

# Combinatórias do Português: uma abordagem *corpus-driven* com fins lexicográficos

Sandra Antunes

Centro de Linguística da Universidade de Lisboa

Grupo de Linguística de *Corpus*

Av. Prof. Gama Pinto, 2 – 1649-003 Lisboa

sandra.antunes@clul.ul.pt

## Resumo

A dificuldade encontrada em delimitar os diferentes tipos de associações de palavras, bem como o estatuto incerto que estas ocupam na gramática conduziram a uma grande variação nas análises realizadas até à data por vários autores. Dada a natureza multifacetada deste fenómeno, a sua definição e categorização só parecem ser possíveis através da combinação de critérios como a frequência (facilmente obtida através dos dados do *corpus*) e a análise intensiva de todas as suas propriedades linguísticas. Neste artigo será possível observar, de um modo sucinto, o método de extracção e selecção das combinatórias do *corpus* (secção 2), as análises e tipologias de alguns dos autores mais proeminentes na área (secção 3) e a sua adaptação a alguns dados analisados até à data (secção 4).

## 1 Introdução

Cada vez mais os investigadores se têm dado conta de que o recurso à intuição e à introspecção de um falante não é suficiente para analisar e descrever as associações de palavras de uma língua. A grande dificuldade encontrada pelos lexicógrafos em identificar e definir as combinatórias usuais que devem constar num dicionário, recorrendo apenas à intuição, tem contribuído para que elas sejam, em geral, negligenciadas na maior parte dos dicionários portugueses. Contudo, o recurso a *corpora* permite, de um modo mais fácil, identificar os diversos padrões de associações de palavras e, conseqüentemente, proceder a uma selecção mais criteriosa. Com base num *corpus* equilibrado com mais de 50 milhões de palavras, o trabalho em curso tem como objectivos a detecção, análise e definição das principais combinatórias do português europeu, o estabelecimento de uma tipologia e uma proposta de selecção e tratamento destas associações de palavras nos dicionários de português.

## 2 Extracção e selecção das combinatórias

Para este estudo foi utilizado um *corpus* equilibrado composto por 50 milhões de palavras no registo escrito<sup>1</sup> – jornal (30.000.000), livro (10.917.889), revista (7.500.000), vária (1.851.828), direito (313.962), política (277.586) e folheto (104.889) – e 1 milhão de palavras no registo oral – informal (912.581), formal (528.187) e conversas telefónicas (24.365).

Para a extracção das combinatórias do *corpus* foi aplicado, em UNIX, um programa informático que permite extrair todos os grupos compostos por 2 a 5 palavras, bem como as concordâncias de cada grupo. A aplicação da medida de associação lexical *Informação Mútua* (Church & Hanks, 1989) – que calcula as probabilidades de ocorrência conjunta dos grupos de palavras, no *corpus*, e de ocorrência independente dessas mesmas palavras – permite evidenciar os grupos mais significativos.

De forma a facilitar a visualização e a selecção das associações de palavras significativas, foi criada uma base de dados, com plataforma SQL e interface em Access, que permite importar automaticamente todos os resultados obtidos através da aplicação do programa informático referido acima.

As combinatórias seleccionadas foram, posteriormente, organizadas de forma a identificar um lema de grupo (que permite reunir, sob uma única forma, as variantes flexionais que ocorreram no *corpus*) e um lema principal (que corresponde ao lema a partir do qual a combinatória foi seleccionada), como se pode observar na tabela 1.

---

LEMA

**fogo**

---

LEMA DE GRUPO

**fogo de artifício**

---

GRUPO

**fogo de artifício**

---

<sup>1</sup> Corpus constituído no âmbito do projecto *COMBINA-PT – Combinatórias Lexicais do Português*, desenvolvido no Centro de Linguística da Universidade de Lisboa pelo grupo de Linguística de Corpus (<http://www.clul.ul.pt>).

não faltou até às três horas, o **fogo de artifício** surpreendeu os novos doze meses e os cachos de **fogo de artifício** transfiguram u contará com uma grande sessão de **fogo de artifício**, a marcar a en costura parisiense. Perante este **fogo de artifício** sentimos por v

GRUPO

#### **fogos de artifício**

de moagem, cerâmica, curtumes e **fogos de artifício**, merece ainda a pelos Chineses, sobretudo para **fogos de artifício**. A pólvora pa aca impediu que 156 toneladas de **fogos de artifício** fossem lançad amariz e na Baía de Cascais. Com **fogos de artifício**, animação de

**Tabela 1.** Organização dos lemas

### 3 Algumas considerações teóricas

Nesselhauf (2004:1) refere que, provavelmente, existem tantas respostas à pergunta ‘o que é uma combinatória?’ quantos os autores que escrevem sobre o assunto. Na verdade, no seguimento de Bartsch (2004:27-28), a existência de uma tão grande variação no uso deste termo pode estar relacionada com:

(i) a dificuldade encontrada em delimitar os diferentes tipos de combinações de palavras;

(ii) o facto de as combinatórias consistirem num fenómeno linguístico que parece situar-se na fronteira entre a gramática e o léxico (o que dificulta a elaboração de uma definição formal e sistemática do ponto de vista da teoria linguística, uma vez que parece não ser possível uma explicação baseada unicamente em princípios sintácticos e semânticos já estabelecidos);

(iii) o facto de existirem diferentes abordagens deste fenómeno, nomeadamente a abordagem baseada em frequências e a abordagem fraseológica.

No âmbito destas duas abordagens foi proposto um vasto número de definições e de tipologias, que se discutirão nas secções 3.1 e 3.2.

#### 3.1 A abordagem baseada em frequências

A abordagem baseada em frequências foi inicialmente desenvolvida por J. R. Firth. Ainda que possa não ter sido o pioneiro no uso do termo combinatória (com o significado de combinação frequente e relativamente fixa de palavras), Firth foi, certamente, o autor que o definiu do ponto de vista teórico e que mais o divulgou, contribuindo, assim, para o aumento do interesse pelo estudo das relações sintagmáticas.

De um modo geral, pode dizer-se que esta abordagem encara as combinatórias como um fenómeno evidenciado pela estatística, definindo-as como uma combinação frequente e relativamente fixa de palavras, que podem ocorrer a uma certa distância. Contudo, a observação das análises efectuadas por diferentes autores mostram a existência de alguma variação no que respeita a determinados aspectos.

#### **Combinatórias e Coligações**

Ainda que muitos autores usem indiscriminadamente o termo combinatória para

referir qualquer tipo de combinação frequente de palavras, também é possível encontrar análises que fazem uma distinção a nível lexical e gramatical.

A falta de uma teoria linguística que conseguisse descrever adequadamente o fenómeno das combinatórias conduziu Firth (1957) à sugestão da introdução de um nível combinatorial. Deste modo, ainda que o evento da linguagem deva ser considerado como um todo, o autor propõe que, do ponto de vista da descrição linguística, ele assente em quatro níveis de significado (*levels of meaning*): (i) contexto situacional; (ii) combinatórias; (iii) sintaxe; (iv) fonologia e fonética. O autor reconhece, assim, que a combinação de palavras apresenta restrições não só a nível gramatical, mas também a nível combinatorial e faz a distinção entre coligações – que correspondem a combinações de palavras sintacticamente restringidas do ponto de vista da subcategorização (*afraid of*) – e combinatórias – que correspondem a co-ocorrências significativas de determinadas palavras com uma certa proximidade.

Uma distinção semelhante é também adoptada por Sinclair (1991).

#### **Distância**

No que respeita à distância a que as palavras que constituem uma combinatória se podem encontrar, também é possível encontrar alguma variação.

Firth (1957) não faz nenhuma referência à distância, classificando como combinatórias quer associações contíguas de palavras (*hold life in contempt*), quer palavras que podem não ocorrer contiguamente (como *dark* e *night* – *one dark and cold night; the night is dark*).

Mitchel (1971) defende que uma combinação de palavras pode ultrapassar fronteiras fráscas, dando como exemplo: *He didn't want the job. I don't think he even applied*.

Finalmente, (Sinclair, 1991:170) é o único autor que define objectivamente a distância a que as palavras podem ocorrer: “The usual measure of proximity is a maximum of four words intervening”.

#### **Lexemas vs. palavras derivadas**

O tipo de elemento que constitui a combinatória é outra questão que também pode variar de autor para autor. Enquanto que Firth (1957) dá exemplos diferentes de combinatórias para cada forma flexionada da palavra *get* (*gets, got, getting*), sugerindo que para cada forma diferente existe uma combinatória diferente, Haliday (1961) e Mitchel (1971) defendem que uma combinatória é constituída pelo lexema (ou raiz) e por todas as suas formas derivadas (expressões como *a strong*

*argument, he argued strongly, the strength of the argument* e *his argument was strengthened* são consideradas instâncias da mesma combinatória).

### Frequência

Ainda que esta abordagem encare as combinatórias como uma combinação frequente de palavras, alguns autores, como Firth (1957) e Sinclair (1991) fazem a distinção entre combinatórias habituais ou significativas (se ocorrerem mais frequentemente) e combinatórias idiossincráticas ou casuais (se ocorrerem mais esporadicamente). Também Greenbaum (1970) propõe a distinção entre  *collocability* (para potenciais coocorrências de palavras) e  *collocation* (para expressões que coocorrem frequentemente).

O recurso à análise de textos reais da língua, defendido por Sinclair, permite não só obter facilmente dados sobre a frequência de uma determinada combinação de palavras, como também observar que diferenças mínimas no significado de uma palavra (incluindo sinónimos quase equivalentes) correspondem, normalmente, a diferentes padrões combinatoriais.

### Relações Sintagmáticas e Semânticas

Ainda que Firth (1957) considerasse o estudo lexical separado da gramática, os seus seguidores, nomeadamente Haliday (1961), Mitchel (1971) e Greenbaum (1970) começam a dar prevalência às relações sintagmáticas entre os itens lexicais e a reconhecer a interdependência entre a gramática e o léxico. Ainda que considere que uma combinatória é constituída por raízes, Mitchel (1971:52) realça o facto de a análise gramatical ter sido fundamental para a descoberta de que nem todas as palavras de uma combinatória ocorrem em todos os padrões morfossintáticos teoricamente possíveis: “*heavy damage* is possible as well as *damage heavily* and *heavily damaged*, but not \**heavy damaging*, whereas *heavy drinking* is possible as well as *drink heavily*, but not \**heavily drunk*”. Apesar de considerar as relações gramaticais como as mais importantes, este autor destaca-se por ter também em atenção as relações semânticas no estudo das combinatórias, tentando estabelecer um  *continuum* entre combinatórias, compostos e expressões idiomáticas.

Também Greenbaum (1970:11) defende uma abordagem integrada, uma vez que também sugere que o estudo das combinatórias deve ter em conta quer os factores semânticos, quer, principalmente, os sintáticos: “for an analysis of collocations that is divorced from syntax, (...) it does not seem possible to establish a criterion for determining whether two items are collocating”.

### O Princípio Idiomático

De todos os autores que trabalharam nesta abordagem, Sinclair foi aquele que mais anos dedicou ao estudo das combinatórias. (Sinclair, 1991:115) estabelece dois princípios organizadores da língua, simultaneamente alternativos e complementares, a partir dos quais é possível interpretar o significado das palavras: (i) o princípio da livre escolha, em que o falante tem como única restrição a gramaticalidade do enunciado; (ii) o princípio idiomático, em que o falante tem à sua disposição um grande número de grupos de palavras pré-construídos (ainda que possam apresentar alguma variação, nomeadamente a nível lexical, flexional ou de ordem das palavras). Efectivamente, tem-se considerado que o falante utiliza as capacidades de memória e as rotinas, sendo os seus discursos preferencialmente constituídos pelas escolhas correspondentes ao princípio idiomático. E, para o autor: “Collocation (...) illustrates the idiom principle”.

### 3.2 A abordagem fraseológica

A abordagem fraseológica, inicialmente desenvolvida pelos estudiosos russos, considera as combinatórias como um tipo particular de unidade fraseológica. Esta abordagem propõe uma descrição sistemática da categorização das unidades fraseológicas elaborada a partir de uma escala de idiomaticidade. Contudo, o facto de as fronteiras entre as diversas categorias não serem bem delimitadas (existindo princípios unificadores entre elas), implica que a categorização das unidades fraseológicas seja elaborada com base num  *continuum*. Pelo que se pode observar através das diferentes análises, os principais princípios unificadores parecem ser, por um lado, o desvio (maior ou menor) do significado totalmente composicional (que implica uma decrescente “analísabilidade”) e, por outro, uma crescente fixidez das expressões (incluindo aquelas que têm significado composicional). Este  *continuum* é baseado numa perspectiva sincrónica e parece representar os graus de estabilidade e de institucionalização com os quais as expressões estão armazenadas no léxico mental.

Nesta abordagem, é possível distinguir quatro grandes categorias, como se pode observar na tabela 2, adaptada de Poulsen (2005:58). Note-se, contudo, que este modelo-base de categorização pode ser interpretado e completado de modo diferente pelos diversos autores.

CATEGORIA	COMPOSICIONAL	ANALISÁVEL	COMBINAÇÃO FRASEOLÓGICA
Associações Livres	Sim	Sim	Não
Combinatórias	Parcialmente	Sim	Sim
Idiomas	Não	Sim	Sim

Figurativos			
Idiomas Puros	Não	Não	Sim

**Tabela 2.** Categorização geral das unidades fraseológicas

As associações livres correspondem a expressões com significado totalmente composicional, onde os seus elementos se podem combinar livremente com outros elementos (*drink tea/water/wine*). Normalmente não são consideradas unidades fraseológicas (nem devem estar listadas nos dicionários (Benson *et alii*, 1986a)).

As combinatórias são normalmente entendidas nesta abordagem como expressões que têm um significado parcialmente composicional. De acordo com a maior parte dos autores, pelo menos uma das palavras que compõem a combinação é livremente escolhida com base no seu significado literal, enquanto que pelo menos outra apresenta um significado figurativo, resultante da combinação. Normalmente, a palavra com significado figurativo não pode ser substituída por

um sinónimo (*heavy smoker/\*weighty smoker; strong coffee/\*powerfull coffee;*).

Os idiomas figurativos correspondem a expressões que têm tanto um significado figurativo, como um significado composicional (*to catch fire*). Note-se, contudo, que o reconhecimento do significado literal pode estar dependente não só do conhecimento linguístico do falante, como de qualquer outro factor externo (idade, cultura, experiência de vida, etc.).

Os idiomas puros correspondem a expressões com significado totalmente idiomático, uma vez que não pode ser calculado através do significado dos elementos que as compõem (*spill the beans*).

A tabela 3, em baixo, ilustra as categorizações das combinações de palavras defendidas por alguns dos autores mais proeminentes desta abordagem (bem como os diferentes termos utilizados) e a variação que apresentam em relação ao modelo-base.

MODELO-BASE	COWIE (1994)	MEL'ČUK (1996)	BENSON ET ALII (1986b)	HAUSMANN (1989)
ASSOCIAÇÕES LIVRES	<i>open collocations</i>	<i>free combinations</i>	<i>free combinations</i>	<i>co-creation/free combinations</i>
COMBINATÓRIAS	<i>restricted collocations</i>	<i>semi-phraseemes/collocations</i>	<ul style="list-style-type: none"> <li>• <i>collocations</i></li> <li>• <i>transitional collocations</i></li> </ul>	<i>collocations/affines</i>
IDIOMAS FIGURATIVOS	<i>figurative idioms</i>			
IDIOMAS PUROS	<i>pure idioms</i>	<i>full phraseemes/idioms</i>	<i>idioms</i>	<i>idioms</i>

**Tabela 3.** Categorização das unidades fraseológicas de acordo com alguns autores

Como se pode observar, o modelo-base representa na perfeição a categorização de Cowie (1994). Note-se, no entanto, que o autor faz ainda uma distinção mais geral entre o que denomina *formulae* (combinações de palavras ao nível da frase, com função pragmática – *how are you?*) e *composites* (combinações de palavras abaixo do nível da frase, com função léxico-sintáctica). As combinações de palavras apresentadas na tabela 3 correspondem a diferentes tipos de *composites*.

A par de Cowie (1994), também Mel'čuk (1996) faz uma distinção mais geral entre o que denomina *pragmatemes* (combinações de palavras semanticamente composicionais mas pragmaticamente restringidas, uma vez que envolvem restrições combinatoriais – pode ocorrer *no talking please*, mas não *\*keep silent please*) e *semantic phraseemes* (expressões cujo significado não é composicional). As combinações de palavras apresentadas na tabela 3 correspondem a diferentes tipos de *semantic phraseemes*.

Como se pode observar, o autor não prevê os idiomas figurativos na sua categorização. Existe, no entanto, outro tipo de unidade fraseológica denominada *quasi-phraseeme/quasi-idiom*. Contudo, esta categoria não parece diferenciar-se muito das combinatórias, uma vez que, segundo o autor, também corresponde a combinações de palavras onde o significado pode ser somente

parcialmente obtido através dos elementos que as compõem (*start a family*).

Inerente à proposta de classificação das combinações de palavras de Mel'čuk encontra-se a sua teoria de Funções Lexicais (LFs), que pretende captar as relações existentes entre os seus constituintes em diferentes níveis de descrição linguística. Para qualquer unidade lexical, as LFs pretendem especificar todos os seus potenciais coocorrentes (cf. Mel'čuk, 1984).

No que respeita a Benson *et alii* (1986a,1986b), como se pode ver na tabela 3, as definições de associações livres e de expressões idiomáticas correspondem às apresentadas para o modelo-base. Contudo, no âmbito da classe das combinatórias, baseando-se em critérios de coesão e frequência, são distinguidas *collocations* de *transitional collocations*. O primeiro caso diz respeito a expressões frequentes que apresentam um certo grau de coesão, como *warmest regards* e *commit murder*. No caso de *warmest regards* pode considerar-se que pode ser incluído na classe de combinatórias defendida por Cowie e Mel'čuk (uma vez que tem um significado parcialmente composicional, não sendo possível substituir *warmest* por expressões equivalentes – *\*hot regards; \*hearty regards*). Contudo, no caso de *commit murder*, apesar de ser uma expressão com significado composicional, os autores consideram

que também apresentam um certo grau de coesão, uma vez que *commit* apresenta restrições relativamente aos seus coocorrentes (selecciona somente nomes com significado de ‘crime’ ou ‘transgressão’). As *transitional collocations* correspondem a expressões com mais ‘transparência’ semântica do que as expressões idiomáticas, mas com menos possibilidade de variação do que as combinatórias.

São, ainda, distinguidas as combinatórias lexicais (que, por sua vez, se subdividem de acordo com as classes gramaticais dos elementos que as compõem) das combinatórias gramaticais (que, por sua vez, se subdividem de acordo com os padrões gramaticais dos seus elementos).

Relativamente a Hausmann (1989), para além de considerar a existência de combinação de palavras com um certo grau de convenção e restrição (*faire une promenade*) – que se enquadram no modelo-base – este autor também considera a existência de um tipo particular de combinação, denominada *counter-creations/counter-affines* e que corresponde à coocorrência de uma palavra com outras com as quais normalmente não se combina (*la route se rabougrit*).

Do ponto de vista lexicográfico, defendendo que as combinatórias devem ser inseridas debaixo do constituinte que constitui o tópico da combinação, o autor considera que estas são constituídas por uma base (que corresponde ao elemento que é semanticamente autónomo e seleccionado em primeiro lugar pelo falante) e por um coocorrente (que corresponde ao elemento que é seleccionado em função da base e que só ao combinar-se com esta é que recebe a sua “identidade semântica”, ou seja, o seu significado exacto – em *pack of dogs* e *cigarette pack*, o coocorrente *pack* é um constituinte polissémico que apresenta diferentes significados de acordo com as bases a que se junta). A base é, assim, o constituinte dominante e aquele que apresentará maior interesse na consulta de um dicionário.

#### 4 Trabalho em curso

As diferentes análises apresentadas ilustram um pouco a variação existente entre os diversos autores na distinção dos diferentes tipos de combinações de palavras (o termo combinatória é, inclusivamente, várias vezes utilizado para referir combinações de palavras distintas). Tendo em conta a natureza multifacetada deste fenómeno, a sua análise só parece ser possível através da combinação de critérios como a frequência e a análise intensiva de todas as suas propriedades linguísticas (nomeadamente sintácticas (fixidez), semânticas (idiomaticidade), fonológicas (com influência na ordem das palavras), lexicais (tipo de

constituintes e paradigmas em que se inserem – hiperonímia, homonímia, sinonímia, etc.), gramaticais (co-ocorrência sistemática com determinado tipo de elemento ou construção), distribucionais (co-ocorrência sistemática de uma palavra com outras de forma contínua ou descontínua), discursivas (associações que ocorrem mais num determinado registo de língua do que noutra) e pragmáticas.

Deste modo, do ponto de vista léxico-sintáctico, os dados do *corpus* analisados até à data mostraram uma grande variação a nível sintáctico e lexical. Na verdade, a ocorrência de expressões totalmente fixas não é tão frequente quanto se poderia esperar. Mesmo no que respeita a expressões aforísticas (supostamente as mais rígidas do espectro), é possível encontrar variação. A criatividade lexical dos falantes resulta na ocorrência de expressões como as encontradas no *corpus* e apresentadas em (1).

(1)	no	poupar	é que está o ganho	(4 ocorr.)
	no	anunciar	é que está o ganho	(1 ocorr.)
	no	atacar	é que está o ganho	(1 ocorr.)
	no	descontar	é que está o ganho	(1 ocorr.)
	no	prejuízo	é que está o ganho	(1 ocorr.)
	no	esperar	é que está o ganho	(1 ocorr.)
	no	cooperar	é que está o ganho	(1 ocorr.)
	no	comparar	é que está o ganho	(1 ocorr.)
	no	economizar	é que está o ganho	(1 ocorr.)
	no	provar	é que está o ganho	(1 ocorr.)

Como também se pode observar nos exemplos em (1), a expressão *no poupar é que está o ganho* é a que tem, de facto, o maior número de ocorrências (4), continuando a ser reconhecida como a expressão institucionalizada que os falantes têm armazenada no seu léxico mental. Contudo, apesar da sua institucionalização e reconhecida fixidez, é possível “brincar” a nível lexical, construindo expressões semanticamente equivalentes.

Os tipos de variação mais frequente correspondem a fenómenos de: (i) flexão (*esfregavam/esfregou as mãos de contentamento*); (ii) passivização (*correr riscos/foram corridos riscos*); (iii) relativização (*correr riscos/los riscos que correm*); (iv) nominalização (*argumento fortela força do argumento*); (v) permuta (*pôr em causa [algo]/pôr [algo] em causa*); (vi) possessivação (*estar nas mãos de [alguém]/estar nas [poss] mãos*); (vii) inserção de artigos com contracção de preposições (*estar atento a/estamos atentos a/estivemos atentos àquela*); (viii) inserção lexical (*alvo de críticas/alvo de fortes críticas*); (ix) variação lexical (*onda/maré/vaga de consternação*).

Do ponto de vista semântico a identificação de uma expressão idiomática (cujo significado não pode ser calculado através dos significados dos seus constituintes) nem sempre é uma tarefa tão linear quanto pode parecer. Ainda que se defenda

que estas expressões obedecem a certas propriedades (opacidade e “não-analisabilidade” semânticas, fixidez e institucionalização), na verdade, estas podem apresentar alguma variação. A interpretação de uma determinada expressão pode variar de falante para falante, por influência de vários factores (enquanto que muitos falantes considerarão a expressão *esticar o pernil* totalmente opaca, outros identificarão facilmente o seu significado literal). Uma vez que o grau de lexicalização e de institucionalização de uma expressão resulta de um processo gradual, é possível encontrar uma determinada combinação de palavras com diferentes graus de coesão. Neste sentido, a expressão *fazer a cama* pode ser: (i) uma combinação livre, com significado composicional (*construir a cama*); (ii) uma expressão fixa, institucionalizada, mas ainda com significado composicional (*arranjar a cama*); (iii) uma expressão altamente lexicalizada, com significado não-composicional (*tramar alguém*).

No que respeita ao critério da frequência, ainda que ignorado por alguns autores, parece que não pode ser deixado de parte, uma vez que se torna essencial na identificação de certas associações de palavras, como as associações privilegiadas – expressões cujo significado é totalmente composicional, mas que se tornaram estilisticamente idiossincráticas e preferíveis a outras, igualmente possíveis no mesmo contexto (*trabalho árduo, ar fresco* ou *chuva torrencial*).

Através deste critério é possível, também, observar que certas expressões apresentam uma preferência na ordem dos seus constituintes (*fixos e móveis, públicas e privadas, preto e branco*).

O recurso a dados do *corpus* fornece, assim, informação importante para a análise das associações de palavras, uma vez que torna visíveis certos aspectos que, de outro modo, poderiam passar despercebidos (como a variação lexical de expressões, como em (1), que, à primeira vista, seriam consideradas como totalmente fixas).

## 5 Trabalho futuro

Após a análise das principais associações de palavras e do estabelecimento de uma tipologia, este trabalho tem também como objectivo criar uma proposta de selecção e de tratamento destas associações nos dicionários de português. Para além da área da lexicografia, os resultados deste estudo poderão constituir, ainda, uma fonte de informação importante para áreas como a psicolinguística, a linguística computacional, a tradução e a didáctica do português.

## 6 Agradecimentos

Estudo financiado pela Fundação para a Ciência e a Tecnologia (SFRH/BD/24905/2005).

## Referências

- A. P. Cowie. 1994. “Phraseology”. In Asher, R. E. (ed.). *The Encyclopedia of Language and Linguistics*, Pergamon. Oxford. pp. 3168-3171.
- F. J. Hausmann. 1989. “Le dictionnaire de collocations”. In F. J. Hausmann, et alii (eds.). *Wörterbücher, dictionaries, dictionnaires. Ein internationaler Handbuch zur Lexikographie*. de Gruyter. Berlin. pp. 1010-1019.
- I. Mel’čuk. 1984. *Dictionnaire explicatif et combinatoire du français contemporain*. Les Presses de L’Université de Montreal. Canada.
- I. MEL’ČUK. 1996. “Lexical functions: A tool for the description of lexical relations in a lexicon” In L. Wanner, (ed.). *Lexical Functions in Lexicography and Natural Language Processing. Studies in Language Companion Series*. John Benjamins. Amsterdam/Philadelphia. pp. 37-102.
- J. Firth. 1957. “A Synopsis of Linguistics Theory 1930-1955”. *Studies in Linguistics Analysis*. Oxford Philological Society.
- J. Sinclair. 1991. *Corpus, Concordance and Collocation*. Oxford University Press. Oxford.
- K. W. Church & P. Hanks. 1989. “Word Association Norms, Mutual Information and Lexicography”. *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*. Vancouver. Canada. pp. 76-83.
- M. A. K. Halliday. 1961. “Categories of the theory of grammar”. *Word* 17. pp. 241-292.
- M. E. Benson et alii. 1986a. *The BBI Combinatory Dictionary of English: a guide to word combination*. John Benjamins. Amsterdam/Philadelphia.
- M. E. Benson et alii. 1986b. *Lexicographic Description of English*. John Benjamins. Amsterdam/Philadelphia.
- N. Nesselhauf. 2004. “What are collocations?”. In D. J. Allerton et alii (eds.). *Phraseological Units: basic concepts and their application*. International Cooper Series in English Language and Literature. vol. 8. Schwabe Verlag Basel. Switzerland. pp. 1-21.
- S. Bartsch. 2004. *Structural and Functional Properties of Collocations in English. A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Gunter Narr Verlag Tübingen. Germany.
- S. GREEBAUM. 1970. *Verb-Intensifier Collocations in English. An Experimental Approach*. Mouton.
- S. Poulsen. 2005. *Collocations as a language resource. A functional and cognitive study in English phraseology*. Ph.D. Dissertation. Denmark.
- T. F. Mitchell. 1971. “Linguistic ‘goings-on’: collocations and other lexical matters arising on the syntactic record”. *Archivum Linguisticum* 2. pp. 35-69.