

DICIONÁRIO DE COMBINATÓRIAS DO PORTUGUÊS (VARIEDADE EUROPEIA)^{1,2}

Maria Fernanda BACELAR DO NASCIMENTO³

- RESUMO: a partir de um *subcorpus* de 12 milhões de palavras do CRPC, extraíram-se todos os pares de palavras contíguas e os pares separados por uma, duas, três ou quatro palavras. Para cada par, é dada a informação seguinte: Frequência e Distribuição (oral, escrito, literatura, jornal etc.); Distância entre as palavras do par; Informação Mútua: IC; Localização do par no *subcorpus*; Concordâncias do par; Grupos fixos de 3, 4, 5 palavras.
- PALAVRAS-CHAVE: Léxico; lexicografia; combinatórias léxicas.

O *Dicionário de Combinatórias do Português* (DCP) é um projecto do Centro de Linguística da Universidade de Lisboa (CLUL), executado no âmbito do Programa Lusitânia, entre 1994 e 1997; constitui a primeira grande exploração, feita com carácter sistemático, do *Corpus de Referência do Português Contemporâneo* (CRPC), projecto também em curso no CLUL, e que conta actualmente 80 milhões de palavras que ocorreram em amostragens de textos escritos (da segunda metade do século XIX até 1998) e de discurso oral (de 1970 a 1997).

1 O texto obedece à ortografia do português de Portugal.

2 Programa "Lusitânia": Instituto Camões e Junta Nacional de Investigação Científica e Tecnológica (JNICT) - Projecto PLUS/C/LIN/816/93.

3 Centro de Linguística da Universidade de Lisboa, Portugal.

De forma sumária, pode dizer-se que um projecto como o do *Dicionário de Combinatórias do Português* determina e é determinado pelo actual desenvolvimento da interacção Linguística/Informática. O aproveitamento de versões electrónicas de textos ou de outros recursos informáticos tornou-se um método clássico na investigação em Ciências Humanas, particularmente em certas disciplinas linguísticas de que se destacam os estudos do léxico.

Inicialmente, a utilização das tecnologias informáticas nos estudos lexicais foi valorizada sobretudo como forma de resolver problemas relacionados com o armazenamento e organização de grandes colecções de dados, sua quantificação, indexação e contextualização. São conhecidos os resultados de publicações dicionarísticas que, seja pela dimensão gigantesca dos materiais recolhidos (veja-se, por exemplo, a obra monumental que é o *Trésor de la Langue Française*), seja pelas operações selectivas e analíticas que exigiram (veja-se, por exemplo, o excelente dicionário de inglês para estrangeiros *Essential English Dictionary* da Collins Cobuild), só foram possíveis graças à interdisciplinaridade Linguística/Informática a que recorreram.

Se os processos informáticos utilizados em Linguística têm permitido a observação e organização de grandes massas de dados textualmente atestados e, conseqüentemente, descrições mais objectivas dos fenómenos linguísticos, essas descrições, por sua vez, servem a concepção de modelos de processamento de língua natural. De facto, estes modelos recorrem, cada vez mais, àquelas descrições linguísticas, no intuito de resolver problemas postos pela automatização dos processos de compreensão, produção e aquisição da linguagem, com o objectivo último de construir processadores de linguagem computacional tão próximos quanto possível da linguagem humana.

No que respeita ao estado actual do desenvolvimento e interacção destas duas disciplinas (que visam, em última análise, a construção de um modelo da cadeia cognitiva) aplicado concretamente ao estudo da co-ocorrência lexical restrita, interessa acentuar a importância que lhe vem sendo atribuída nas hipóteses sobre a representação do léxico mental dos indivíduos; se considerarmos, como, por exemplo, Sinclair (1991), a existência de dois princípios alternativos e complementares, o princípio da *livre escolha* (o falante tem como única restrição a gramaticalidade do enunciado) e o *princípio idiomático* (o falante tem à sua disposição um grande número de "grupos de palavras" pré-construídos ou semi-pré-construídos que constituem escolhas únicas) e se, também como ele, considerarmos a hipótese de que os indivíduos

começam por usar as capacidades da memória e as rotinas, sendo os seus discursos preferencialmente constituídos por escolhas únicas, correspondentes, portanto, ao princípio idiomático, facilmente compreenderemos as razões que levaram ao grande desenvolvimento que actualmente se observa em estudos e trabalhos sobre combinatórias lexicais. É de salientar, contudo, a inexistência, até há bem pouco tempo, de um grande desenvolvimento nesta área para o português europeu.

Convém assinalar que o problema das combinatórias a incluir num dicionário, seja ele um dicionário de língua monolíngue ou plurilíngue, é dos mais difíceis problemas da lexicografia tradicional, precisamente porque a intuição dos falantes não é suficiente para determinar o grau de fixidez das combinatórias, nem a extensão, estabilidade e vitalidade do seu uso.

O *Dicionário de Combinatórias do Português* consiste num inventário de associações lexicais do português contemporâneo, estabelecido a partir de um *corpus* desenhado para o efeito (extraído do *Corpus de Referência do Português Contemporâneo* do CLUL) contendo diversos tipos de discurso falado e escrito do português europeu. As associações lexicais, que designámos por combinatórias, apresentam graus de coesão diversos: grupos totalmente cristalizados, semi-cristalizados ou apenas constituídos por co-ocorrentes privilegiados. Um repertório de associações lexicais assim concebido é, em si mesmo, de grande utilidade, mas é-o também como fonte de informações a incluir noutros dicionários unilíngues, bilíngues e multilíngues, na tradução tradicional ou assistida por computador, no levantamento de grupos neológicos vernáculos ou importados, no estabelecimento de tipologias diversas (por exemplo, da complementação verbal ou nominal), no estudo contrastivo das variantes do português e, naturalmente, na programação que visa a produção e reconhecimento automático da língua natural. Quando da apresentação do Projecto, estava previsto que o Dicionário se desenvolvesse dentro das formas clássicas de observação e análise de co-ocorrências lexicais em *corpora* (Concordâncias, Frequência, Repartição) e desse origem a um dicionário tradicional, no seu género (cf., por exemplo, Benson et al. 1986); contudo, a conjugação de vários factores fez alterar as previsões iniciais: o conhecimento aprofundado da vasta Bibliografia sobre modelos de associações lexicais (de que se destacam Church & Hanks, 1990; Calzolari & Bindi, 1990; Sinclair, 1991 e 1995; Smadja, 1993 e Biber, 1993), o contacto com outros projectos europeus, a colaboração neste projecto, aquando das suas deslocações a Portugal, de Sinclair da Universidade de Birmingham e

de Calzolari do Instituto de Linguística Computacional de Pisa, e as experimentações de novas metodologias de análise realizadas pela equipa de linguistas e informáticos que participaram neste projecto permitiram a perspetivação de objectivos qualitativa e quantitativamente mais ambiciosos. Consistiram estes novos objectivos em permitir ao utilizador o acesso informático directo a TODAS as combinatórias que ocorreram num *corpus* de 12.282.392 palavras com diferentes posições relativamente à palavra em estudo (palavra-nó), assim como o acesso directo aos contextos em que ocorrem (contextos restritos ou alargados), Frequência de Ocorrência e Repartição por tipo de discurso e significância da combinatória no *corpus* (Índice Combinatório determinado estatisticamente), conjunto de informações a seleccionar e a manipular directamente pelos utilizadores de acordo com os seus objectivos particulares. Isto porque se considerou que o Projecto, tal como se ia redefinindo, se tornava uma fonte inesgotável de informações que permitiam ultrapassar grandemente as finalidades inicialmente previstas. Este dicionário, o primeiro do seu género, em Portugal, se beneficia do facto de ser realizado com base num *corpus equilibrado*, representativo de variedades do português falado e escrito, estabelecido especificamente para este efeito e, tal como já se disse, extraído do CRPC. Uma das vantagens que decorrem da utilização do *corpus* tem que ver com os dados quantitativos e estatísticos obtidos, dados absolutamente determinantes para, em conjunto com outros critérios de análise linguística, se poder vir a atribuir a um grupo de palavras o "estatuto" de combinatória com estabilidade na língua e isto, principalmente, no caso de formações recentes.

Existe, pois, uma estreita relação entre a qualidade última do trabalho e os dados observados no *corpus*, o que significa que as experiências para a selecção do *corpus* são determinantes nos resultados da investigação.

Depois de várias experiências estabeleceu-se o *corpus* definitivo do DCP, constituído por 11.426.197 palavras de discurso escrito e 856.195 palavras de discurso oral. *Corpora* com dimensão de cerca de 10 milhões de palavras têm sido considerados capazes de assegurar a fiabilidade dos resultados na extracção de combinatórias (Cf., por exemplo, Smadja, 1993).

O *corpus* ficou constituído da seguinte forma:

LÍNGUA ESCRITA:

Discurso jornalístico

60%

Discurso literário	24%
Discurso técnico, científico e didático	10%
Discurso político	4%
Discurso jurídico	2%

LÍNGUA FALADA:

Discurso espontâneo	80%
Discurso formal	20%

Iniciou-se a exploração do *corpus* com extracção de concordâncias de tipos diferentes:

- contextos de diversas dimensões: 1, 2, 3, 4 linhas antes e depois da palavra em estudo (palavra-nó);
- contextos de 1 linha, incluindo a palavra-nó a meio da linha e 5 palavras antes e 5 palavras depois do nó;
- contextos ordenados de diferentes formas, quer alfabeticamente antes ou depois do nó, quer de acordo com as frequências dos co-ocorrentes.

Com base nas propostas teóricas e nas técnicas descritas em Church & Hanks (1990) e em Sinclair (1991) procedeu-se à tentativa de caracterização dos grupos lexicais, a partir de “medidas de informação mútua” que definem indicadores estatísticos sobre índices combinatórios, índices de tipicidade, distância média entre o nó e o co-ocorrente, variância e fixidez das combinatórias.

A programação informática realizada para este Projecto permite, actualmente, a extracção de pares de palavras com a indicação da distância a que se encontra um do outro os elementos dos pares (distâncias 1-2-3-4), posição que ocupam relativamente ao nó (posição anterior ou posterior) e a respectiva Frequência de ocorrência (sempre igual ou superior a 2).

O Programa CONCORD.CB percorre o *corpus* e determina todas as ocorrências de palavras que co-ocorrem a determinada distância. O utilizador pode indicar: a distância máxima a que pretende observar os pares e a frequência a partir da qual pretende considerar os pares. Obtém-se, pois, com este programa, uma lista de todos os pares co-ocorrentes no *corpus* com a respectiva frequência e distância a que co-ocorrem, conjuntos especificados de pares e os contextos em que ocorrem. Elaboraram-se, ainda, programas que agrupam dados (palavras e seus pares, palavras e suas frequências) e efectuaram-se cálculos estatísticos que

relacionam a frequência das palavras com a frequência dos pares e com as frequências dos elementos no total do *corpus*.

No Quadro 1 apresentam-se resultados obtidos num *corpus* experimental quanto a indicadores estatísticos que permitem seriar combinações da palavra-nó pressão, de uma forma rigorosa.

Quadro 1 – “Medidas de Informação Mútua da Palavra-nó **pressão**”³.
Total Corpus

Lema	Pressão										
f(x)	1147										
	Palavra y	fa (x,y)	f (y)	f-2	f-1	f+1	f+2	m	s ²	IC	IF
Pressão	exercer	13	325	5	8	0	0	-1.385	0.256	8.445	23.694
Pressão	ceder	6	159	6	0	0	0	-2	0	8.361	83.609

Legenda:

f(x) Número de vezes que o lema x (pressão/pressões) ocorre no *corpus*.

y Co-ocorrente.

fa(x,y) Número de vezes que o co-ocorrente y ocorre na janela, i. e., na proximidade de -2 a +2 palavras do nó.

f(y) Número de vezes que o co-ocorrente y ocorre no *corpus*.

f-/+ Número de vezes que o co-ocorrente y ocorre na posição i, com i = -2,-1,1,2.

m Média das distâncias da palavra y.

s² Variância da palavra y.

IC Índice Combinatório da palavra y com o lema x.

IF

Índice de Fixidez da palavra y com o lema x.

No Quadro 2 apresenta-se um excerto de concordâncias da palavra **pressão** ordenadas pela frequência dos grupos, num *corpus* experimental.

3 “O Índice Combinatório (IC) entre a palavra-nó x e a palavra co-ocorrente y é o índice calculado entre as probabilidades de ocorrência conjunta dos pares de palavras de x e de y (x, y) e de ocorrência independente das mesmas palavras. O IC é indicador do grau de significância da combinação entre as mesmas palavras.

O Índice de Fixidez (IF) da combinação obtém-se da correção do Índice Combinatório pela consideração da variância e permite distinguir, de entre palavras com idênticos Índices Combinatórios, aquelas em que esta é de natureza mais fixa ou menos fixa.

O Índice de Tipicidade (IT) da combinação entre as palavras x e y (x, y) é o índice calculado pela relação entre a frequência de ocorrência conjunta de x e y e a frequência de ocorrência independente ponderada de y. O IT é indicador da Tipicidade da combinação de y com x.

A variância (s²) da distância de cada palavra da janela relativamente ao nó ... é um bom indicador do grau de fixidez dos grupos de palavras. (PEREIRA, L. A. S., 1994, p. 137-49).

Quadro 2 – Concordâncias por ordem crescente da frequência do grupo

F-2	F-1	F+1	F+2	m	S ²	IC	
0	0	9	0	1	0	11,451	As altas pressões subtropicais , funcionando
0	0	9	0	1	0	11,451	as altas pressões subtropicais juntamente
0	0	9	0	1	0	11,451	as altas pressões subtropicais pela
0	0	9	0	1	0	11,451	de altas pressões subtropicais , sendo
0	0	9	0	1	0	11,451	de altas pressões subtropicais , é
0	0	9	0	1	0	11,451	às altas pressões subtropicais . Os
0	0	9	0	1	0	11,451	Das altas pressões subtropicais nomeadamente
0	0	9	0	1	0	11,451	das altas pressões subtropicais . Na
0	0	9	0	1	0	11,451	das altas pressões subtropicais que
5	8	0	0	-1,385	0,256	8,445	a exercer pressão o que
5	8	0	0	-1,385	0,256	8,445	a exercer pressão sobre a
5	8	0	0	-1,385	0,256	8,445	a exercer pressão sobre elas
5	8	0	0	-1,385	0,256	8,445	e exercer pressões junto de
5	8	0	0	-1,385	0,256	8,445	de exercer pressões sobre a
5	8	0	0	-1,385	0,256	8,445	exercer uma pressão excessiva sobre
5	8	0	0	-1,385	0,256	8,445	exercer uma pressão sobre aquele
5	8	0	0	-1,385	0,256	8,445	para exercer pressão sobre os
5	8	0	0	-1,385	0,256	8,445	para exercer pressões sobre a
5	8	0	0	-1,385	0,256	8,445	exercer maior pressão sobre Gaza
5	8	0	0	-1,385	0,256	8,445	exercer alguma pressão Este é
5	8	0	0	-1,385	0,256	8,445	exercer fortes pressões no sentido
5	8	0	0	-1,385	0,256	8,445	conseguir exercer pressões suficientes para

Considerámos que os materiais a disponibilizar deveriam conter os dados exaustivos e informações extensivas sobre os mesmos, por forma a não cortar nenhuma das possibilidades de pesquisa oferecidas por este dicionário que tendo sido inicialmente previsto como um dicionário "corpus-based", veio a tornar-se um dicionário "corpus-driven". Como afirmou Sinclair, no resumo da comunicação que apre-

sentou ao XI Encontro da Associação Portuguesa de Linguística, em Outubro de 1995:

Very few dictionaries are "corpus-driven", that is the lexicographers try to record without distortion the most common patterns of the language, selecting examples directly from the data, highlighting natural phraseology and pointing out regular contextual restrictions. It is argued that corpus evidence is so dramatically different from our expectations that the humbler position of the corpus-driven methodology is at present more reliable than the corpus-based one.

O produto que obtivemos pode aproximar-se do *COBUILD English Collocations on CD-ROM*, publicado em 1995 sob a direcção de Sinclair. Esta obra parte de um inventário exaustivo como aquele que constitui o Dicionário de Combinatórias do Português e, na impossibilidade de incluir num CD-ROM uma tal quantidade de dados (tanto mais que o *corpus* observado era de dimensão superior ao do português), procedeu-se a uma selecção automática dos exemplos a publicar. Sendo este produto de excepcional importância, a selecção automática dos exemplos levou a que ele contenha combinatórias de menor interesse do ponto de vista quantitativo ou mesmo linguístico:

The actual selection of examples was made at random by computer with the only consideration being the space available on the CD-ROM. Because of this, you may find some unusual examples, and possibly a few offensive ones, for which we apologize. In general as we have come to expect from the Bank of English, the examples are helpful and indicate the typical phraseology of the collocation. In corpus work you quickly learn that there are some odd and unusual occurrences in natural language. Sometimes the context is not sufficiently long to explain why a word sequence has arisen, and sometimes there are mistakes, misprints, and peculiar usages. Hundreds of millions of people speak and write English all the time, and they are all different and come from all over the world, and differ in age by up to a century. The miracle is not that they don't always agree on how a English word is used, but that they agree at all (op. cit. Introdução).

No intuito de comprovar a eficácia deste dicionário *corpus-driven* para determinadas aplicações, considerou-se que teria interesse apresentar na fase final do Projecto exemplos que resultassem das análises sistemáticas de alguns lemas, tendo-se colocado, a equipa, na perspectiva de utilizadores interessados, essencialmente, quer em seleccionar dados a integrar em dicionários gerais da língua de formato tradicional,

electrónicos ou informatizados, unilingues ou multilingues, quer em seleccionar dados para estudos de padrões combinatórios lexicais e sintácticos do português. Estas experiências de aplicação foram realizadas na fase final do dicionário, manualmente, ou seja, partindo das selecções automáticas a equipa observou e analisou dados do ponto de vista linguístico, seleccionando-os em conformidade com os resultados dessa análise.

Neste ponto importa assinalar que a investigação sobre um *corpus* permitiu a observação controlada de grandes massas de dados, de forma a determinar tendências fortes no uso da língua, factor fundamental em trabalhos de linguística teórica, descritiva e aplicada e em trabalhos interdisciplinares, designadamente em linguística computacional, estatística lexical, psicolinguística e sociolinguística.

A investigação sobre *corpora* tem tornado patente que as línguas naturais são, quer do ponto de vista paradigmático, quer do ponto de vista sintagmático, mais padronizadas do que os estudos baseados na intuição faziam prever. É também o estudo de *corpora* que permite identificar esses padrões (léxico-sintáctico-semânticos) e a regularidade com que ocorrem e interagem nos diversos usos.

No caso do *Dicionário de Combinatórias do Português*, os fenómenos associativos contínuos e descontínuos que determinámos parecem ser de natureza diversa: *textual* (observada repetidamente em determinados textos ou tipos de textos, potencializam análises textuais e estilísticas), *discursiva* (observados em situações enunciativas diversas, terão papel importante na análise do discurso) e, em grande medida, *cognitiva* (serão fundamentais em estudos sobre a aquisição e a memória lexical dos indivíduos); estes fenómenos representam, pois, relações diversas que os utilizadores podem analisar em termos quantitativos e qualitativos, utilizando as ferramentas informáticas disponíveis, para identificar a extensão e os tipos particulares de relação associativa.

Estes conhecimentos, como se disse, constituem elementos de grande interesse nos estudos linguísticos, em geral, sendo fundamentais na construção teórica e na prática lexicográfica. Assim, os materiais que actualmente se encontram disponíveis para consulta, no Centro de Linguística da Universidade de Lisboa, constituem, mais do que um dicionário, um dicionário dos dicionários, fornecendo informações sobre a frequência de uso de lemas e de formas lexicais isoladamente ou em associação, e sobre factores morfológicos, sintácticos, semânti-

cos e colocacionais que caracterizam padrões associativos do português contemporâneo.

Não se trata de fornecer “restrições de selecção” das associações baseadas em conceitos gramaticais ou de congruência semântica ou, ainda, de natureza pragmática, noções essas intuitivas e que supõem a correlação permitido/*vs*/proibido. O que está à disposição dos utilizadores são *selecções preferenciais observáveis e procedimentos informáticos e estatísticos* que os ajudarão a distinguir factores relevantes do puro ruído e também a isolar fenómenos ou induzir generalizações.

Os padrões associativos são por nós entendidos como uma extensão da noção tradicional de colocação.⁴ Ampliando a noção firthiana de colocação, propomos que a identificação de *combinatórias* (modelos associativos gradativos e mais ou menos padronizados que excluem a associação livre e atingem a frase idiomática) se faça de acordo com processos de natureza quantitativa (medidas de informação mútua) e qualitativa (recorrendo à observação e interpretação dos dados reais) o que permite caracterizar as palavras:

- na sua relação com outras palavras com as quais sistematicamente ocorrem de forma contínua ou descontínua (padrões co-ocorrências);
- na sua relação com traços gramaticais morfossintácticos e sintácticos – próprios ou dos co-ocorrentes; por exemplo, determinado lema ou forma flexionada do lema ocorre sistematicamente com determinado tipo de verbo, tempo verbal ou construção sintáctica (padrões gramaticais);
- nas suas relações extralinguísticas – enunciativas, situacionais, contextuais – já que associações fortes em determinado registo de língua podem ser associações fracas noutro registo de língua (padrões discursivos).

Sendo manifestamente impossível caracterizar as palavras segundo estes padrões por recurso à intuição e dado que eles se evidenciam com demasiada sistematicidade para serem tidos como acidentais, tornou-se clara a enorme importância do seu estudo com base em *corpora* assim como a construção de instrumentos de análise adequados.

4 Termo introduzido por FIRTH nos anos 50. Este autor chamou fortemente a atenção para o estudo das co-ocorrências lexicais, mostrando que a caracterização de aspectos fundamentais do significado de uma palavra decorre do conjunto de todas as outras palavras que tipicamente ocorrem com ela.

Os resultados obtidos são os seguintes:

Os utilizadores podem seleccionar todas as palavras que ocorreram em pares com F igual ou superior a 2, aceder ao tipo de informação disponível sobre elas e manipular as concordâncias de acordo com os seus próprios objectivos.

Estão disponíveis, para consulta, os seguintes materiais:

a) *Corpus DCP*: 12.282.392 palavras (*sub-corpus* oral: 856.195, *sub-corpus* escrito: 11.426.197).

b) Índice de 233.543 (*subcorpus* escrito) e de 33.030 (*subcorpus* oral) formas lexicais que ocorreram no *corpus* e respectiva Frequência (F) de ocorrência, apresentadas de forma especificada conforme se trata do *corpus* de língua falada ou do *corpus* de língua escrita.

c) Índice dos lemas teóricos⁵ correspondentes às formas lexicais referidas na alínea b).

d) Índice dos 2.428.809 *pares* diferentes de formas lexicais que ocorreram no *corpus* com F igual ou superior a 2. A distância entre as formas que constituem os pares pode ser de 1 (formas contíguas), 2, 3 e 4. Sendo a palavra nó representada por A e o seu par por B, teremos:

AB

A - B

A — B

A — B

BA

B - A

B — A

B — A

Destes índices constam:

- Frequência de ocorrência da palavra-nó no *corpus*;
- Frequência de ocorrência da palavra-nó no par, ou seja, a frequência do par;
- O Índice Combinatório (IC) do par.

e) Índices de lemas co-ocorrentes.

f) Listas brutas de concordâncias de todos os pares com 1 linha de contexto, contendo toda a informação quantitativa e estatística relati-

⁵ Teóricos porque não sendo o *corpus DCP* um *corpus* anotado, os lemas extraídos automaticamente estão inflacionados por sobregeração, daí ser irrelevante fornecer aqui o n° de lemas. Contudo, a lematização automática constituiu uma fase interessante do Projecto pois permitiu estabelecer índices associativos que têm como nó um vocábulo e todas as suas formas flexionadas.

vamente ao lema e às formas e a identificação do tipo de discurso a que pertence, de acordo com a tipologia estabelecida e enunciada na alínea a).

As concordâncias aparecem indexadas por ordem decrescente dos Índices Combinatórios. Também é possível indexá-las com outras ordenações. Dentro de cada conjunto de concordâncias de um par, os contextos podem ser ordenados por ordem alfabética das palavras que vêm à direita ou à esquerda do par ou, simplesmente pela ordem de ocorrência no *corpus*.

Optámos por não classificar morfossintacticamente os lemas para não impor limitações às possibilidades de análise que o Dicionário oferece. Por exemplo nos grupos nominais, constituiria uma limitação para o utilizador a separação dos contextos segundo a sua realização como substantivos ou como adjectivos, ou ainda, no caso dos Particípios Passados, será certamente muito produtivo que o próprio utilizador possa estabelecer os seus critérios quanto à realização adjectival, de acordo com a observação de todas as ocorrências.

g) Listas de concordâncias com contexto expandido até à dimensão adequada aos objectivos do utilizador.

h) Listas de grupos contíguos de 5, 4, 3 e 2 palavras com as respectivas Frequências de ocorrência e contextos em que ocorreram.

i) Resultados de análises linguísticas realizadas manualmente sobre 26.568 pares diferentes (correspondentes a 60 nós) com o objectivo de demonstrar o interesse da reutilização deste dicionário de base em aplicações lexicográficas quer de carácter mais geral (selecção visando a delimitação de grupos lexicalizados ou com tendência para a lexicalização) quer de carácter mais especializado (selecção tendo em vista, por exemplo, dicionários de tipo analógico (co-ocorrências descontínuas) ou sintáctico (co-ocorrências que apontam para a selecção de estruturas sintácticas).

Excerto de resultados obtidos para a palavra-nó **pressão** em textos escritos do *corpus* definitivo

*** 9 SUBTROPICAL (real:9) ***

Pressões subtropicais 8.731

9 pressões subtropicais 1

201065274 mais próximo das áreas de altas
201065253 am. O ar polar alimenta as altas
201065218 westerlies que provêm das altas
201065232 ssões no Mediterrâneo. Das altas
201065260 entos dos jet streams e às altas
201065239 s e, à direita, origina as altas
201065267 devido às influências das altas
201065225 nvolvidos por centros de @baltas

pressões subtropicais, funcionam
pressões subtropicais juntamente
pressões subtropicais. Na região
pressões subtropicais, nomeadame
pressões subtropicais@b. Os fenó
pressões subtropicais pela subsi
pressões subtropicais que migram
pressões subtropicais@b, sendo o

*** 44 ATMOSFÉRICO (real:46) ***

Atmosférica pressão 8.643

40 pressão atmosférica 1

201038714 resão exercida pelo ar chama-se
201038896 ormenor em capitulos anteriores.
201038756 s sob a acção da pressão do ar -
201038763 em que é corrente exprimirmos a
201038784 es, melhor equipadas, analisam a
201038791 a temperatura, a precipitação, a
201038805 diminuindo. b) O @bdecrécimo da
201038798 itude, da composição química, da
201038686 li não havia qualquer mudança de
201038875 emperatura sobe ligeiramente e a
201038931 e admissão é P-I- (normalmente a
201038952 do cilindro cai para o valor da
201038833 altitude está relacionada com a
201038742 erna dos pulmões fica inferior à
201038924 ias de temperatura e humidade, a
201038812 eso. Ou, mais simplesmente, se a
201038882 porque se verifica um aumento da
201038721 ma-se @bpressão atmosférica. - A
201038945 ar comprimido introduzido até à
201038826 s variações da temperatura com a
201038679 nações: barómetros, para medir a
201038840 de de superfície dá-se o nome de
201038917 lugar terrestre que determina a
201038777 em-se à temperatura ambiente e à
201038770 neus de trás, de 27 lb/pol+2+. A
201038735 A compressão do ar é superior à
201038910 grandes altitudes por redução da
201038938 xpansão para as condições livres
201038903 de vista climatológico, a baixa
201038728 m influência sobre a variação da
201038693 como provar a sua existência?@b
201038700 resão exercida pelo ar chama-se
201038819 sobre nós; daí, a diminuição da
201038889 to maior quanto mais baixa for a
201038707 s a esta permitem concluir que a
201038868 ecebeu ao longo do dia; assim, a
201038749 sobre todos os corpos originam a
201038854 gual modo, podemos afirmar que a
201038861 enor a pressão. Concluímos que a
201038847 do; assim, podemos afirmar que a

*** 81 EXERCER (real:90) ***

Pressões exercidas 8.521

4 pressões exercidas 1

201061860 Mundo de 1998, em França. Daí as
201061874 atribuiu a atitude tailandesa a

@b**pressão atmosférica.** - A press
@b**Pressão atmosférica**@b - diminui
pressão atmosférica - porque est
pressão atmosférica; 1 atm= 760
pressão atmosférica, a estrutura
pressão atmosférica, a nebulosid
pressão atmosférica com a altitu
pressão atmosférica, da fuga moi
pressão atmosférica, De resto, a
pressão atmosférica diminui. d)
pressão atmosférica) e a pressão
pressão atmosférica e a força da
pressão atmosférica e com a temp
pressão atmosférica, e o ar entr
pressão atmosférica é de 1, 03kg
pressão atmosférica é o @bpeso q
pressão atmosférica. Está-se, po
pressão atmosférica exerce-se em
pressão atmosférica exterior. Um
pressão atmosférica. Fácil é con
pressão atmosférica; manómetros,
@b**pressão atmosférica** @b; mede-se
pressão atmosférica nesse mesmo
pressão atmosférica normal. || @t
pressão atmosférica normal tem o
pressão atmosférica. O ar compri
pressão atmosférica. O diagnósti
(pressão atmosférica), o que rep
pressão atmosférica. ocasiona uma
pressão atmosférica. Os aparelho
@b**Pressão atmosférica**@b Os resul
@b**pressão atmosférica**@b. Pois cl
pressão atmosférica. Por curiosi
pressão atmosférica. Se a evapor
@b**pressão atmosférica** se exerce
pressão atmosférica será mais al
pressão atmosférica. @t@bConclus
@b**pressão atmosférica** varia na r
pressão atmosférica varia na raz
pressão atmosférica varia segund

pressões exercidas pela embaixad
pressões exercidas pela embaixad

201061867 de opinião, mas sim as ameaças e
201061853 factos, como aponta, "das enormes

25 pressão exercida 1

201047234 , do que um casamento falhado. A
201047115 ão. Tento rodar o carreto, mas a
201047108 cado abandonado. Recentemente, a
201047255 aranda dos Windsor, não? DJOB- A
201047101 encontro. E nem quando baixou a
201047129 ocupar o lugar que ele deixou. À
201047136 xerce pressão sobre os corpos. À
201047143 rce-se em todos os sentidos@b. A
201047199 essão. Podemos assim dizer que a
201047206 essão exercida pelo ar é igual à
201047185 a Polónia, tem sido importante a
201047150 reia, é maior na situação I. @lA
201047094 em mais 70 milhões de dólares. A
201047213 eco (sem humidade) adicionada da
201047087 o entanto, alguns problemas pela
201047164 s frequentemente tensão. || @t@bA
201047248 randa dos Windsor, não? DJOB - A
201047192 a-se cada vez mais em virtude da
201047227 Franco considerou inaceitável a
201047178 elo capital e, por outro lado, a
201047122 encontra a massa magmática e da
201047171 líquido resulta na diminuição da
201047220 vamente, um termo próprio para a
201047157 o devida à acção de uma força de
201047241 a da filha. É difícil imaginar a

8 pressão exercidas 1

201047481 re as intensidades das forças de
201047488 e as intensidades das forças de
201047446 evidentemente fechada, as forças de
201047460 ncia mostra-nos que as forças de
201047439 e@i. Anteriormente, as forças de
201047453 inatamente sujeita às forças de
201047467 explicar o facto de as forças de
201047474 estão representadas as forças de

Exerceram pressão 7.287

2 exerceram pressão 2

116064858 vitória dos Estados Unidos, que
116064851 perações a meio-campo, zona onde

exercendo pressão 7.142

4 exercendo pressão 1

116057229 actuarem em todas as direcções,
116057222 sangue entra nessas cavidades e,
116057215 correctamente estas e apertam-se
116057208 em nada conter. O ar está sempre

2 exercendo pressão 3

pressões exercidas publicamente

pressões exercidas sobre a Inгла

pressão exercida contra esta mul
pressão exercida lá em baixo tir
pressão exercida pela edilidade
pressão exercida pela mídia, tin
pressão exercida pela selecção p
pressão exercida pelo ar chama-s
pressão exercida pelo ar chama-s
pressão exercida pelo ar sobre o
pressão exercida pelo ar é igual
pressão exercida pelo ar seco (s
pressão exercida pelo movimento
pressão exercida pelo peso do ti
pressão exercida pelo poder para
pressão exercida pelo vapor de á
pressão exercida pelos adversári
pressão exercida pelos líquidos
pressão exercida pelos "media" t
pressão exercida pelos @imovimen
pressão exercida pelos sindicali
pressão exercida pelos trabalhad
pressão exercida por essa massa
pressão exercida por esse gás na
pressão exercida por grupos que
pressão exercida sobre uma dada
pressão exercida sobre uma mãe,

pressão exercidas na face superi
pressão exercidas na face inferi
pressão exercidas na sua superfi
pressão exercidas pela água sobr
pressão exercidas pelo ar e vapo
pressão exercidas pelo ar atmosf
pressão exercidas por um fluido
pressão exercidas sobre o corpo

exerceram forte pressão para que
exerceram grande pressão. Quando

exercendo pressão nas superfície
exercendo pressão nas válvulas a
exercendo pressão sobre as super
exercendo pressão sobre os corpo

116057238 pre toda a superfície da grosa e **exercendo uma certa pressão.** @tR
 116057245 eve até ao intervalo. A Ucrânia, **exercendo uma grande pressão** no
pressão exerce 6.795
4 exerce pressão 1
 116052990 a de vapor de água o qual também **exerce pressão.** Podemos assim di
 116052976 do tijolo. Dizemos que o tijolo **exerce pressão** sobre a areia. Ma
 116052983 , à superfície do saco: @ia água **exerce pressão** sobre a parede do
 116052969 sípido@b e @binvisível@b. – O ar **exerce pressão** sobre os corpos.
exercer pressões 6.377
4 exercer pressões 1
 116062864 Silvestre, é acusado de estar a **exercer pressões** junto dos profé
 116062871 ando que Ancara "pretende apenas **exercer pressões** para que a Síri
 116062885 o social, resistam à tentação de **exercer pressões,** respeitem a au
 116062878 agita o tema da dissolução para **exercer pressões** sobre a conduta
2 exercer pressões 2
 116062901 e a França, país que tem vindo a **exercer fortes pressões** no senti
 116062894 problema é que Clinton não pode **exercer grandes pressões** sobre o
pressões exercem 6.246
2 exercem pressão 3
 116055940 co@i. Não são só os líquidos que **exercem forças de pressão,** em to
 116055947 em contacto com qualquer corpo, **exercem forças de pressão,** força
exercem pressão 6.083
2 exerceu pressão 2
 116066230 o maior domínio da Irlanda, que **exerceu forte pressão** no meio-ca
 116066237 s, no Estádio Fournier, Portugal **exerceu maior pressão** atacante e
exercer pressão 5.848
6 exercer pressão 1
 116062774 empresários burlados passaram a **exercer pressão** junto de Alexand
 116062781 admitir suspeitas de que estou a **exercer pressão,** o que para mim
 116062788 avizinham para os pescadores, e **exercer pressão** para que o Gover
 116062802 oduzidos na combustão, de modo a **exercer pressão** sobre a parte mó
 116062795 a-se às peças a soldar de modo a **exercer pressão** sobre elas. Dura
 116062767 . A resistência anunciou que vai **exercer pressão** sobre Portugaal
4 exercer pressão 2
 116062825 as são tão numerosas que poderão **exercer alguma pressão.** Este é o
 116062818 e!» No fim de contas trata-se de **exercer alguma pressão** psicológi
 116062811 cio do segundo tempo conseguiram **exercer alguma pressão** sobre a d
 116062832 corre à infidelidade quando quer **exercer uma pressão** sobre aquele
 *** 26 **ARTERIAL (real:26) *****
pressão arterial 8.256
26 pressão arterial 1
 201038251 fumam têm valores mais baixos de **pressão arterial.** À partida, par
 201038335 s de alto risco, como tabagismo, **pressão arterial** alta ou um enfa
 201038258 s vai conduzir a uma elevação da **pressão arterial.** c) Álcool – Pe
 201038202 intestinais, náuseas e subida da **pressão arterial.** Contudo, se in
 201038209 caracteriza por uma elevação da **pressão arterial.** De qualquer mo
 201038244 bulatórios, que permitem medir a **pressão arterial** do indivíduo du

201038300 ntamento de pesos que aumentam a
 201038307 es (contribuindo para diminuir a
 201038195 ia severa, palpitações, queda da
 201038328 ia que o tabaco origina. Eleva a
 201038174 ai pela aorta, que tem o nome de
 201038286 rtamento, peso equilibrado, melhor
 201038237 confiança e rigor na medição da
 201038293 exercício físico regular reduz a
 201038321 lesterol em circulação, quando a
 201038314 tinha efeitos favoráveis sobre a
 201038265 ão ter qualquer interferência na
 201038279 ar um medicamento para reduzir a
 201038342 experimentam uma baixa súbita de
 201038230 ipais factores que influenciam a
 201038181 s com que contactam. 56. 1 - @iA
 201038272 e álcool contribui para elevar a
 201038167 vasos sanguíneos e faz baixar a
 201038188 ção dos pesos; a determinação da
 201038223 ção e, obviamente, com o peso. A
 201038216 De salientar ainda que o padrão

*** 67 ALTA && ALTO (real:71) ***

altas pressões 8.075

48 altas pressões 1

18647758 os - que, em altitude, deixam as
 18647653 osfera. Assim, ao descer origina
 18647940 urbações alternam com centros de
 18647891 s, pois, além de predominarem as
 18647905 rológico mereça ser estudado. As
 18647646 a - e nesse caso, às baixas e às
 18647835 ar descendentes que originam as
 18647863 o Inverno. Durante o Inverno, as
 18647947 em que predominam os núcleos de
 18647807 zonas mais secas correspondem às
 18647919 r da Europa e da Ásia centros de
 18647884 ninegrado à latitude de 60°N: As
 18647730 do continente asiático originava
 18647842 deve-se sempre a uma situação de
 18647660 a **baixas pressões** à superfície e
 18647793 bre o continente (em situação de
 18647751 udes subtropicais, originando as
 18647674 no norte; no hemisfério sul, as
 18647912 centro da Europa. Este centro de
 18647877 assim, no Inverno, os núcleos de
 18647933 damentalmente, daquele centro de
 18647821 s ocorrem na estação quente e as
 18647744 ntos acontecem quando se centram
 18647681 apercebemo-nos da existência de
 18647800 eja, em condições, por norma, de

pressão arterial durante o esfor
pressão arterial e melhorar a pe
pressão arterial e náuseas. @iNi
pressão arterial e o ritmo cardí
pressão arterial e que se pode d
pressão arterial. e) Sedentarism
pressão arterial. Ele é o padrão
pressão arterial em cerca de 10
pressão arterial está elevada e
pressão arterial, etc. Actualmen
pressão arterial. No entanto, da
pressão arterial. Portanto, peso
pressão arterial, que pode mesmo
pressão arterial. Realizaram-se
pressão arterial resulta das for
pressão arterial, sendo portanto
pressão arterial. @t@iinfluência
pressão arterial (valores de 18
pressão arterial varia até ao lo
pressão arterial varia constante

altas pressões à direita e as ba
altas pressões à superfície, mas
altas pressões (anticiclones) co
altas pressões, as massas de ar
altas pressões caracterizam o cl
altas pressões chamam-se @bcentr
altas pressões continentais (ant
altas pressões continentais e as
altas pressões, da frente intert
altas pressões das altas latitud
altas pressões de origem térmica

altas pressões do Inverno, espec
altas pressões e as massas de ar
altas pressões, e esta situação
altas pressões em altitude. São
altas pressões). Este ar chega d
altas pressões. Este esquema fic
altas pressões ficam para a dire
altas pressões forma-se em conse
altas pressões impedem as precip
altas pressões. Mas há outros fa
altas pressões na estação fria.
altas pressões na Itália e baixa
altas pressões nas altas latitud
altas pressões. Nas latitudes mé

18647870 . Se durante o Inverno se formam
 18647632 ntros de acção. Numa situação de
 18647625 anticiclónicos@b ou @bcetros de
 18647856 om os máximos no Verão devido às
 18647618 ões tropicais (desertos secos de
 18647667 de frente no rosto, ficam-nos as
 18647723 uesa. c) @bZonas polares@b - nas
 18647695 udes e na região equatorial. Das
 18647688 es nas altas latitudes (chamadas
 18647639 na-se pesado e desce, originando
 18647828 midos, do mar para a terra, e as
 18647898 e calor por radiação originam as
 18647926 ca, já mais próximo das áreas de
 18647772 ubtropicais pela subsidiência. As
 18647779 t stream. O ar polar alimenta as
 18647702 e@b ou westerlies que provêm das
 18647737 as pressões no Mediterrâneo. Das
 18647786 slocamentos dos jet streams e às
 18647765 titudes e, à direita, origina as
 18647814 Verão, devido às influências das
 18647709 estão envolvidos por centros de
 18647849 s, provocando assim uma faixa de
 18647716 nós. b) @bZona temperada@b - das

3 pressões altas 2

201059295 mais secas correspondem às altas
 201059281 que a pressão do líquido atinja
 201059288 ebemo-nos da existência de altas

alta pressão 4.729

13 alta pressão 1

18625816 es de fácies, nomeadamente as de
 18625760 nsgás, concessionária da rede de
 18625774 dos, a concessionária da rede de
 18625823 ntem possível recristalização de
 18625767 hos de construção do gasoduto de
 18625795 erilizados por ebulição, vapor a
 18625788 iclónica, é uma área alongada de
 18625781 de humidade do ar, com baixa ou
 18625830 scarga um volume de ar (V-C), a
 18625809 o ao deslocamento dos núcleos de
 18625802 mudança rápida duma atmosfera de
 18625844 te sol a brilhar! A tão esperada
 18625837 os; - @lgases incandescentes@i a

*** 50 BAIKA (real:50) ***

pressões baixas 8.063

33 baixas pressões 1

31010500 as altas pressões à direita e as
 31010430 -se mais leve e sobe, originando
 31010444 s em altitude; ao subir, provoca

altas pressões
altas pressões,
altas pressões@b
altas pressões
altas pressões)
altas pressões
altas pressões
altas pressões
@baltas pressões
@baltas pressões@b.
altas pressões
altas pressões
altas pressões
altas pressões
altas pressões
altas pressões
altas pressões
altas pressões
altas pressões
altas pressões
@altas pressões
altas pressões;
altas pressões

pressões das altas
pressões mais altas
pressões nas altas

alta pressão;
alta pressão,
alta pressão
alta pressão
alta pressão
alta pressão,
alta pressão.
alta pressão
alta pressão
alta pressão
alta pressão
alta pressão
alta pressão.

baixas pressões
@bbaixas pressões@b
baixas pressões

no continente, o
 o ar desce e div
 (o valor decre
 oceânicas que ori
 onde chegam a ta
 para a esquerda e
 polares, de orige
 polares divergem
 polares@b) e ta
 Portanto, a
 provocam ventos c
 siberianas e as d
 subtropicais, é m
 subtropicais, fun
 subtropicais junt
 subtropicais. Na
 subtropicais, nom
 subtropicais@b. O
 subtropicais pela
 subtropicais que
 subtropicais@b,
 sugado para as b
 tropicais desloca

latitudes e v
 que o normal
 latitudes (ch

- Coombs @iet al. @
 a nova fonte dever
 do gás natural, que
 e, por isso, poster
 entre Setúbal/ Brag
 formalização ou su
 O @btalvegue@b, ou
 ou com mudança do c
 (P-2-), fica no vol
 para latitudes mais
 para uma atmosfera
 parecia que tinha v
 Trata-se de @bespe

à esquerda no h
 à superfície
 à superfície e a

31010640 intas: – uma em consequência das
 31010626 tuguês ficar sob a influência de
 31010605 está na presença de um centro de
 31010598 te no Inverno, nas zonas onde as
 31010584 este nas orientais. No Verão, as
 31010507 de oeste tem, à sua esquerda, as
 31010591 é quase constante. No Verão, as
 31010577 ta» proporcionando a invasão das
 31010542 do do zénite, arrasta consigo as
 31010521 nvergência, isto é, das zonas de
 31010437 pressões à superfície, mas cna
 31010549 são de tipo convectivo, pois as
 31010570 e altas pressões; sugado para as
 31010465 equador está assim envolvido por
 31010633 ações frontais. Estes centros de
 31010458 m nas latitudes subtropicais. As
 31010451 aparentemente, das altas para as
 31010493 ntram altas pressões na Itália e
 31010479 cal originado pela existência de
 31010423 esce e diverge e em situações de
 31010556 ue suplantam todos os outros. As
 31010528 quedas coincidem com as zonas de
 31010563 tas pressões na estação fria. As
 31010647 mmando uma cintura de centros de
 31010514 de @bconvergência@b ou zonas de
 31010612 rovocam a formação de centros de
 31010486 m o centro, especialmente quando
 31010472 os de leste@b que sopram para as
 31010535 lidade e da migração das altas e
 31010619 ores; – outra condicionada pelas

baixa pressão 4.929

10 baixa pressão 1

30995018 ste. Retorno por ar comprimido a
 30995039 m tubo de vidro contendo o gás a
 30994983 ponto de vista climatológico, a
 30995004 a-Aracena, com o metamorfismo de
 30994990 ta pressão para uma atmosfera de
 30995032 s@b. Também os gases e vapores a
 30995011 aconteceu com gradiente de mais
 30994976 5 mb correspondem a @bcentros de
 30994997 média (Barroviana) ao regime de
 30995025 ste. Retorno por ar comprimido a

*** 21 CEDER (real:21) ***

cedeu pressões 7.432

4 cedeu pressões 2

36403157 r Jorge perdeu dois jogos porque
 36403164 spot de provas de que a Comissão
 36403178 toconsumo é ainda forte. Ninguém

baixas pressões às quais se asso
baixas pressões associadas a per
baixas pressões centrado na Peni
baixas pressões conseguem pene
baixas pressões continentais são
baixas pressões das médias latit
baixas pressões de origem térmic
baixas pressões devastadoras. Os
baixas pressões, e daí as chuvas
baixas pressões, e por isso é es
baixas pressões, em altitude; ao
baixas pressões equatoriais, aco
baixas pressões equatoriais, aí
@bbaixas pressões equatoriais@b,
baixas pressões, localizados no
baixas pressões localizam-se nas
baixas pressões. Logo, o resulta
baixas pressões na Europa centra
baixas pressões no Mediterrâneo.
baixas pressões, o ar sobe e con
baixas pressões ocorrem na estaç
baixas pressões, onde se verific
baixas pressões, originam sempre
baixas pressões, ou de depressõe
baixas pressões. Quando estas co
baixas pressões. São estas @bdep
baixas pressões se localizam no
baixas pressões subpolares e que
baixas pressões. @t@bCLIMA: DEFI
baixas pressões térmicas da Pení

baixa pressão (0, 5 a 1kgf/ cm+
baixa pressão .Analisando a luz
baixa pressão atmosférica ocasio
baixa pressão .c) Geologia de Tr
baixa pressão e é a mais importa
baixa pressão emitem luz quando:
baixa pressão ,mas com temperatu
baixa pressão @b,ou @bcentros ci
baixa pressão ,para o intervalo
baixa pressão (por exemplo 0, 5k

cedeu a pressões e colocou em ca
“cedeu a pressões exteriores”. A
cedeu às pressões da autarquia,

36403171 ou por querer a conciliação, mas ceder pressões 7.149 9 ceder pressões 2	cedeu às pressões	de Pacheco Per
36400434 isco de ser visto como estando a 36400455 sidente do IBL acusa ministro de 36400448 ensão, acusando a CBI de estar a 36400427 isco de ser visto como estando a 36400441 orto, que o seu partido não deve 36400469 a, a nível governamental, está a 36400462 has e revolta-se, disposta a não 36400413 dia. «É por isso que não se pode 36400420 grama global, o Estado acaba por 3 cedido pressões 2	ceder a pressões. ceder a pressões ceder a pressões. ceder a pressões ceder às pressões ceder às "pressões ceder às pressões ceder às pressões ceder às pressões ceder às pressões	Caso não haja da Maçonaria Aj "O .que aqui es políticas, circ de alguns sect dos interesse dos juizes, do existentes e r locais", atalh
36405025 a, Manuel Maria Carrilho, de ter 36405032 Herzegovina parece ter sobretudo 36405039 das. Se tivéssemos sido fracos e cedendo pressão 6.737 3 cedendo pressão 2	cedido a pressões cedido às pressões cedido às pressões	da Maçonaria, do Presidente socialistas,
36397120 privada dos sistemas municipais, 36397127 caminho-de-ferro, nalguns casos, 36397134 caminho-de-ferro, nalguns casos	Cedendo à pressão Cedendo à pressão Cedendo à pressão	das multinacio dos operadores dos operadores

BACELAR DO NASCIMENTO, M. F. Portuguese Combinatory Dictionary (European variety). *Alfa (São Paulo)*, v.42, n.esp., p.183-203, 1998.

- **ABSTRACT:** *Taking a subcorpus of 12 million words from CRPC, the extraction of all pairs of contiguous words and the pairs with one, two, three or four words between the two words of the pair was made. All these pairs are provided with the following information: Frequency and Distribution of the pair; Mutual information: IC; Localization of the pair in the subcorpus; Concordances of the pair; Groups of 3, 4, 5 words in fixed sequence.*
- **KEYWORDS:** *Lexicon; lexicography; lexical combinatories.*

Referências bibliográficas

- BENSON, M., BENSON, E., ILSON, R. *The BBI combinatory dictionary of English: a guide to word combination*. Amsterdam: John Benjamins, 1986.
- BIBER, D. Co-occurrence patterns among collocations: a tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, v.19, n.3, p.531-38, 1993.

- CALZOLARI, N., BINDI, R. Acquisition of lexical information from a large textual italian corpus. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 13, 1990, Helsinki. *Coling 90: papers...* Helsinki: Helsinki University, 1990.
- CHURCH, K. W., HANKS, P. Word association norms, mutual information and lexicography. *Computational Linguistics*, v.16, n.1, p.22-9, 1990.
- PEREIRA, L. A. S. *Como se combinam as palavras? Contributo para um Dicionário de Combinatórias do Português*. Lisboa, 1994. Dissertação (Mestrado) – Faculdade de Letras, Universidade de Lisboa.
- SINCLAIR, J. M. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.
- _____. (Ed.) *Cobuild English collocations, on CD-ROM*. London: Harper Collins Publishers, 1995.
- SMADJA, F. Retrieving collocation from text: extract. *Computational Linguistics*, v.19, n.1, p.143-77, 1993.

Bibliografia consultada

- BACELAR DO NASCIMENTO, M. F. Aplicação de resultados de análises linguísticas sobre *corpora* ao ensino do Português, LE. In: CONGRESSO INTERNACIONAL DO ENSINO DE PORTUGUÊS COMO LÍNGUA ESTRANGEIRA, 3, 1994, México. *Memórias...* México: Centro de Estudos Brasileiros, Universidade Nacional Autónoma de México, 1994. p.104-20.
- _____. Contribuição da análise de *corpora* para a descrição lexicográfica. In: *Sentido que a vida faz: estudos para Óscar Lopes*. Porto: Campo das Letras, 1987. p.734-44.
- _____. A exploração de *corpora* linguísticos no ensino/aprendizagem do português. In: SEMINÁRIO INTERNACIONAL DE PORTUGUÊS COMO LÍNGUA ESTRANGEIRA, 1997, Macau. p.391-407.
- _____. Exploitation de *Corpus*: exemples de combinatoires établies pour l'écrit et pour l'oral. In: COLLOQUE INTERNATIONAL QUESTIONS DE MÉTHODE DANS LA LINGUISTIQUE SUR CORPUS, 1998. (No prelo).
- BACELAR DO NASCIMENTO, M. F., CARVALHO, A. Preto e branco ou branco e preto? Como se combinam os nomes de cores. In: ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE LINGÜÍSTICA, 11, 1996, Lisboa. *Actas...* Lisboa: APL, 1996. v.1, p.367-80.
- BACELAR DO NASCIMENTO, M. F. PEREIRA, L. A. S. Dicionário de Combinatórias do Português: associações frequentes observadas num *corpus* de Português contemporâneo. In: ENCONTRO NACIONAL DA ASSOCIA-

- ÇÃO PORTUGUESA DE LINGUÍSTICA, 11, 1996, Lisboa. *Actas...* Lisboa: APL, 1996. v.2, p.43-54.
- FIRTH, F. R. Modes of meaning. *PAPER in Linguistics 1934-1951*. London: Oxford University Press, 1995. p.190-215,
- GARCIA MARQUES, M. L. Da competência à realização: existência potencial e existência real de algumas combinatórias a partir da análise de um *corpus*. In: ENCONTRO DA ASSOCIAÇÃO PORTUGUESA DE LINGUÍSTICA, 13, 1997. *Actas...*
- NETO, P. M. *Combinatórias lexicais no discurso da astronomia: um estudo em estatística lexical*. Lisboa, 1995. Dissertação (Mestrado) – Faculdade de Letras, Universidade de Lisboa.
- PEREIRA, L. A. S. Para um Dicionário de Combinatórias do Português. In: CONGRESSO INTERNACIONAL SOBRE O PORTUGUÊS, 1996, Lisboa. *Actas...* Lisboa: APL, Colibri, 1996. p.197-206.
- RENOUF, A. J., SINCLAIR, J. M. Collocational frameworks in English. In: AIJMER, R., ALTENBERG, B. (Ed.) *English corpus linguistics*. London: Longman, 1991.