalmente, também, a importância que se atribui aos parâmetros a estabelecer no desenho dos diversos *corpora* especializados e à avaliação dos *tipos* de textos que os constituem, *géneros* e *domínios* do conhecimento que abrangem.

Como se tem provado na prática real e está amplamente documentado na literatura da especialidade (por exemplo, em CABRÉ, 1998 e 1999, PEARSON, 1998 ou MEYER e MACKINTOSH, 1996), os dados empíricos não confirmam, antes claramente infirmam, uma pretendida homogeneidade e universalidade de denominações no âmbito dos conhecimentos especializados assim como a adequação do processo de unificação metodológica estabelecido pelas normas internacionais; estas, limitando o seu campo de trabalho à normalização dos termos e dos conceitos denominados, consideram a existência de um só registo, não atendendo à dimensão comunicativa dos termos que, no uso real, tornam patente a existência de diversos graus de variação directamente relacionados com os contextos discursivos em que se inserem.

Para dar conta da pluridade dos usos reais nesta matéria tão complexa foi determinante a revisão dos modelos mais reducionistas da Teoria Geral da Terminologia (TGT) que não descrevem satisfatoriamente a complexidade dos léxicos especializados.

Os princípios da Teoria Comunicativa da Terminologia (TCT) vão, precisamente, no sentido da revisão desses modelos da TGT à luz dos dados empíricos. CABRÉ (1999) identifica claramente esses princípios e aponta como pontos fracos da TGT o facto de esta não considerar que os *termos* pertencem plenamente – e não só nos seus aspectos denominativos – à língua natural com todas as consequências que isso implica. Outros dos pontos fracos que a autora aponta à TGT são o facto de esta teoria considerar irrelevante a análise sintáctica dos termos nos contextos em que ocorrem e, ainda, a defesa que a TGT faz da univocidade e monossemia dos termos, conceitos que são claramente contrariadas pela observação dos dados reais.

Considerando que, em terminologia, a investigação empírica se encontra no início, e que existem muito poucos dados sobre a frequência dos *termos* em textos especializados, CABRÉ (1999) afirma que todo o trabalho terminológico é, pelo menos inicialmente, descritivo e que fazer terminologia significa descrever os *termos* usados efectivamente na comunicação especializada.

Assim, só depois de realizado o trabalho descritivo se podem conceber propostas de carácter mais prescritivo no sentido de orientar o uso.

Nesta abordagem comunicativa, pretende-se dar conta dos *termos* como "unidades singulares y a la vez similares a otras unidades de comunicación, dentro de un esquema global de representación de la realidad, admitiendo la variación conceptual y denominativa, y teniendo en cuenta la dimensión textual y discursiva de los términos" CABRÉ (1999: 120).

Considerando o carácter polissémico das unidades lexicais, a sinonímia na comunicação especializada é tida como um facto real. Assim, a TCT prevê processos de *banalização* de unidades especializadas, de *terminologização* de unidades gerais e de *pluriterminologização*: passagem das unidades de um âmbito para outro ou outros, com a consequente aquisição de traços associados a esse ou a esses âmbitos.

Enfim, concebida 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*, esta última dará conta das unidades terminológicas. Estas formam parte integrante da língua natural e participam das suas características; na comunicação, activa-se e singulariza-se o carácter terminológico dos itens lexicais em função do seu uso em situações e contextos adequados. Na TCT, as unidades terminológicas são, pois, encaradas como parte da língua natural e da gramática que descreve cada língua.

A utilização de recursos linguísticos – nomeadamente *corpora* e léxicos – na criação de bases terminológicas encontra fundamento nestas propostas teóricas. Deter-me-ei num tema de importância crucial para o sucesso dos trabalhos terminológicos – o desenho de *corpora* especializados, com referências também aos *corpora* gerais, indispensáveis para controlar e aferir fenómenos próprios das *línguas para fins específicos*.

O desenho de um corpus especializado

A qualidade de um projecto terminológico ou terminográfico está directamente relacionada com a qualidade do *corpus* especializado em que se baseia. Essa qualidade adquire aqui uma importância ainda maior do que num *corpus* geral. De facto, um *corpus* geral, pela enorme quantidade e diversidade dos documentos que inclui, atinge facilmente dimensões tão grandes¹ que pequenos desequilíbrios na qualidade e representatividade dos textos que o constituem acabam por ser neutralizados pela quantidade dos dados, não afectando a qualidade dos resultados.

Em contrapartida, dificilmente se consegue constituir um *corpus* especializado de dimensões muito extensas pelo que, na constituição destes *corpora* – que, em geral, não ultrapassam 1 milhão de palavras – se torna imprescindível uma grande exigência no que respeita à qualidade e à representatividade dos textos seleccionados. Só assim se poderá assegurar que sejam fiáveis os resultados das análises e a frequência de ocorrência dos fenómenos linguísticos prototípicos da língua em estudo. Além disso, o objectivo que mais frequentemente se tem em vista quando se constitui um *corpus* especializado é a sua utilização em terminologia e em terminografia. Ora a capacidade de introspecção de terminólogos e terminógrafos face à língua que descrevem é geralmente muito reduzida; por isso, ainda mais do que noutras áreas, o sucesso dos seus trabalhos depende, em grande medida, do *corpus* em que se baseiam.

Sendo, obviamente, necessário que se constituam equipas mistas de linguistas e de profissionais das áreas abrangidas pelos *corpora* especializados, não se pretende que estes últimos designem os termos a propor nem se espera que forneçam descrições linguisticamente consistentes dos seus usos e variações. A participação dos peritos é, no entanto, fundamental para a validação dos resultados finais (principalmente no que se refere à relação entre os termos e os conceitos referidos) e é absolutamente insubstituível quando se trata de garantir a qualidade das fontes de informação documentais.

Numa síntese dos aspectos mais gerais a considerar no desenho e constituição de um *corpus* especializado, é de referir que, contrariamente a um *corpus* de referência que por definição contém apenas amostragens dos documentos que o constituem, aquele deve ser, de preferência, um *corpus* textual, isto é, deve conter documentos completos para que não se perca toda a riqueza de termos e de conhecimentos, relativos aos domínios científicos ou técnicos, que os textos especializados maximamente contêm; deve ser, também, um *corpus* aberto, em constante actualização, ao qual se vão acrescentando os textos mais recentemente produzidos e do qual se vão retirando textos obsoletos, permitindo com isso acompanhar as rápidas alterações que se registam nas terminologias pertencentes à maior parte dos domínios científicos e técnicos.

É sabido que o desenho de um *corpus* geral deve contemplar uma grande diversidade de áreas do conhecimento, de temas, géneros e tipos de textos. A inclusão de documentos muito variados, em *corpora* de grandes dimensões, permite identificar e caracterizar – especialmente nas altas e médias frequências de ocorrência – fenómenos que são comuns a todos eles; estes fenómenos recorrentes num *corpus* assim constituído serão característicos de um "uso médio" representativo da generalidade dos usos. Porém, como já foi dito, também o discurso científico e técnico está longe de ser homogéneo. "Un simples análisis de la comunicación especializada real en situaciones profesionales de distinto signo muestra una multiplicidad importante de registros, en los que, sin abandonar el carácter especializado del conocimiento y su transmisión, se ponen de manifiesto una serie de características que coinciden con las que muestran otras unidades utilizadas en otros tipos de situación comunicativa" CABRÉ (1999: 118).

Assim, tal como os *corpora* gerais, também os *corpora* especializados devem contemplar as variedades do discurso que representam, sendo que uma identificação adequada dos termos científicos e técnicos depende crucialmente da classificação criteriosa dos textos a seleccionar.

De facto, em *corpora* especializados, o estabelecimento de critérios externos de selecção dos textos deve supor a inclusão de textos representativos de variados graus de tecnicidade, incluindo textos orais recolhidos em aulas, programas radiofónicos ou televisivos de divulgação e conferências dirigidas a especialistas assim como textos escritos recolhidos em manuais, artigos publicados em jornais ou revistas de grande divulgação, e artigos e livros altamente especializados. Deve ser sempre assegurado o aval de especialistas à qualidade dos autores de todos estes textos.

Assim, uma classificação dos textos no que respeita ao *género* pode ser feita através de critérios externos (informações referentes ao emissor, ao seu público, às intenções comunicativas e a outros aspectos socioculturais e profissionais) mediante, até, a utilização de hierarquias de atributos definidas pelas normas internacionais do TEI (TEXT ENCODING INITIATIVE).

Contudo, como afirma SINCLAIR (1996), não há nesta classificação, nenhuma alínea de relevância linguística que oriente o utilizador do *corpus* na procura de componentes internas adequadas, ou seja, de características linguísticas diferenciadoras que contribuam de maneira satisfatória para a definição de uma tipologia textual. Ainda segundo SINCLAIR (1996) uma tipologia adequada consistiria numa combinação equilibrada de critérios externos e internos que se reflectissem mutuamente.

Não é fácil definir *a priori* critérios internos para a constituição de *corpora*, por isso, na generalidade dos casos, eles são compilados na base de critérios não-linguísticos. Contudo, várias experiências têm sido feitas no intuito de se classificarem os textos de acordo com critérios linguísticos, ou seja, internos. Situam-se nesta linha HALLIDAY e MARTIN (1993) que propõem a identificação dos tipos de linguagem através da especificação de critérios léxico-gramaticais. Associam, por exemplo, o texto científico a passivas e a nominalizações.

BIBER (1989) propõe uma tipologia de textos baseada numa análise linguística e estatística das associações lexicais e léxico-gramaticais que aqui designaremos genericamente por

combinatórias; a este propósito afirma SINCLAIR "a combinatória é ao mesmo tempo linguisticamente poderosa e fácil de identificar através de processos informáticos" considerando, ainda, que "o agrupamento de co-ocorrentes ajuda a desambiguar as palavras individuais, e possibilita uma identificação mais exacta do tema do texto do que a que é disponibilizada por simples palavras-chave. Mais recentemente, a combinatória tem sido utilizada para classificar géneros, revelando que a mesma palavra está caracteristicamente associada a certos co-ocorrentes em determinados tipos de língua escrita e falada" SINCLAIR (1996: 80).

Uma metodologia semelhante, de base quantitativa, assente nas características linguísticas do texto, fora também já proposta por PHILLIPS, em 1983, e consiste, precisamente, na identificação de associações entre *lemas*, em textos científicos, com o objectivo de se estabelecer a macroestrutura dos textos definida através da análise de padrões combinatórios.

Parece, então, adequado que, na construção de *corpora* especializados, a uma pré-selecção de textos feita de acordo com critérios externos se siga a extracção de conhecimentos de natureza lexical ou léxico-gramatical, a qual contribui para validar a selecção definitiva dos documentos, num processo cíclico de constituição de bases fiáveis para reconhecimento das unidades terminológicas.

Os resultados a que têm chegado autores de experiências assim concebidas demonstram as potencialidades dos *corpora* especializados em estudos terminológicos, sendo a identificação, análise e interpretação das combinatórias um instrumento poderoso a utilizar nas descrições prévias ao estabelecimento de terminologias. Com este objectivo também têm sido feitas algumas experiências em Portugal, utilizando o *Corpus* de Referência do Português Contemporâneo (CRPC) ² e as ferramentas de extracção de léxicos e de combinatórias ³ que lhe estão associadas.

A título meramente exemplificativo, e sem usar dados estatísticos mas apenas informação de frequências, apresento resultados comparativos das ocorrências do verbo *deduzir* observadas em vários sub-*corpora* de dimensões comparáveis, extraídos do CRPC. A dimensão de cada um destes sub-*corpora* é de, aproximadamente, 1 milhão de palavras e a breve análise realizada teve em vista, apenas, observar frequências de um lema e respectivas formas flexionadas e também frequências dos seus usos sintáctico-semânticos e lexicais; neste sentido, são apresentados quadros comparativos das ocorrências do verbo *deduzir* em diferentes contextos discursivos.

Quadro I

Corpora e respectivas dimensões em número de palavras

<i>CORPORA</i>	ORAL	JORNALÍSTICO (JORN.)	LITERÁRIO (LIT.)	DIDÁCTICO (DID.)	ECONÓMICO (ECON.)	JURÍDICO (JUR.)
Nº de Palavras	917.921	1.031.752	1.033.457	1.021.390	1.030.602	1.029.747

As frequências lexicais

As frequências lexicais podem ser determinantes na caracterização de tipos de textos. Como diz BAAYEN (1998: 90) "Word frequencies constitute a rich source of information on the relation between lexis and text type. General summary measures capture global properties such as repetitiveness and vocabulary richness, but much more detailed information is available if we allow ourselves to consider individual words in specific frequency ranges". Na realidade, a especificidade dos textos e dos temas é evidenciada pela observação de palavras que tipicamente

ocorrem em textos homogêneos. O verbo *deduzir* ocorre com as seguintes frequências nos *corpora* em estudo:

Quadro II

Frequências globais do v. *deduzir* nos *corpora* analisados

CORPORA	ORAL	JORN.	LIT.	DID.	ECON.	JUR.
Frequência de <i>deduzir</i>	6	9	7	57	40	242

Quadro III

Distribuição das formas do v. *deduzir* pelos *corpora* analisados

Formas do v. <i>deduzir</i>	Distribuição por <i>Corpus</i>					
	Oral	Jorn.	Lit.	Did.	Econ.	Jur.
deduz		2		8	1	9
deduz-se	1	2	1	3		1
deduza						1
deduzem				2		1
deduzi		1	1			
deduzia			1			
deduziam	1					
deduzida		2		6	2	48
deduzidas				1	3	11
deduzido	1		1	1	6	56
deduzidos					8	
deduzimos				4		12
deduzindo				2	5	6
deduzir	1	2	1	21	13	35
deduzir-se					1	1
deduzirá						1
deduziram						9
deduzirem	1					
deduziria				1		
deduzirmos				2		
deduziste				1		
deduziu			1	5	1	50
deduziu-se						1
deduzo	1		1			
Totais	6	9	7	57	40	242

Como se pode observar, a frequência de *deduzir* no *corpus* Jurídico é muitíssimo superior às frequências deste verbo nos outros *corpora*. Vale a pena mencionar que num *corpus* geral extraído do CRPC, o CORLEX, com 16 milhões de palavras, o verbo *deduzir* ocorre 274 vezes,

ou seja, pouco mais vezes do que neste *corpus* jurídico de 1.029.747 palavras. Também é de esclarecer que do CORLEX consta um pequeno sub-*corpus* jurídico.

Os contextos sintáctico-semânticos e lexicais de uso

Combinando o estudo da interacção entre o verbo *deduzir* e, por um lado, as construções gramaticais e acepções a que está associado nos diferentes *corpora* e, por outro lado, as suas combinatórias, caminha-se para a identificação deste verbo como termo do discurso jurídico para a qual apontava já a frequência de uso no *corpus* correspondente. As diferentes construções sintácticas de *deduzir* estão relacionadas com diferentes sentidos deste verbo, facto que corrobora a conclusão de SINCLAIR (1991: 65) "It seems that there is a strong tendency for sense and syntax to be associated".

Nos *corpora* mencionados, observam-se três estruturas sintácticas que nos quadros IV, V e VI são representadas com o verbo seguido dos elementos à sua direita. A estas estruturas sintácticas correspondem três acepções do verbo, simplifcadamente identificadas pelos sinónimos: *concluir*, *subtrair* e *apresentar*.

Quadro IV

Deduzir \cong Concluir: estrutura sintáctica e frequências de ocorrência nos *corpora*

Estrutura sintáctica		ORAL	JORN.	LIT.	DID.	ECON.	JUR.
V	N F (de, por, a partir de,... N) Que F	4	7	5	50	10	20

Exemplos:

deduz-se das suas palavras que também as potencialidades... (ORAL)

*Nestas circunstâncias, é legítimo **deduzir** que não se trata de um projecto consensual.* (JORN.)

*Querem evitar a tropa, **deduziu** ele* (LIT.)

*As conclusões aí **deduzidas** não são aplicáveis imediatamente* (DID.)

*esse saldo, depois de **deduzido** este encargo, determina aquilo que o subscritor re* (ECON.)

***deduz-se**, com suficiente clareza, que a indemnização devida...* (JUR.)

Quadro V

Deduzir \cong Subtrair: estrutura sintáctica e frequências de ocorrência nos *corpora*

Estrutura sintáctica		ORAL	JORN.	LIT.	DID.	ECON.	JUR.
V	N de, a, em N	2	0	1	7	29	23

Exemplos:

*se puderem chegar a casa com mais alguns dólares para **deduzirem** daquela despesa* (ORAL)

Soma-se depois a despesa, **deduz-se** do dinheiro entregue na véspera (LIT.)

Se a esta matéria orgânica **deduzirmos** a que é gasta por esses operadores (DID.)

O titular terá de somar os montantes que **deduziu** no IRS (ECON.)

tinha o direito de **deduzir** na Retribuição do trabalhador o montante correspondente (JUR.)

Quadro VI

Deduzir \equiv Apresentar: estrutura sintáctica e frequências de ocorrência nos *corpora*

Estrutura sintáctica	ORAL	JORN.	LIT.	DID.	ECON.	JUR.
V N (contra N)	0	2 ¹	0	0	0	222

¹ - Nos dois casos, texto jornalístico sobre temas jurídicos.

Exemplos:

Na acusação **deduzida** pelo Ministério Público, quatro dos arguidos... (JORN.)

O Ministério Público na comarca de Lisboa **deduziu** acusação contra os arguidos (JUR.)

O conhecimento das estruturas prototípicas de cada item lexical e a desambiguação semântica que os contextos de uso possibilitam são factores de identificação dos termos, a que se junta o papel determinante que desempenham as associações lexicais. Estes três factores associados definem os padrões de um item lexical (lema ou forma); segundo HUNSTON e FRANCIS (2000: 37) "The patterns of a word can be defined as all the words and structures which are regularly associated with the words and which contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it".

Restringindo a observação, agora, apenas às combinatórias lexicais no *corpus* jurídico, observa-se claramente que, neste *corpus*, a estrutura sintáctica mais frequente de *deduzir* é V N (contra N) sendo que o verbo, nesta estrutura, tende a associar-se com um número restrito de itens lexicais, configurando um sentido particular que não ocorre noutros tipos de discurso. Os cinco lemas que neste *corpus* co-ocorrem mais frequentemente com *deduzir* são: *pedido* (54) - *pedido de indemnização* (19), *pedido cível* (25) -, *acusação* (41), *embargo* (21), *oposição* (19), *excepção* (14), *recurso* (10). Apresentam-se excertos da lista de concordâncias de *deduzir* que ilustram este padrão sintáctico-semântico e lexical.

Concordâncias de *deduzir*

stentes não tomaram qualquer atitude, pois **deduziram** acusação autónoma do Ministério Público, XXXXXXXXXXXX, esposa do arguido, que havia **deduzido** acusação contra XXXXXXXXXXXXXXXXXXXX pelo O Ministério Público na comarca de Lisboa **deduziu** acusação contra os arguidos XXXXXXXXXXXXXXXXXXXX, ovlhã, o Magistrado do Ministério Público **deduziu** acusação contra XXXXXXXXXXXXXXXXXXXXXXXXXXXX, s, em processo comum, o Ministério Público **deduziu** acusação contra os arguidos XXXXXXXXXXXXXXXXXXXX ulado por XXXXXXXXXXXXXXXXXXXX. 4) Foi contra ele **deduzida** acusação em processo comum e com intervenç oferecendo provas, requerendo diligências, **deduzir** acusação independente da do Ministério Públ de intervenção, pois podem os assistentes **deduzir** acusação independentemente da do Ministério

77, n. 3 do Código de Processo Penal e não rfilhado pelo Ministério Público desde que o desta acusação os assistentes podiam ter o do processo acima referido e no qual foi usação pública, e ter-se desinteressado de o corria seus termos, só depois tendo sido riação e Comercialização de Porcos, Lda. " uando se tenha em atenção que, podendo ser Direcção-Geral de Educação do Norte, veio , n. 2, do Código de Processo Civil, tendo citação, é perfeitamente legítimo que ele 4. Perante estes factos, o recorrente veio XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, er o seu direito. Em nenhuma delas poderão move a União de Bancos Portugueses S. A., e XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX rca de Guimarães, XXXXXXXXXXXXXXXXXXXXXXX m sede na Avenida 5 de Outubro, em Lisboa, ilelo, Lda. " contra XXXXXXXXXXXXXXX, este XXXXXXXXXXXXXXX, casada com o XXXXXXXX, XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX tivessem filha menor XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX, XXXXXXXX, também com os sinais dos autos, do Ministério Público, e podiam ainda ter XXXXXXXXXXXXXXX e a Autora, contra quem foi a XXXXXXXXXXXXXXXXXXXXXXX e que nele foi ículo, os herdeiros da falecida XXX vieram ravado -, todos do Código Penal de 1995) . iram à acusação do Ministério Público, nem idos, se constituíram assistentes. E ambos se constituíram assistentes nos autos, nem olha 190) , perfilhou a acusação pública e go Penal. O ofendido XXXXXXXXXXXXXXXXXXXXXXX cho: "Dado que XXXXXXXXXXXXXXXXXXXXXXX o em que disse aderir à acusação pública e a na sua petição inicial. O R. contestou e contestou por excepção e por impugnação e velada no acórdão recorrido, de terem sido velada no acórdão recorrido, de terem sido 2 do Código Penal. Contra ele foram também Como o poder de dispor da relação jurídica ia em factos ainda não provados para deles

deduziram acusação, nem pedido civil. Efectuado o j **deduziu** acusação. O crime de homicídio privilegiado **deduzido** acusação pelos factos acusados pelo Minist **deduzida** acusação por homicídio negligente contra o **deduzir** acusação própria. Finalmente, o direito de **deduzida** acusação pública. Com a notificação desta **deduziu** embargos a execução movida pelo "Banco Fons **deduzidos** embargos à execução em que a matéria aleg **deduzir** embargos à acção executiva que a esta mover **deduzido** embargos com fundamento em que a execução **deduza** embargos de terceiro para discutir a naturez **deduzir** embargos de terceiro para defender a sua po **deduzir** embargos de terceiro pedindo o levantamento **deduzir** embargos de terceiro (artigo 1037 do Código **deduzir** embargos de executado, alegando, por um lad **deduziram** embargos de terceiro contra XXXXXXXXXXXXXXX **deduziu** embargos de executado por apenso à execução **deduziu** embargos de terceiro na execução para entre **deduziu** embargos de executado, invocando, no essenc **deduziu** embargos de terceiro com fundamento em a dí **deduzido** pedido cível, foram os arguidos também sol **deduziu** pedido cível contra ambos os arguidos, por **deduziu** pedido cível contra aquele arguido pedindo **deduzido** pedido civil de indemnização, o qual por f **deduzido** pedido de indemnização cível, enquanto seg **deduzido** pedido de indemnização cível pelos herdeir **deduzir** pedido de indemnização cível nos termos do **Deduziram** pedido de indemnização civil contra o arg **deduziram** pedido de indemnização civil, pelo que nã **deduziram** pedido de indemnização civil: XXXXXXXXXXXXXXX, **deduziram** pedido de indemnização civil. 23. Tal con **deduziu** pedido de indemnização civil contra o XXXX, **deduziu** pedido de indemnização civil contra o argui **deduziu** pedido de indemnização civil, está por essa **deduziu** pedido de indemnização cível. Mas o Meritís **deduziu** pedido reconvenicional, replicando o A. no s **deduziu** pedido reconvenicional, pedindo a condenação **deduzidos** pedidos cíveis entretanto decididos ou de **deduzidos** pedidos cíveis entretanto decididos ou de **deduzidos** pedidos de indemnização civil pelos famil **deduzida** em juízo cabe, em geral, aos respectivos s **deduzir** consequências desfavoráveis ao acusado, não

Nota: Os nomes próprios foram substituídos por X.

Actualmente, a investigação terminológica baseada em *corpora* está em expansão e cada vez mais se procuram critérios apropriados à selecção dos textos especializados. Examinando as ocorrências dos termos em contextos autênticos, identificam-se factos linguísticos que lhes estão associados e que, tal como os próprios termos, serão considerados nas descrições terminológicas.

Em Portugal vêm sendo feitos esforços no sentido de se criarem *corpora* especializados que possibilitem vários níveis de análise dos *termos*, elementos que, como qualquer outro elemento de comunicação linguística, fazem parte das línguas naturais. Considera-se, nesta perspectiva, que a formulação de recomendações terminológicas só será adequada se se basear em descrições linguísticas dos termos propostos.

Agradecimentos

O meu agradecimento à Sandra Amendoeira, Florbela Barreto e Raquel Amaro pela constituição dos *sub-corpora*, tratamento e organização dos dados.

Referências Bibliográficas

BAAYEN, R.H. (1998), "Lexis, Word Frequencies and Text Types", *IV-V Jornades de Corpus Linguistics 1996-1997*, Barcelona, Institut Universitari de Lingüística Aplicada: Universitat Pompeu Fabra, pg. 87-102.

BANDEIRA DA SILVA, Maria Ana (1993), *Aspectos da Sintaxe e da Semântica de verbos de Língua Comum usados em Linguagem Informática (Contribuição para a sua descrição lexicográfica)*, Lisboa, Dissertação de Mestrado, FLUL.

BIBER (1989), *A Typology of English Texts*, Amsterdam: Mouton de Gruyter.

CABRÉ, M.T. (1998), "El discurs especializat o la variació funcional determinada per la temàtica", *Revista Caplletra: Variació lingüística*.

CABRÉ, M.T. (1999), *La terminologia: representación y comunicación. Elementos para una teoría de base comunicativa y outro artículos*, Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 1999.

CARVALHO, Anabela (1994), *Nomes de cor num corpus especializado. Moda e vestuário nos últimos trinta anos (Contribuição para a descrição lexicográfica dos nomes de cor)*, Lisboa, Dissertação de Mestrado, FLUL.

HALLIDAY, M.A.K. and J.R. MARTIN (1993), *Writing Science*, Pittsburgh, University of Pittsburgh Press.

HUSTON, S. e Gill FRANCIS (2000), *Pattern Grammar, A corpus-driven approach to the lexical grammar of English*, Amsterdam/Philadelphia, John Benjamins Publishing Company.

MEYER, I. and MACKINTOSH, K. (1996), "The *Corpus* from a Terminographer's Viewpoint", *International Journal of Corpus Linguistics*, Vol. I, Nº2, John Benjamins Publishing Company, Amsterdam/Philadelphia, pg. 255-285.

NETO, Paula Marquez (1995), *Combinatórias Lexicais no Discurso da Astronomia. Um Estudo em Estatística Lexical*, Lisboa, Dissertação de Mestrado, FLUL.

NUNES, Helena (2000), *Contributo para a caracterização da Linguagem Jurídica no domínio lexical*, Lisboa, Dissertação de Mestrado, FLUL.

PEARSON, J. (1998), *Terms in Context*, Studies in Corpus Linguistics, John Benjamins Publishing Company, Amsterdam/Philadelphia.

PHILLIPS, M.K. (1983), *Lexical Macrostructure in Science Text*, Ph.D thesis, University of Birmingham.

REBELLO DE ANDRADE, Ana Margarida (1995), *As palavras importadas no Léxico de Decoração*, Lisboa, Dissertação de Mestrado, FLUL.

SAIANDA, Maria Helena Rosado (1998), *A Palavra: instrumento de acesso ao poder. O caso particular do debate eleitoral em Portugal entre 1986 e 1991*, Évora, Dissertação de Doutoramento, Universidade de Évora.

SANTOS PEREIRA, Luísa Alice (1994), *Como se combinam as palavras? Contributo para um Dicionário de Combinatórias do Português*, Lisboa, Dissertação de Mestrado, FLUL.

SINCLAIR, J. (1996), "Tipologia Textual Eagles" in BACELAR NASCIMENTO, M.F.; RODRIGUES, C. e BETTENCOURT GONÇALVES, J. (Org.), *Actas do XI Encontro Nacional da APL*, Vol. I, *Corpora*, Lisboa, Colibri, pg. 39-91.

SINCLAIR, J.M. (1991), *Corpus, Concordance, Collocation*, Oxford: OUP.

TEUBERT, W. (1995), "Language Resources: The Foundations of a Pan-European Information Society", *Proceedings of the First European Seminar*, "Language Resources for Language Technology", Hungary, 1995, TELRI.

¹ TEUBERT, 1995, considera 100 milhões de palavras como dimensão *standard* para *corpora* gerais.

² Este *corpus* do Centro de Linguística da Universidade de Lisboa, actualmente com 152 milhões de palavras, é um *corpus* geral, monitor, que inclui importantes sub-*corpora* especializados. Foram realizados trabalhos académicos sobre estes sub-*corpora*, contrastando os resultados com os do *corpus* geral (Cfr. BANDEIRA DA SILVA, 1993, CARVALHO, 1994, REBELLO DE ANDRADE, 1995).

³ Uma destas ferramentas, O Dicionário de Combinatórias do Português, tem sido utilizada em trabalhos académicos (Cfr. PEREIRA, 1994) e, nomeadamente, aplicada a *corpora* especializados (Cfr. NETO, 1995, SAIANDA, 1998, NUNES, 2000).