



Universidade do Minho
Instituto de Letras e Ciências Humanas

Mariana Fernandes Barbosa

**A potencialidade didática das colocações
lexicais - Um dicionário monolíngue de
colocações para alunos de PLE**

**A potencialidade didática das colocações lexicais - Um dicionário
monolíngue de colocações para alunos de PLE**

Mariana Fernandes Barbosa

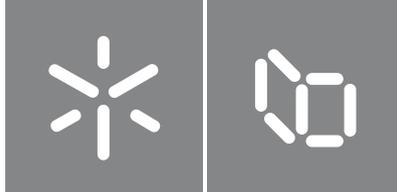
UMinho | 2021



With the support of the
Erasmus+ Programme
of the European Union



junho de 2021



Universidade do Minho

Instituto de Letras e Ciências Humanas

Mariana Fernandes Barbosa

**A potencialidade didática das colocações
lexicais - Um dicionário monolíngue de
colocações para alunos de PLE**

Dissertação de Mestrado
Mestrado Europeu em Lexicografia

Trabalho efetuado sob a orientação do
Professor Doutor Álvaro Iriarte Sanromán

Direitos de Autor e Condições de Utilização do Trabalho por Terceiros

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.



Este trabalho está licenciado com uma Licença Creative Commons - Atribuição-NãoComercial-CompartilhaIgual 4.0 Internacional.

Declaração de Integridade

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Agradecimentos

Primeiramente, gostaria de agradecer ao Programa EMLex por me presentear, ao longo destes dois últimos anos, uma experiência valiosa e enriquecedora ao nível interpessoal, cultural e didático, que jamais irei esquecer.

Agradeço a todos os meus amigos que me apoiaram incessantemente durante o desenvolvimento dos meus estudos e no decorrer da pesquisa e produção desta dissertação. Em particular, um obrigado especial ao Carlos, ao João, à Tânia, à Beatriz e à Fernanda por todo o carinho. De igual modo, agradeço aos meus pais e à minha Selma pelo amor e sacrifício.

Finalmente, pelo auxílio exemplar que me proporcionaram durante a formulação do meu trabalho lexicográfico, apresento a minha sincera gratidão ao Prof.º Dr.º Álvaro Iriarte Sanromán e à Prof.ª Dr.ª Idalete Maria Silva Dias. Obrigado pela paciência e pela sabedoria.

Resumo

Numa primeira fase, esta dissertação preocupa-se em explorar a teoria lexicográfica intrínseca ao estudo das colocações, sendo que pretendemos analisar o assunto de forma conceptual e oferecer uma via esclarecedora à problemática associada à representação especializada da combinatória em dicionários.

Seguidamente, planeamos contextualizar o potencial das colocações na esfera pedagógica do Português Língua Estrangeira (PLE). Reconhecemos que alunos de língua segunda têm a natural dificuldade na perceção deste fenómeno linguístico. Assim sendo, achamos oportuna a construção de uma ferramenta lexicográfica que ofereça apoio à produção de PLE.

Como objetivo principal, baseando-se numa literatura pertinente, este trabalho procura apresentar um planeamento metodológico a favor da compilação de um dicionário monolíngue especializado em colocações portuguesas, destinado a alunos de PLE. Acreditamos que esta ferramenta em particular pode ser essencial para o mercado lexicográfico, remetendo-se ao contexto da didática do léxico português.

Desta forma, fundamentamos o nosso estudo através de exemplos especializados do ramo, onde prestamos atenção não só à análise teórica das colocações, mas também à construção pormenorizada de um dicionário a seu favor.

Em suma pretendemos demonstrar a componente associativa do projeto descrito em relação ao ensino de PLE, de forma a contribuir para o impacto deste tipo de ferramenta especializada no campo da Lexicografia e da didática portuguesa.

Palavras-chave: colocação lexical, Português Língua Estrangeira, dicionário monolíngue, metodologia

Abstract

Firstly, this dissertation concerns with the exploration of the lexicographical theory inherent to the study of collocations, as we intend to analyse the subject in conceptual terms and shine a light onto the problems associated with the specialized representation of the combinatory in dictionaries.

Adjacently, we plan on contextualising the potential of collocations in the pedagogy of Portuguese as a Foreign Language (PFL). We recognize that second language students have a natural difficulty in understanding this linguistic phenomenon. Therefore, we believe that it is imperative to build a lexicographical tool that supports the production of PFL.

As its main objective, based on the established literature, this work aims to present a methodological plan in favour of the compilation of a monolingual dictionary specialized in Portuguese collocations, aimed at PFL students. We believe that this lexicographic tool can be essential for the dictionary market, in reference to the context of the teaching of the Portuguese lexicon.

Therefore, we base our study in particular examples of the field, taking into account not only the theoretical analysis of collocations, but also the detailed construction of a dictionary in its favour.

In short, we intend to showcase the associative component of our project in relation to the discipline of PFL in order to contribute for the impact of this type of specialized tool in the sphere of Lexicography and of Portuguese didactics.

Keywords: lexical collocation, Portuguese as a Foreign Language, monolingual dictionary, methodology

Índice

Agradecimentos	iv
Resumo	v
Abstract	vi
Introdução	1
1. O conceito de colocação	3
1.1. Expressões idiomáticas <i>versus</i> colocações	3
1.2. A identificação das colocações	6
2. Revisão literária	8
2.1. Firth	8
2.2. Halliday	11
2.3. Sinclair	16
2.4. Cowie	21
2.5. Hausmann	24
2.6. Corpas Pastor	26
2.7. Porzig	27
2.8. Coseriu	28
2.9. Mel'čuk	30
2.10. Benson	36
3. O ensino e tratamento das colocações. Metodologia para um dicionário monolíngue de colocações portuguesas	42
3.1. A importância das colocações para a didática do Português Língua Estrangeira	42
3.2. Conceções gerais da metodologia	47
3.2.1. Corpora e software de extração	52
3.2.2. Macroestrutura	54
3.2.3. Microestrutura	56
3.3. Considerações finais	61
Conclusão	63
Bibliografia	65
Webgrafia	71
Anexos	72

Lista de Figuras

- Figura 1** – Tabela representativa da noção de *frozenness hierarchy* (Fraser, 1970). Retirada de Wulff (2010, p. 69).
- Figura 2** – Gráfico representativo do *lexicogrammar cline*. Retirado de Halliday (2014, p. 64).
- Figura 3** – Excerto representativo das relações sintagmáticas e lexicais das colocações. Retirado de Halliday (2014, p. 59).
- Figura 4** – Tabela demonstrativa das dimensões da linguagem e dos seus princípios de ordem. Retirada de Halliday (2014, p. 20).
- Figura 5** – O *cline* de instanciação. Retirado de Halliday (2014, p. 28).
- Figura 6** – Esquema representativo da estratificação do léxico. Retirado de Halliday (2014, p. 26).
- Figura 7** – Gráfico representativo da classificação de combinações de palavras (Cowie, 1988). Retirado de Pereira da Costa (2017, p. 22).
- Figura 8** – Excerto do processamento do método NVELE (Spencer, 1975). Retirado de Cowie (1985, p. 67).
- Figura 9** – Esquema ilustrativo das estruturas lexicais (Coseriu, 1970). Retirado de Lipka (1992, p. 11).
- Figura 10** – Classificação das combinações (Mel'čuk, 1998). Retirado de Pereira da Costa (2017, p. 39).
- Figura 11** – Abreviações e anotações da simbologia usada na versão do DECFC para o léxico inglês (ECD). Retiradas de Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 3).
- Figura 12** – Subentrada de “walk a dog”. Retirada da versão do léxico inglês (ECD) do DECFC Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 9).
- Figura 13** – Artigo do lema *assignment*. Retirado do BBI (Benson *et al.*, 2010, p. 22).
- Figura 14** – Artigo do lema *about-turn*. Retirado do BBI (Benson *et al.*, 2010, p. 1).
- Figura 15** – Artigo do lema *able*. Retirado do BBI (Benson *et al.*, 2010, p. 1).
- Figura 16** – Artigo do lema *act up*. Retirado do BBI (Benson *et al.*, 2010, p. 5).
- Figura 17** – Artigo do lema *bush*. Retirado do BBI (Benson *et al.*, 2010, p. 52).
- Figura 18** – Artigos dos lemas *bow I*, *bow II*, *bow III*, e *bow down*. Retirados do BBI (Benson *et al.*, 2010, p. 45).

Figura 19 – Gráfico temático relativo à palavra *tangi* do léxico maori. Retirado de Thornbury (2002, p. 16).

Figura 20 – Esquema ilustrativo dos fatores a serem considerados para a escolha de lemas. Retirado de Atkins (2008, p. 178).

Figura 21 – Tabela dos nove estímulos utilizados para favorecer a produção textual de alunos de PLE. Retirada de <https://www.uc.pt/fluc/rcpl2/metodologia> (Universidade de Coimbra, 2008).

Figura 22 – Excerto que exemplifica a alfabetização, o *nitching* e o *nesting* dos artigos de dicionário. Retirado do BBI (Benson *et al.*, 2010, p. 1).

Figura 23 – Artigos dos lemas *acquaintanceship*, *acquainted* e *acquiesce*. Retirados do BBI (Benson *et al.*, 2010, p. 4).

Figura 24 – Artigo do lema *ad*. Retirado do BBI (Benson *et al.*, 2010, p. 5).

Figura 25 – Artigo do lema *acumen*. Retirado do BBI (Benson *et al.*, 2010, p. 5).

Figura 26 – Artigo do lema *canon*. Retirado do BBI (Benson *et al.*, 2010, p. 56).

Introdução

“A word is a microcosm of human consciousness.”

Vygotsky, 1987

Desenvolver capacidades linguísticas focadas na combinatória de uma língua estrangeira nem sempre é uma tarefa fácil. E por esta razão, uma componente vital para a codificação de, neste caso, Português Língua Estrangeira (PLE), é a compreensão e a assimilação particular do uso das colocações lexicais portuguesas.

Tal tarefa obriga à conjugação correta das formas lexicais de maneira a construir frases contextualizadas e padronizadas de acordo com o nível nativo do português, segundo o Quadro Europeu Comum de Referência para as Línguas (QECR).

Embora um estudante avançado de PLE consiga elaborar frases sintaticamente corretas, naturalmente nem sempre estas frases irão corresponder ao sentido que o aluno está a tentar produzir em termos pragmáticos. Estas construções peculiares podem ser entendidas por qualquer nativo do português, no entanto, podem soar de forma equívoca perante o seu contexto de uso pois não são fiéis ao costume da língua.

Fontenelle (1994) nomeia dois exemplos ilustrativos desta precisa situação, no léxico inglês, que podem ser transcritos para a língua portuguesa. Com o uso de “sour milk” (leite azedo), muitas vezes, aprendentes da língua trocam esta combinação por “rotten milk” (leite podre) (p. 1). E, de facto, podemos afirmar que qualquer nativo português conseguiria compreender a mensagem que estaria a ser veiculada por “leite podre”, porém iria também concluir que quem usou esta colocação não é um conhecedor da normativa linguística portuguesa.

Este fenómeno da aprendizagem de língua estrangeira revela a dificuldade universal em contextualizar as colocações, nomeadamente aquelas que conferem concordância com os aspetos linguísticos e socioculturais do léxico. Estas combinações que entram em conformidade com o costume linguístico são chamadas de colocações lexicais. Mas o que são elas? Como podem ser aprendidas? E onde podem ser representadas?

As colocações, como entidades únicas do léxico, não possuem uma representação sistemática e valorizada dentro do leque pedagógico, muito por causa dos ensinamentos da semiologia do século XX relativos ao estruturalismo linguístico, originários de autores como Saussure e de Chomsky, que levaram ao abstracionismo da vertente comunicativa das colocações.

Por conseguinte, o estudo da combinatória na esfera pedagógica linguística é regularmente descartado pela sua suposta irrelevância. Não só por esta razão, mas também porque, ainda hoje, é difícil apontar uma única definição para o termo de *colocação* e o que tal interpela, sendo que os estudos sobre as colocações acabam por ser subjetivos ao indivíduo. Em função disto, o tratamento de colocações em ferramentas de pesquisa, na sua maioria, é escasso e incompleto. Por conseguinte, como podem os alunos de PLE beneficiar de um campo tão complexo, mas fundamental, como o das colocações?

Em virtude ao referido, esta tese pretende primeiramente estabelecer uma revisão e análise sobre a literatura associada ao estudo da combinatória lexical. Destacamos a função primordial das colocações no léxico, de forma a demonstrar uma perspetiva mais clara relativamente à diversidade da conceção do termo e das suas ramificações nos campos da linguística, da metalexigrafia e da linguística informática.

De seguida, propomos enquadrar a relevância das colocações para o estudo concreto de PLE, e o que tal proporciona aos estudantes desta disciplina. Assim sendo, pretendemos explorar teoricamente o tipo de desafios que tal aprendizagem impõe, assim como averiguar um sistema didático a seguir de forma a desfazer as dificuldades no reconhecimento e uso das colocações por parte dos alunos.

Finalmente, e de forma elaborada, concedemos, em função da literatura estudada ao longo deste projeto, a organização de uma metodologia benéfica para uma melhor representação das colocações numa ferramenta de aprendizagem, nomeadamente num dicionário especializado em colocações portuguesas. Consecutivamente, pretendemos idealizar as bases teóricas para uma ferramenta lexicográfica monolíngue, capaz de fornecer informação construtiva a favor da produção da combinatória lexical portuguesa, focada em ajudar didaticamente alunos de português língua estrangeira.

O resultado desta pesquisa tem como objetivo conceder aspetos relevantes não só para a esfera didática do português, mas também com ele aspiramos projetar um impacto instrutivo no ramo de outras áreas associadas ao espectro das humanidades. Realçamos ainda que com esta metodologia procuramos contribuir para o enriquecimento e progresso contínuo da lexicografia portuguesa.

1. O conceito de colocação

1.1. Expressões idiomáticas *versus* colocações

Primeiramente, é importante mencionar a existência de expressões idiomáticas e o que essencialmente as faz diferir das colocações. Sendo um tópico diversamente discutido no ramo da lexicografia, as características de ambas podem ser um tanto ambíguas não só para alunos, como também para especialistas.

Podemos considerar uma primeira linha divisória entre os diferentes tipos de combinações lexicais: a linha que separa as combinações fixas e as livres. Claramente, as fixas transcrevem-se na forma de expressões idiomáticas e as livres correspondem às colocações. A validação das últimas é enquadrada em três pontos essenciais: a sua valência semântica, a lógica da junção dos seus elementos singulares e a frequência de uso destas combinações no léxico.

Desenvolvendo um pouco a homologação das colocações, a frequência das mesmas é um fator imperativo para o seu estudo e representação. Posto isto, se um bigrama de duas palavras coocorrerem frequentemente então isto mostra evidência de que esta combinação possui relevância na pré-fabricação da língua. Estas combinações podem, ou não, ser filtradas para propósitos de investigação, como por exemplo na averiguação das combinações que importam para o estudo de padrões frásicos. Atualmente, esta retração e filtração de combinações pode ser conseguida a partir de *corpus* linguísticos através da lexicografia computacional, que se encontra em contante aperfeiçoamento.

Quando lidamos com colocações que são fluídas quanto ao seu colocativo, ou seja, o acompanhante da palavra-chave da combinação, a extração por frequência deste elemento pode assistir pobremente na pesquisa lexicográfica. Neste caso é importante definir um espaço de intervalo, digamos com mais de duas palavras dentro de cada combinação, para perceber o que é relevante no léxico. Este recurso em particular é relativo ao método da análise por variância.

Partindo para a caracterização de expressões idiomáticas, Fontenelle (1994) diz que “What characterizes idioms is the fact that they constitute a single semantic entity, and that their meaning is not tantamount to the sum of the meanings of the words they are made up of” (p. 1). Por “tantamount”, o autor exprime que não há adjacência entre o significado literal de cada palavra da combinação em relação ao seu significado metafórico. Para ilustrar isto, o autor utiliza a expressão idiomática de “to lick somebody’s boots” (p. 1). Nesta expressão idiomática, o sujeito não vai de facto “lamber” as botas de alguém. Esta combinação fixa tem deveras um significado figurativo. Seguidamente, o autor acrescenta

uma característica interessante sobre as expressões idiomáticas. Estas, por norma, não podem ser manipuladas sintaticamente, como podem ser as colocações lexicais, pois apresentam maiores restrições sintáticas. Adicionalmente, nas expressões idiomáticas não há espaço para a aplicação da voz passiva, da adjetivação, da pronominalização, entre outros (Fontenelle, 1994).

Esta aplicação da voz passiva, adjetivação participial indo até à presença/ausência de artigo, como entendido por Álvaro Iriarte (2001), é figurada e comparada entre a expressão idiomática “perder a cabeça” **(a)**, que apresenta maiores restrições sintáticas, e a colocação “prestar atenção” **(b)**, que já admite algumas transformações. Tal comparação pode então ser revista seguidamente (Iriarte, 2001, pp. 176-177):

(a) O João perdeu a cabeça

(b) O público prestou atenção

1. Passivação:

*A cabeça foi perdida pelo João **(a)**

*A atenção foi prestada ao ministro pelo público **(b)**

2. Adjectivação participial:

*A cabeça perdida ... **(a)**

*O ministro agradeceu a atenção prestada **(b)**

3. Relativização:

*A cabeça que perdeu o João **(a)**

*Surpreendeu-nos a atenção que as crianças prestavam **(b)**

4. Pronominalização:

*O João perdeu-a **(a)**

*O público prestou-a **(b)**

5. Modificação adjectival:

*O João perdeu a impaciente cabeça **(a)**

*O público prestou grande atenção **(b)**

6. Modificação nominal:

*O João perdeu a cabeça da serenidade **(a)**

*O público prestou atenção de grande intensidade **(b)**

7. Modificação adverbial:

*O João perdeu a cabeça intensamente **(a)**

O público prestou atenção ininterruptamente **(b)**

8. Determinação:

*O João perdeu aquela cabeça **(a)**

*O público prestou aquela atenção **(b)**

¹ Exemplos retirados de Iriarte (2001), adaptados de Aguilar-Amat (1993, pp. 67-68).

9. Quantificação:
- *O João perdeu muito a cabeça **(a)**
 - O público prestou muita atenção **(b)**
10. Indefinição:
- *O João perdeu uma cabeça **(a)**
 - *O público prestou uma atenção **(b)**
11. Pluralização:
- *O João perdeu as cabeças **(a)**
 - *O público prestou atenções **(b)**
12. Presença/ausência de artigo:
- *O João perdeu cabeça **(a)**
 - *O público prestou a atenção (b)
- (embora: O público prestou a devida atenção).

Este conceito de restrição sintática na combinatória provém de teorias iniciais no estudo de frasemas, nomeadamente de Fraser (1970) e do seu conceito de “frozenness hierarchy”.

Wulff (2010) explora este conceito, explicando que tal é verificado pela intuição do autor e não pela análise de “naturally occurring data language” (p. 68) e adianta que é difícil descrever todos os frasemas como expressões excluídas de flexibilidade. No entanto, o modelo de Fraser, na Figura 1, vem complementar o estudo da idiomática lexical, em forma de escala de níveis. Estes níveis variam desde o nível 0 sendo este “completely frozen”, de extrema fixidez, até ao nível 6 “unrestricted”, completa flexibilidade.

Figura 1

Tabela representativa da noção de *frozenness hierarchy*

Level	Transformation	Example
6	Unrestricted	
5	Reconstitution	Nominalization
4	Extraction	Particle movement, passive, preposing of PPs
3	Permutation	Indirect object movement, particle movement
2	Insertion	Insertion of non-idiomatic constituent(s)
1	Adjunction	Gerundive nominalization transformation
0	Completely frozen	

Nota: de Fraser (1970), retirada de Wulff (2010, p. 69).

1.2. Identificação de colocações

Seguindo o conceito hierárquico das combinações, entra também a questão da caracterização detalhada das colocações, que, segundo Mel'čuk (1981), são essencialmente combinações lexicais restritas.

De forma sintetizada, defendemos, ao longo deste trabalho, que as colocações são combinações não fixas, mas lexicalmente restritas. Daí o exemplo de “rotten milk” (Fontenelle, 1994) não fazer sentido no léxico inglês. A colocação “rotten milk” não tem viabilidade no uso da língua, pois “rotten” simplesmente não possui sentido lexical junto a “milk”. Isto demonstra que as colocações não podem sofrer substituição por sinónimos próximos sem perder o seu valor semântico. Consequentemente, um conhecedor do léxico inglês sabe que “sour milk” (Fontenelle, 1994) é a colocação correta a ser utilizada em contexto, sabendo que “sour” é deveras um elemento restrito e literal. Este simples exemplo mostra então as implicações sociais da nossa língua, tais implicações que merecem ser assunto de pedagogia, especialmente para alunos de língua estrangeira.

Adicionalmente, é oportuno enfatizar a prosódia semântica das colocações. Se, por exemplo, um verbo e um nome não são recorrentes em termos de contexto, estes transcrevem uma prosódia semântica negativa, como “rotten milk”. Caso se verifique sentido contextual numa combinação, como “sour milk”, então esta classifica-se como uma combinação idiosincrática – pois testa que os seus elementos fazem sentido semanticamente. É importante reter que não se pode alternar um elemento de uma colocação com outro tipo de item lexical sem que a colocação perca o seu sentido e função lexical inicial.

Com a teoria Sentido-Texto de Mel'čuk (1981) vemos que nas colocações um dos seus elementos é assim restrito em termos da combinação de palavras. Apenas uma pequena porção de variantes fazem sentido com a base da colocação, a palavra-chave. E, por essa razão, as colocações são imprevisíveis pois requerem conhecimento do fabrico do léxico.

Nesta teoria, Mel'čuk (1981) salienta que os constituintes de uma colocação não são livres em termos de significado, pois necessitam de ter os seus acompanhantes por afinidade, para conterem sentido contextual e semântico – a palavra-chave de uma colocação precisa do seu outro respetivo componente subordinado, o colocativo. O colocativo por si só, precisa também da sua palavra-chave correspondente de forma a possuir um significado apropriado. Conclui-se, por conseguinte, que as colocações são elementos de composicionalidade limitada, ao contrário das expressões idiomáticas.

Tomemos a colocação “amor cego”, de forma a ilustrar a situação no português. Ao analisar esta colocação entendemos que os seus elementos são o substantivo “amor” acompanhado pelo adjetivo “cego”, e, conseqüentemente, é perceptível compreender que o significado desta colocação é o amor incondicional que uma entidade sente por outra. O adjetivo “cego” modifica completamente o tipo de amor que o sujeito sente, dando-lhe um significado de extrema paixão. Inclusivamente a sua tradução para o inglês, “blind love”, seria algo facilmente identificado por qualquer pessoa possuidora de conhecimento bilingue, e por essa razão seria fácil de estabelecer uma ligação ao significado deste semi-frasema. Em contrapartida, a expressão idiomática “esticar o pernil”, reconhecemos o verbo “esticar” e o substantivo “pernil”, mas neste caso um aluno de português língua estrangeira (PLE) poderia não entender o seu sentido metafórico correspondente ao verbo “morrer”.

Conclusivamente, o significado da palavra-chave de uma colocação está, de certo modo, restrita à existência do seu colocativo, que podem ser vários. No entanto, quanto mais restrita a palavra-chave é do seu colocativo, mais exclusiva se torna a direccionalidade da colocação, revelando uma frequência de baixo valor no léxico.

Porém, a identificação e defesa das colocações no léxico não fica apenas por autores como Fraser ou Mel’čuk. Dito isto, juntamente com Mel’čuk, outros autores fundamentais para este estudo serão seguidamente explorados.

Em síntese, deduzimos que as colocações fazem parte da convencionalidade de qualquer léxico, pois através delas é possível nos exprimirmos de forma estruturada e concisa, e ainda identificar e possuir conhecimento cultural sobre uma língua – “[...] dictionaries are as much reference works aimed at particular groups of users as they are linguistic descriptions of the lexical resources of the language.” (Jackson, 2002, p. 177). Por esta razão, a capacidade da combinatória do léxico é pertinente ao estudo do fenómeno linguístico e social da nossa comunicação, principalmente para falantes não-nativos, na qual merece uma devida representação em ferramentas de referência e de consulta como os dicionários especializados.

2. Revisão literária

2.1. Firth

Firth, fundador da *London School of Linguistics*, é considerado uma das figuras mais importantes na área da linguística descritiva inglesa, tendo contribuído para o desenvolvimento não só do estudo da fonologia e da semântica, mas também, de uma forma mais progressiva, fundamentou a relevância do estudo empírico do contexto social dos *corpora*.

Firth, com a sua pesquisa sobre as colocações, proporcionou uma visão primordial e intrínseca para o estudo das mesmas, nomeadamente quanto à sua importância sintagmática e contextual bem como o uso de estruturas pré-fabricadas na língua, vitais para o entendimento linguístico de qualquer léxico.

O termo “collocation” foi primeiramente introduzido e analisado no seu trabalho *The Technique of Semantics* (1935) e consecutivamente explorado noutras obras como *Modes of Meaning* (1951), onde a unidade representante das colocações e os aspetos idiomáticos do léxico foram estudados pela sua dimensão lexical, semântica e morfológica. Desta forma, o autor demonstrou a relevância da combinação como uma entidade fundamental na construção da língua. Muitos dos seus axiomas permitiram uma maior compreensão na contextualização das colocações e de como o significado de um item lexical é influenciado pelas palavras que o acompanha.

A sua teoria visa nos mais diversos aspetos de construção frásica. Em *The Technique of Semantics*, o autor apresenta a noção de colocação e elabora cinco dimensões correspondentes ao seu significado. Estas dimensões fundamentaram-se na fonética e nos aspetos lexical, morfológico, sintático e semântico (Léon, 2007, p. 2).

Firth não se preocupava com o poder polissémico das palavras, e via os elementos de uma colocação como recipientes de “mutual expectancy” – referindo-se, por exemplo, aos elementos de “dark” e “night”. O autor viria a acrescentar que “one of the meanings of night is its collocability with dark, and of dark, of course, collocation with night” (Firth, 1957, p. 196). Isto significa que, para Firth, as colocações podiam ser vistas como elementos de ocorrência restrita. A referência de “meaning by collocation” não capturava o significado contextual das palavras, mas sim a reflexão sintagmática dos elementos,

The moment a conversation is started, whatever is said is a determining condition for what, in any reasonable expectation, may follow. What you say raises the threshold against most of the

language of your companion, and leaves only a limited opening for a certain likely range of responses. This sort of thing is an aspect of what I have called contextual elimination. (Firth, 1957, p. 31)

Sendo assim, o autor abordava as colocações não como uma justaposição de palavras, mas como uma unificação de mutualidade entre os itens lexicais, sendo que “The collocation of a word or a ‘piece’ is not to be regarded as mere juxtaposition, it is an order of mutual expectancy.” (Firth, 1968, p. 181).

Em *Modes of Meaning*, Firth aplica a sua visão empírica sobre colocações do léxico inglês utilizando poemas do escritor inglês Algemon Charles Swinburne assim como correspondência anotada dos séculos XVIII e XIX. Estas fontes de texto viriam a ser o seu *corpus* para descrever a estilística da combinatória inglesa.

O que Firth (1968) também viria a refletir nos seus estudos posteriores, é que “collocations of a given word are statements of the habitual or customary places of that word” (p. 181). Isto mostra que as colocações são então combinações convencionais da linguagem. O autor via as combinações como unidades linguísticas contrárias à noção de palavras isoladas, pois para ele a temática de “one morpheme one meaning” não fazia sentido.

Para Firth (1968), o texto era essencial para o estudo das colocações, sendo que estas deveriam de ser agrupadas, dando assim significado à palavra específica em estudo,

In the study of selected words, compounds and phrases in a restricted language for which there are restricted texts, an exhaustive collection of collocations must first be made. It will then be found that meaning by collocation will suggest a small number of groups of collocations for each word studied. The next step is the choice of definitions for meanings suggested by the groups. (p. 181)

Este tipo de análise continua a possuir uma utilidade didática para estudantes de língua estrangeira, pois um conjunto de colocações associado a uma determinada palavra acaba por fornecer ao aluno a compreensão do significado da mesma, do seu uso e do contexto em que se insere.

Através do seu estudo das colocações em poemas e cartas dos séculos passados, Firth verificou que os textos equivalem a um *corpus* de investigação, sendo que o tipo de colocações seria restrito ao significado do texto estudado, como era, por exemplo, no caso da poesia de Swinburne. Firth considerava assim o estudo das colocações em conjunto com o seu fator de linguagem restrita, pois linguagens restritas limitavam o contexto das colocações. O estudo das linguagens especializadas ajudaria

consequentemente no estudo isolado das colocações ao contrário da investigação da combinatória de um modo homogêneo.

Firth (1968) insistia então que “A restricted language can be said to have a micro-grammar and a micro-glossary.” (p. 106). O autor acreditava no estudo de colocações em linguagens especializadas, desde a científica à literária, de forma a elaborar uma melhor compreensão sobre combinações do ponto de vista metodológico,

The study of English is a very vague expression referring to a whole universe of possibilities which must be reduced and circumscribed to make exact study and disciplined teaching possible. Hence the notion of a restricted language. Restricted languages function in situations or sets or series of situations proper to them, e.g. technical languages such as those operative in industry, aviation, military services, politics, commerce or, indeed, any form of speech or writing with specialized vocabulary, grammar and style. (Firth, 1968, p. 112)

Posto isto, Firth (1968) abordava assim o texto especializado, analisado manualmente, como um *corpus inscriptionum*:

In support of any linguistic analysis formally presented, there should always be texts. It is perhaps never possible nor desirable to present the whole of the material collected during the observation period, but some sort of ‘corpus inscriptionum’ seems to me essential in almost all studies. (p. 32)

É possível assim averiguar que as colocações existentes em textos especializados proporcionam uma maior compreensão em termos de descrição sintática e linguística. Tais textos especializados, segundo Firth, “provide citations in support of dictionary definitions” (Firth, 2007, p. 7). O agrupamento em série de colocações oferece assim a resposta ao significado das palavras, em contraste com o julgamento subjetivo de que as palavras possuem significado simplesmente por existirem.

Embora Firth com o seu empirismo não avance no estudo de *corpora* referente à estatística linguística no enquadramento da computerização lexicográfica, a sua teoria, a sua abordagem de linguagens restritas, e a sua visão de que a soma de textos equivale a um *corpus*, proporcionou uma abertura para o pensamento da testagem de axiomas linguísticos em linguagem real. Uma linguagem real que acabaria por ser unificada em *corpora* através de muitos dos seus seguidores, como Sinclair e Halliday.

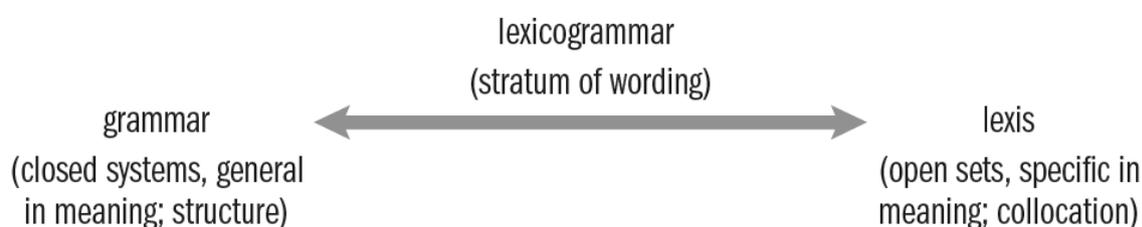
2.2. Halliday

Halliday foi um dos maiores contribuidores para o contextualismo teórico na área da linguística inglesa, e um notável seguidor de Firth. Quanto à sua interpretação das colocações, o linguista via estas como uma parte da vertente lexical da língua e o seu estudo gramatical, defendido por Firth, como um acréscimo. Halliday blindava a esfera da “lexis” como uma máxima independente da “grammar”, cuja conjugação culminava na padronização linguística. Consequentemente, para Halliday, as vertentes lexicais e gramaticais acabariam por representar o “lexicogrammar cline”, que por sua vez reformula a língua e a comunicação como um “continuum” (Halliday, 2014, p. 7). Este *continuum* explora desde a estruturação das palavras com o seu valor semântico geral até à formação de colocações, com significado específico e contextual.

O *lexicogrammar cline*, representado na Figura 2, beneficia das diferentes características de cada vertente das formações combinatórias e pode ser aplicado de forma espectral tanto para o estudo intrínseco das colocações, em termos de extração analítica, como para a construção benéfica de ferramentas de pesquisa (Halliday, 2014).

Figura 2

Gráfico representativo do *lexicogrammar cline*



Nota: Retirado de Halliday (2014, p. 64).

Nos seus trabalhos mais remotos de 1935, Halliday propõe um estudo focado na padronização dos itens lexicais. Desde o início da sua investigação como linguista do léxico inglês, Halliday apelava que os itens lexicais e os aspetos gramaticais dos mesmos possuíam uma relação de continuidade sem fronteiras definidas (Halliday, 2014).

Para Halliday, uma ideia fundamental, no entanto disputada como restritiva por outros autores, é a de que as colocações lexicais são fenómenos estatísticos e de probabilidade. Em 1966, juntamente com McIntosh, o linguista escreveu *Patterns of Language* e, assim, apresentou aspetos comparativos entre os padrões gramaticais e lexicais das colocações defendendo que estas, pela sua perspetiva lexical, possuem pouca colocabilidade, e que, por isso, mereciam um estudo sistemático como organismos

contextualizados, sem hierarquia entre palavra-chave e colocativo, inseridos dentro de um determinado “span” (intervalo) de palavras (McIntosh e Halliday, 1966).

Referindo-se à análise dos elementos das colocações como itens lexicais, o autor declara que, The measure of collocation is the degree to which the probability of a word (lexical item) increases given the presence of a certain other word (the node) within a specified range (the span). This can be measured in the *corpus*. (Halliday, 2014, p. 59)

Consoante o nível de probabilidade obtida de um item lexical estipulado, segundo o autor, é possível extrair uma determinada percentagem de concordâncias de um *corpus* e obter colocações que possuem relações sintagmáticas e lexicais entre si.

Na extração da Figura 3, Halliday (2014) sustenta que os “sets”, ou seja, cada colocação agrupada, partilham particularidades semânticas assim como o mesmo tipo de padronização colocativa.

Figura 3

Excerto representativo das relações sintagmáticas e lexicais das colocações

(i) syntagmatic/lexical [collocation]

change ... mood, season
grass ... flower, green
flower ... bloom
move ... sense ... joy, exhilaration
soil ... land, country
tree ... jacaranda, mimosa
say ... hesitation
country ... beautiful

Nota: Retirado de Halliday (2014, p. 59).

Adicionalmente, ao identificar as palavras como itens lexicais e não como palavras gramaticais, que possuem o mesmo estatuto lexical, o autor nomeia os exemplos de “strong” (forte), “strongly” (fortemente), “strength” (força) e “strengthened” (fortalecido), e explica que

Strong, strongly, strength and strengthened can all be regarded for this present purpose as the same item: a strong argument, he argued strongly, the strength of his argument and his argument was strengthened all as instances of one and the same syntactic relation. (Halliday, 1966, p. 151)

Segundo o autor, “The linguist’s object of study is the language and his object of observation is the text: he describes language, and relates it to situations in which it is operations” (Halliday, 1960, p. 18).

E, deste modo, dentro da esfera dos estudos lexicais ingleses, inspirado pelo trabalho explorado por Firth, Halliday aborda as colocações de uma forma mais moderna que o seu predecessor.

Halliday fundamenta o seu trabalho na investigação da frequência das colocações de forma a representar a sua lexicalização, sempre com um *corpus* a sustentar a sua metodologia, pois para o autor “The *corpus* is fundamental to the enterprise of theorizing language” (Halliday, 2014, p. 51).

Com a mesma intensidade teórica, Halliday (1976) viria a defender três funções linguísticas imperativas ao funcionamento da língua. A linguagem teria assim uma “ideational function”, uma “interpersonal function” e uma “textual function”. O autor acreditava que,

We cannot give anything approaching a definitive answer to the question why human semiotic should have taken precisely this form or no other; but we can begin to look into it; and for this purpose we need a theory of linguistic functions. In principle, language is as it is because of the functions it has evolved to serve. (p. 26)

Existe então uma defesa sobre a metafuncionalidade do significado dos itens lexicais e que a linguagem e o diálogo são conformidades da natureza comunicativa.

Igualmente representado na Figura 4, Halliday caracteriza o fenómeno da organização da linguagem como “The system of a language is ‘instantiated’ in the form of text” (Halliday, 2014, p. 27). O termo “instantiation” refere-se à representação de itens lexicais na sua esfera de contexto dentro de um específico texto,

A text may be a trivial service encounter, like ordering coffee, or it may be a momentous event in human history, like Nelson Mandela’s inaugural speech; in either case, and whatever its intrinsic value, it is an instance of an underlying system, and has no meaningful existence except as such. (Halliday, 2014, p. 27)

Figura 4

Tabela demonstrativa das dimensões da linguagem e dos seus princípios de ordem

	Dimension	Principle	Orders
1.	structure (syntagmatic order)	rank	clause ~ group/phrase ~ word ~ morpheme [lexicogrammar]; tone group ~ foot ~ syllable ~ phoneme [morpheme]
2.	system (paradigmatic order)	delicacy	grammar ~ lexis [lexicogrammar]
3.	stratification	realization	semantics ~ lexicogrammar ~ phonology ~ phonetics
4.	instantiation	instantiation	potential ~ subpotencial/instance type ~ instance
5.	metafunction	metafunction	ideational [logical ~ experiential] ~ interpersonal ~ textual

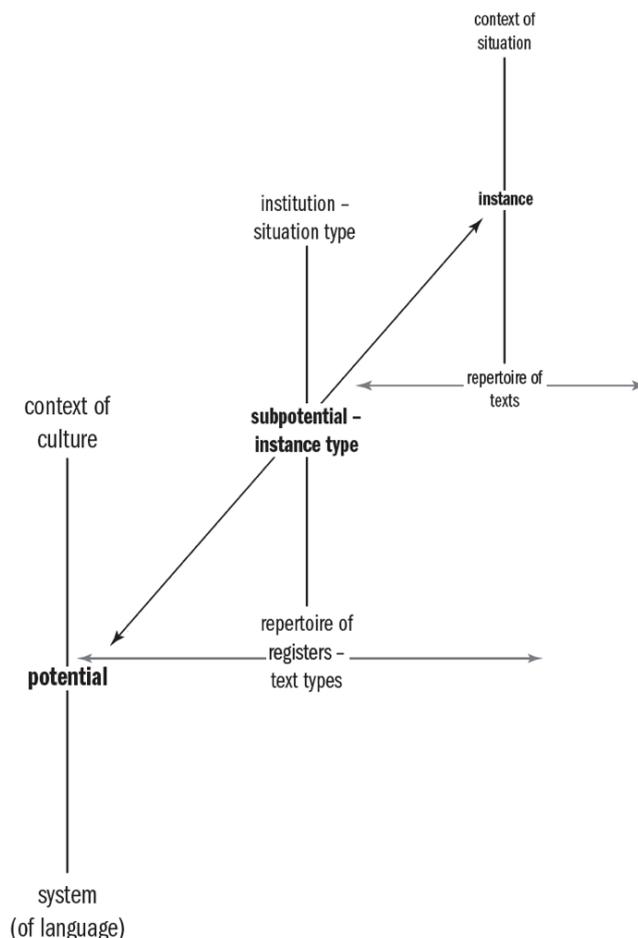
Nota: Retirada de Halliday (2014, p. 20).

Posto isto, a semiótica de um texto só tem validade quando comparada com os seus aspetos de referência no sistema linguístico, sendo tal representado pelo “the cline of instantiation”.

De uma abordagem analítica, a orientação de estudo apresentada na figura 5, inicia-se pelo “potential pole” e termina no “instance pole”, isto sugere uma descrição cultural da língua baseada em situações contextuais da mesma.

Figura 5

O *cline* de instanciação



Nota: Retirado de Halliday (2014, p. 28).

No entanto, Halliday também defende três outras variáveis capazes de assistir à descrição do potencial cultural numa língua. São estes o “field”, o “tenor” e o “mode”.

O “field” é descrito como a situação de estudo e que referencia “ideational meanings” (Halliday, 2014, p. 34). O “tenor” pertence aos elementos que participam na situação, pois colabora com “interpersonal meanings” (Halliday, 2014, p. 34). E o “mode” revela a importância da linguagem e de outros “semiotic systems” que colaboram na construção da situação particular, sendo que culminam em “textual meanings” (Halliday, 2014, p. 33). Sendo assim, Halliday (2014) descreve que,

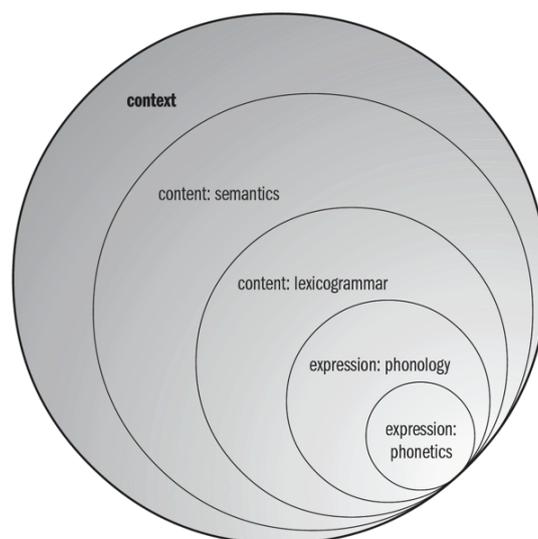
Together they define a multi-dimensional semiotic space – the environment of meanings in which language, other semiotic systems and social systems operate. The combinations of field, tenor and mode values determine different uses of language – the different meanings that are at risk in a given type of situation. (p. 34)

Consequentemente, é possível entender que, o autor apoia-se na pragmática inerente nas colocações dentro de um *corpus* textual, pois para ele “Language is, in the first instance, a resource for making meaning; so text is a process of making meaning in context.” (Halliday, 2014, p. 3).

Halliday via a abordagem das colocações pelo caminho da coesão do texto. Esta coesão é referente nomeadamente ao significado das colocações em relação à sua fonte textual, e, como já referenciado, é dividida em dois setores – o da gramática (*grammar*) e o do léxico (*lexis*). As colocações são então elementos essenciais à estrutura coesa do texto pois existe uma tendência de elas partilharem o mesmo ambiente lexical. O autor exemplifica tal axioma com exemplos relacionais como “ill – doctor” e “sky – sunshine – cloud – rain” (Halliday e Hasan, 1976, p. 286). Os exemplos enunciados demonstram a visão de Halliday em como a coesão gramatical, lexical e o contextual são o que verificam a fiabilidade das colocações num texto. Esta teoria viria a complementar os seus estudos empíricos. Sendo que todas as funções dos itens comunicativos do léxico, como as colocações, verificam a transcendência do contextualismo que absorve os aspetos de “content” e “expression” na sua estratificação (Halliday, 2014), como demonstrado na figura 6.

Figura 6

Esquema representativo da estratificação do léxico



Nota: Retirado de Halliday (2014, p. 26).

Surge assim uma evolução no ramo da Linguística e da Lexicografia. É iniciada uma investigação mais à base do poder dos *corpora* em sistemas computacionais, que viria a ser crucial para o trabalho prático de Sinclair, e de outros especialistas.

Para Halliday tudo era possível graças à potencialidade dos *corpora*, pois estes contêm informação autêntica, incluindo linguagem falada dos mais diversos tipos, capaz de coordenar o estudo linguístico em termos quantitativos. Realisticamente, Halliday (2014) acaba por defender que,

The commitment to data and to the study of small-scale phenomena, in semiotic systems as in systems of any kind. But to banish the macro and the system from one's thinking is simply to indulge in another kind of grand design; being 'atheoretical' disguises a particular theoretical conviction which in our view is ill-judged and ill-informed. (p. 53)

2.3. Sinclair

Sinclair possibilitou grandes contribuições para a expansão tecnológica da Lexicografia, e, nomeadamente, para a análise intrínseca das colocações.

Baseando-se em princípios de Firth e de Halliday, o linguista extraía, eletronicamente, as colocações segundo o seu campo contextual, e salientava a representação das mesmas em ferramentas lexicográficas (Beneduzi, 2008).

O autor dividiu o conteúdo das colocações entre “node”, a palavra-chave, “collocate”, o colocativo, e “span”, sendo este o intervalo de palavras a acompanhar a palavra-chave. A palavra-chave seria então o foco do estudo da combinatória e o colocativo, entendido como a(s) outra(s) palavra(s) que ocorreriam na mesma esfera da palavra-chave, ajudava a caracterizar o significado da mesma. Sinclair via os elementos das colocações como itens de igual importância, sem prioridade em estatuto, pois ambos são considerados relevantes para o estudo da pragmática.

Vendo aptidão nos benefícios estatísticos da análise linguística, Sinclair representa, figurativamente, a palavra-chave por “**a**” e o colocativo por “**b**”. E por essa ordem a colocação (**a+b**) é denominada como “downward collocation” pois a palavra-chave (**a**) é acompanhada pelo seu colocativo, uma palavra com menos frequência (**b**). Adicionalmente, quanto ao inverso (**b+a**), este tipo de colocação é vista como “upward collocation” (Sinclair, 1991, p. 115). E o que as distingue é que,

Upward collocation, of course, is the weaker pattern in statistical terms, and the words tend to be elements of grammatical frames, or superordinates. Downward collocation by contrast gives us a semantic analysis of a word. (Sinclair, 1991, p. 115)

O que Sinclair apresentou também como valioso para a lexicografia foi a importância da investigação empírica das colocações *versus* o estudo das mesmas apenas ao nível teórico. Para o linguista, a lexicografia “was held up by the practitioners to be a largely practical matter, and theories got in the way” (Sinclair, 1991, p. 115). Desta forma, o autor defendia o uso dos computadores e de software com o objetivo de estabelecer um ritual empírico da pesquisa. Sinclair acreditava nesta análise pragmática da língua para atingir detalhes da fraseologia contemporânea, que outrora seria usada consoante o leque do abstracionismo teórico.

Em obras como *Beginning of the Study of Lexis* (1966), o autor abriu a perspectiva da autoridade das colocações no campo linguístico e apresentou, progressivamente, em como as multitudes de *corpora* poderiam vir a ser vitais para a exploração dos fraseologismos.

Sinclair (1991), para produzir a sua metodologia de estudo das colocações, defendia que,

It is, therefore, necessary to have access to a large *corpus* because the normal language is highly specific, and good representative examples are hard to find. This is as true of grammar as of lexis, because grammar is not made of just the patterns of the common grammatical words, but relies on the whole vocabulary of the language. (p. 101)

Assim, obtendo *corpora* com milhões de palavras, a sua abordagem para com os elementos das colocações fundamentava-se na flexibilidade e probabilidade das mesmas, em vez do foco dos significados de cada porção da colocação. Com isto, o linguista apoiou então uma nova abordagem de estratificação dos dicionários. Realçando o estudo da gramática, da polissemia e do lema ao nível linguístico, ao mesmo tempo que desafiando modelos normativos previamente estipulados pela produção dos dicionários, as ferramentas lexicográficas passaram assim a serem vistas pela sua potencialidade como instrumentos de pesquisa deveras prático e informativo.

Sinclair (1991), assumindo as funcionalidades do conhecimento do léxico, averiguava a normatividade dos dicionários como,

The implicit stance of a conventional dictionary entry is that most of the words in daily use have several meanings, and any occurrence of the word could signal any one of the meanings. If this were actually the case, communication would be virtually impossible. (p. 103)

O autor preocupava-se em como funcionava a linguagem e como ocorria a sua pré-fabricação, e para isso usava o meio da lexicografia computacional para obter resultados determinantes. Desta forma, Sinclair colaborou com uma melhor estruturação dos dicionários e defendeu a devida representação de colocações nos mesmos. Um exemplo que culminou desta axiologia, foi uma das ferramentas com mais potencial no campo da Lexicografia – o projeto COBUILD.

Com a construção do COBUILD², iniciado nos anos 80 na Universidade de Birmingham, Sinclair procurou simplificar a apresentação dos lemas, atribuindo a cada um deles informação prática e linguisticamente contextualizada, de forma a melhorar a experiência do usuário. Esta apresentação foi possível ser então sistematizada com recurso a software, cuja informação seria extraída do *corpus* Collins, um *corpus* do inglês contemporâneo.

Para Sinclair (1991) era importante que,

A valid generalization about data must be relate to the data in a systematic way; each relevant instance must either support the generalization or exhibit features which make the generalization subordinate to some other descriptive statement. (p. 101)

Tendo em conta que o COBUILD é uma ferramenta de teor eletrónico, não foi necessário seguir as características restritivas dos dicionários em papel. Sinclair, com a sua equipa de especialistas, apresentou um instrumento de consulta capaz de providenciar conteúdo esclarecedor, sem ser influenciado pelo ciclo unidimensional do mercado comercial dos dicionários de papel. Seguindo as palavras de Bartsch e Evert (2014),

This approach has paved the way for the automatic investigation of collocations on the basis of relatively little linguistic pre-processing other than parts of speech tagging, and by means of statistical methods modelling the characteristic co-occurrence of lexical items in terms of significance of co-occurrence as well as statistical measures of association. (p. 2)

Partindo para o seu método de análise, Sinclair (1991) via os elementos frásicos das combinações como concordâncias,

A concordance has many of the properties of a natural text, and it is reasonable for the purposes of statistical analysis to treat each cited line as if it were a sentence, and so to examine the vocabulary of the concordance. (p. 105)

As concordâncias seriam então compiladas numa listagem em ordem decrescente ao nível da frequência. Mas para tal, a lista de frequências passaria então por diversas etapas de extração e organização.

Primeiramente, Sinclair considerava apenas combinações com menos de cinco elementos frásicos de intervalo. O linguista via que para além de cinco elementos o nível de relevância dos elementos para com a palavra-chave seria mínimo. Em maior pormenor, Sinclair, na sua obra colaborativa *English*

² O COBUILD é um acrónimo para Collins Birmingham University International Language Database, fundado em 1980 e criado por HarperCollins Publishers.

lexical collocation: A study in computational linguistics (1974), aborda as colocações como “the occurrence of two or more words within a short space of each other in a text” (Sinclair, 1991, p. 170).

Seguidamente, o nível de frequência seria estudado, e quanto maior o seu nível mais pertinente seria a sua extração. Posto isto, a listagem inicial seria então minuciosamente revista e reorganizada. O topo da lista apresentaria as colocações mais importantes, ou seja, aquelas com maior frequência de uso no *corpus*.

Tal metodologia e extração viriam mostrar o essencial da separação de “lexis” e de “syntax”. Sendo “lexis” a aglomeração de todas as palavras de um respetivo texto ou língua e a “syntax” a estruturação e organização sintática das palavras no mesmo texto ou língua. Juntamente a esta distinção seria também ideal a desagregação da semântica, pois para o autor “The realization of meaning is much more explicit than is suggested by abstract grammars.” (Sinclair, 1991, p. 108).

Relativamente a extrações feitas a partir de um *corpus* pequeno, o autor viria a exemplificar que estas poderiam ser analisadas pelo “Fisher’s Exact Test”³. Este teste calculava as associações assíduas e não-aleatórias entre palavras de diferentes categorias gramaticais. Para o autor, estas associações “co-occur more often than their respective frequencies and the length of text in which they appear would predict” (Sinclair, 1991, p. 108).

Porém este e outros testes estatísticos não eram o fundamento do trabalho de Sinclair. E isto explica-se pela razão de que a extração de colocações deveria de ser reciclada consoante a arbitrariedade do autor, fundamentando-se este na relevância das colocações para o usuário da ferramenta em construção (Beneduzi, 2008).

Partindo para o método de identificação e de valorização colocativa, Sinclair (1991) mostra que poderiam existir dois princípios de interpretação, completamente dispares, que beneficiam à construção de uma ferramenta de colocações.

O autor menciona, em primeiro lugar, o “open-choice principle”, ou seja, o princípio de escolha aberta. Sinclair (1991) explica que este princípio se baseia no raciocínio de que “at each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness” (p. 109). Para tal princípio seria usado o modelo “slot-and-filler” que se traduz na distribuição das mais diversas palavras em “slots” (espaços) que conduziria à construção das combinações. Sinclair aponta que é através do princípio de escolha aberta que as gramáticas seriam normalmente construídas.

³ O Teste Exato de Fisher é um teste estatístico feito para averiguar a contingência ou independência de valores.

O segundo princípio é o “idiom principle”, o princípio idiomático. Visto que as palavras têm regras para surgirem acompanhadas por outras, este princípio ajuda a elevar o nível de organização de uma língua, coisa que o princípio de escolha aberta não faz. Desta forma, “Things which occur physically together have a stronger chance of being mentioned together” (Sinclair, 1991, p. 110). Por conseguinte, entra então o fator da pré-fabricação da língua, que correlaciona expressões e frases previamente contruídas passando estas a fazer parte da norma linguística na qual “may illustrate a natural tendency to economy of effort” (Sinclair, 1991, p. 110). Sinclair (1991) conclui que para o estudo de textos modernos este é o princípio que mais favorável.

Relativamente ao nível semântico das colocações, existem algumas noções que são rotineiras na análise de *corpora*. Numa extração de *corpora*, é costume encontrar-se uma frequência elevada de determinadas palavras ou significados. E, muitas vezes, estas palavras frequentes são difíceis de catalogar pois nem sempre demonstram uma independência semântica. A dependência de determinados significados de palavras é atribuída, maioritariamente, ao princípio idiomático pois não há possibilidade para a livre conjugação de outras palavras, sendo as colocações um exemplo concreto disso. A mesma situação demonstra também que a sua análise gramatical não é viável, pois segue precisamente o princípio idiomático.

Sinclair (1991) realça que o melhor a fazer na etapa da análise de *corpora* pelo princípio idiomático é o lexicógrafo se preocupar com “uses rather than meanings” (p. 113). E assim acontece uma “progressive delexicalization” dessas mesmas palavras, de forma a descodificar o seu valor real no texto. Todavia, a palavra-chave não pode passar pelo processo de deslexização pois a teoria apoia que a palavra-chave tem um significado independente por si só.

Quando o lexicógrafo tem que saber conjugar os dois princípios de análise referidos, Sinclair (1991) salienta que,

It thus appears that a model of language which divides grammar and lexis, and which uses the grammar to provide a string of lexical choice points, is a secondary model. It cannot be relinquished, because a text still has many switch points where the open-choice model will come into play. It has an abstract relevance, in the sense that much of the text shows a potential for being analysed as the result of open choices, but the other principle, the idiom principle, dominates. (p. 114)

O autor vê assim o dicionário como “a systematic description of a language” (Sinclair, 1991, p. 123), capaz de mostrar o funcionamento fraseológico, ofuscando a utilidade de algumas regras teóricas.

O objetivo desta análise intrínseca seria evidentemente o de representar colocações com o melhor *outlook* possível – sendo indispensável explicar contextualmente os elementos inerentes a uma colocação, de forma a estabelecer familiaridade linguística para o usuário e oferecer uma leitura pedagógica da ferramenta lexicográfica.

2.4. Cowie

Cowie é uma figura vital para o estudo da criação de dicionários. A sua filosofia relativa à interpretação e condicionamento de colocações em dicionários demonstra uma imperatividade no tratamento das mesmas, principalmente na aprendizagem de novas línguas.

É com Cowie que, a partir de 1981, é explorada uma noção mais simplificada do conceito de colocação. O autor propõe quatro ramificações de colocações: as colocações abertas, as colocações restritas, as expressões idiomáticas figurativas e as expressões idiomáticas puras.

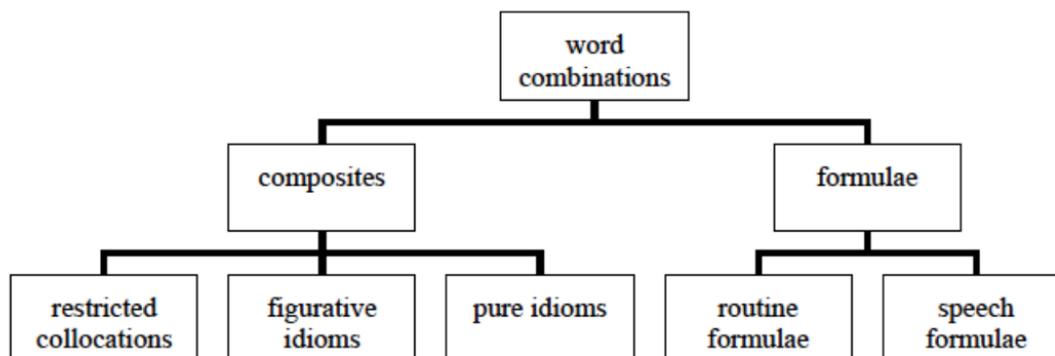
Cowie (1981) vê as colocações abertas como uma junção de elementos transparentes cuja mutabilidade pode ser aceite sem que ocorra mudança no aspeto semântico da colocação. Por exemplo, na colocação “beber água”, o objeto direto “água” poderia ser trocado por uma lista múltipla de outros objetos diretos como “chá” ou “café” sem que a intenção do verbo “beber”, ou da colocação, fosse ofuscada por esta mudança de elemento. O mesmo poderia acontecer com o verbo “comer” por exemplo, assim como com outros sinónimos do verbo “beber” que poderia ser substituído por “ingerir”, “tomar” e afins. De uma forma generalizada, este tipo de colocações é visto como combinações livres pela comunidade lexicográfica.

Já as colocações restritas, estas possuem um elemento especializado acompanhado por outro item com significado literal sendo que determina o uso do último. Por exemplo, na colocação “tocar flauta”, o verbo “tocar” possui uma vertente figurativa pois depende do significado do elemento que o acompanha – “flauta”. Se substituíssemos “flauta” por “a cara”, o verbo perderia o seu significado anterior pois “tocar flauta” refere-se ao ato restrito de soprar ar no instrumento para produzir um determinado som enquanto “tocar a cara” refere-se ao ato restrito de tatear ou sentir a face de alguém.

Tendo em conta as colocações restritas, Cowie (1988) ordena uma representação entre “formulae” e “composites”, sendo que os *composites* referem-se às colocações restritas e expressões idiomáticas e *formulae* fundamenta-se na (pré) fabricação da língua.

Figura 7

Gráfico representativo da classificação de combinações de palavras



Nota: Esquema de Cowie (1988), retirado de Pereira da Costa (2017, p. 22).

Na figura 7, o termo *formulae* é usado pelo autor para se referir às combinações que salientam elementos socioculturais e pré-fabricados, como a forma de saudação “Bom dia”. Já os *composites* são as combinações sintáticas da língua – as colocações restritas e as expressões idiomáticas (figurativas e puras, que são normalmente fraseologicamente estudadas em conjunto).

Posto isto, Cowie (1988) questiona como é possível representar tais ramificações sintáticas deveras imprevisíveis e arbitrárias em dicionários didáticos, sem elevar a dificuldade de consulta para um público principalmente composto por aprendizes de uma língua estrangeira.

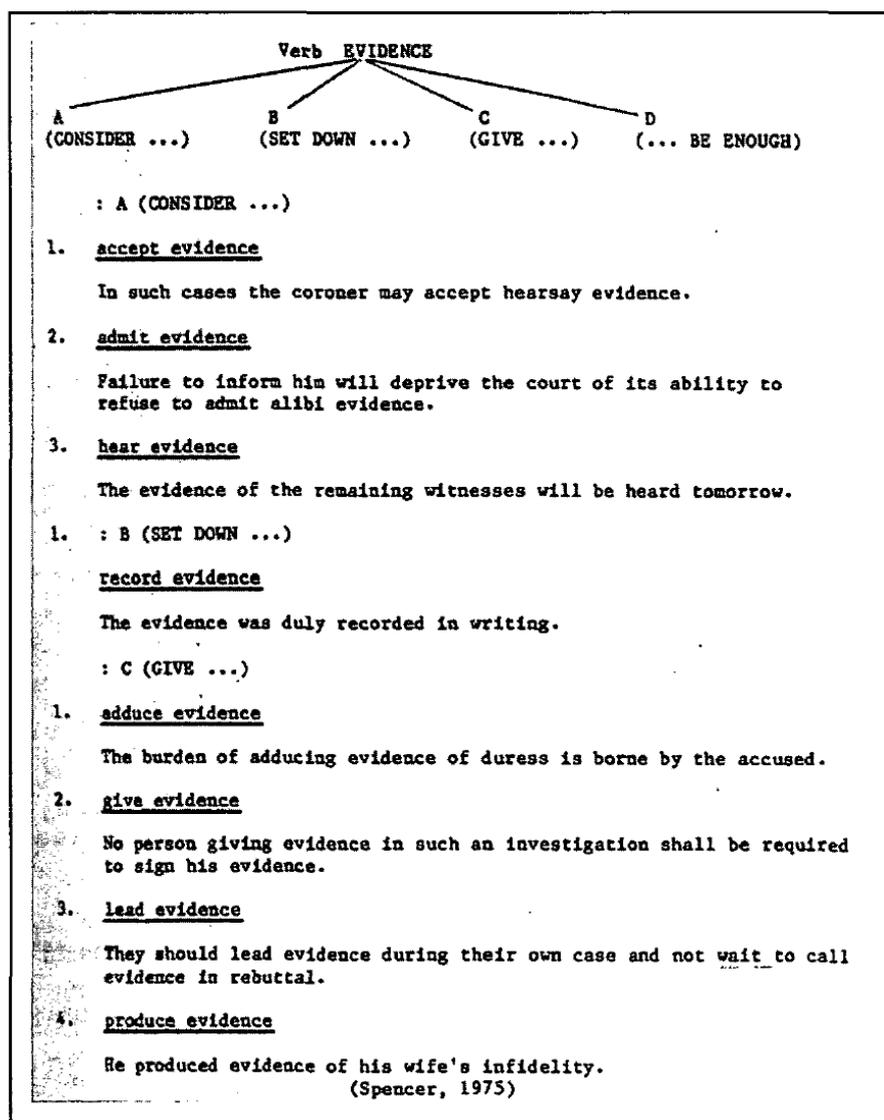
O autor defende um processamento de colocações consideradas úteis para alunos de língua estrangeira e a criação de dicionários com foco na compreensão das colocações. Segundo Cowie (1988), num dicionário, estas colocações deveriam de ser representadas pelas suas palavras-chave de forma a apresentar uma certa familiaridade na macroestrutura usual dos dicionários.

Com a sua obra *Oxford Dictionary of Current Idiomatic English* (1975), em conjunto com Mackin, Cowie conseguiu produzir uma ferramenta centrada na fraseologia descritiva das colocações. Explorando o tipo de processamento das colocações para serem incluídas em dicionários, Cowie defende o método NVELE (*Noun-Verb Expressions in Legal English*) de Spencer (1975)⁴, para a extração seletiva e apresentação condensada das colocações. O excerto recolhido na Figura 8 ilustra tal método.

⁴ Em 1975, Arnold Spencer procurou descrever lexicograficamente palavras associadas à esfera da linguagem legal através do método NVELE.

Figura 8

Excerto do processamento do método NVELE



Nota: de Spencer (1975), retirado de Cowie (1985, p. 67).

Na figura 8 é possível perceber que Spencer (citado por Cowie, 1985) tentou procurar palavras-chave significativas compatíveis com determinados verbos. Idealmente, a organização do dicionário pelas palavras-chave de cada colocação conduziria o usuário a absorver toda a informação inerente ao núcleo da combinação de forma eficaz.

Consequentemente, no *Oxford Dictionary of Current Idiomatic English* (1975), Cowie procurou expor uma organização lematizada que apresentaria a explicação dos próprios lemas assim como dos elementos que, em concordância com as palavras-chave, apresentariam possíveis colocações. Embora o autor achasse o método NVELE eficaz, o seu uso não era conveniente para questões de economia de

tempo. Por esta razão, o NVELE viria a ser usado por Cowie em apenas extrações a partir de *corpus* pequenos.

2.5. Hausmann

Hausmann é um nome solene na Lexicografia e um grande defensor da cultura do uso do dicionário em ambientes de aprendizagem.

O autor possui uma visão do dicionário focada no acesso à sua estrutura – a sua organização e apresentação de informação – assim como no tipo de usuário que consulta dicionários específicos, como aqueles direcionados para alunos. Para Hausmann (1989), a composição e disposição de informação num dicionário seria o que avaliaria a eficácia do trabalho lexicográfico como ferramenta de pesquisa.

Quanto à representação de colocações em dicionários, Hausmann (1989) defendia seis diversos tipos de estruturas:

- **Tipo a** – substantivo + adjetivo (“solteiro endurecido”)
- **Tipo b** – substantivo + verbo (“atenuar a raiva”)
- **Tipo c** – verbo + substantivo (“levantar dinheiro”)
- **Tipo d** – verbo + advérbio (“chover torrencialmente”)
- **Tipo e** – adjetivo + advérbio (“gravemente ferido”)
- **Tipo f** – substantivo + preposição + substantivo (“uma explosão de raiva”⁵)

Estes modelos estruturais propõem axiomas importantes para a produção de dicionários de colocações. Apesar de não abrangerem combinações como advérbio + frase proposicional (“bêbado como um barril”), advérbio + adjetivo (“completamente bêbado”), advérbio + infinitivo ou advérbio + advérbio (“gravemente ferido”), os tipos de colocações defendidos por Hausmann demonstram uma importância na associação sistemática de elementos sintáticos, particular em colocações com dois elementos: “On appellera collocation la combinaison de deux mots” (Hausmann, 1989, p. 1010).

Em 1997, Hausmann viria a referenciar a hierarquia e o binarismo entre a palavra-chave e o colocativo como duas fações que refletem as normas da linguagem na sociedade. Hausmann denuncia que as palavras têm assim um estatuto hierárquico lexical como membros nobres e as locuções são vistas como plebeus, com um significado menos transparente que a palavra-chave e dependente da mesma, e que, por esta razão, acabam por ser negligenciadas.

⁵ Tradução dos exemplos em francês dados por Hausmann (1989).

Por conseguinte, a filosofia de Hausmann estabelece a hierarquia entre palavra-chave e colocativo, sendo que a palavra-chave possui uma maior autonomia semântica em comparação com o seu colocativo. Este, por sua vez, é um acessório descritivo da palavra-chave. O colocativo encontra-se então à mercê da classificação e significado da palavra-chave para possuir uma determinada interpretação. Assim, é demonstrada a hierarquia de valor entre palavra-chave e colocativo: a palavra-chave é o topo da pirâmide hierárquica, o elemento mais importante da colocação. Assiste-se aqui a um comportamento morfossintático, pertencente a uma relação direcional e a uma estruturação de combinação restrita.

Tendo em conta a noção de identificação hierárquica entre elementos das colocações, o autor assinala a dificuldade em distinguir numa colocação qual dos elementos é a palavra-chave e qual é o colocativo. Hausmann (1989) dá o exemplo de substantivo + preposição + substantivo, com “a pack of dogs”, e com verbo + preposição + substantivo, com “blush with shame”. Em ambas as estruturas, é difícil de identificar quais elementos seriam as palavras-chave. Para Hausmann, a solução para representar tais casos em dicionários seria a criação de lemas para ambos os predicados.

Relativamente ao binarismo (colocações com dois itens lexicais) associado à representação de colocações, Hausmann (2004) classificou como “collocational chains” os exemplos onde as colocações possuíam mais que dois elementos sintáticos, como por exemplo em “pay close attention”.

O importante para Hausmann seria então o tratamento das colocações após a identificação da(s) sua(s) base(s). A partir daqui, existiria uma interação entre a macroestrutura e a microestrutura do dicionário. A estrutura de cada artigo do dicionário mostraria a existência de itens e indicadores importantes à explicação dos lemas conduzindo assim à harmonização entre a macro- e microestrutura do dicionário.

Para Hausmann (2004) é importante prestar atenção à distribuição da informação num dicionário, sendo importante a inclusão de uma lista dos lemas antes da sua apresentação em artigos de dicionário, uma espécie de index de todas as palavras-chave do dicionário.

Sendo assim, a macroestrutura é essencialmente vista como a que ordena toda a informação da ferramenta lexicográfica, onde os lemas são os protagonistas.

Hausmann considera o colocativo como sendo o elemento mais difícil de intelectualizar para um aprendiz de uma língua estrangeira. Isto acontece precisamente pela sua dependência semântica – o cariz do colocativo relaciona-se com a sistemática da palavra-chave de uma colocação. A palavra-chave é mais fácil de ser interpretada por um aluno de línguas graças à sua autonomia de valor textual, característica que o colocativo não possui.

Por conseguinte, para o autor, este caráter arbitrário, independente e idiossincrático das colocações deve então ser ilustrado nas ferramentas orientadas à produção de texto.

Para Hausmann (2004), os dicionários podem ser macro-estruturados de forma diferente, conforme as necessidades do usuário. Num dicionário para recepção de texto, a organização do dicionário deve ser enfatizada em volta do colocativo, pois é visto como o elemento mais indispensável para a codificação de uma língua. Mas quando o dicionário é focado na produção de texto, idealmente, este deve de apresentar uma estruturação visando na palavra-chave. A palavra-chave é vista como essencial para a codificação da língua, pois o usuário orienta o seu léxico com base na mesma, que por sua vez conduz à ligação com o colocativo.

No entanto, se referirmos dicionários bidirecionais, Hausmann apoia que as colocações devem de ser catalogadas no dicionário tanto a nível da sua palavra-chave como a nível do colocativo. Desta forma, o dicionário apoia a compreensão textual e a sua produção em simultâneo. Hausmann defende também a formulação de uma introdução instrutiva e esclarecedora, de forma a orientar o usuário como tirar melhor proveito da ferramenta lexicográfica.

Segundo Hausmann (2004), embora não existam normas estáticas quanto à organização intrínseca de um dicionário, independentemente do tipo de microestrutura, a simplicidade dos lemas deve de ser o elemento mais crucial para a viabilidade do dicionário. O autor defende assim um formato simples, condensado, mas explícito o suficiente direcionado para cada palavra-chave lematizada.

2.6. Corpas Pastor

Corpas Pastor é um dos nomes contemporâneos mais importantes para o estudo atual de Lexicografia.

Em 1996, a autora propôs a análise das combinações lexicais focada no estudo da Fraseologia. Para Corpas Pastor (1996), o estudo das colocações concentra-se na análise da frequência estatística, nas potencialidades da pré-fabricação da língua, na semânticidade combinatória, na averiguação da idiomatidade das combinações, nas suas variações diatópicas, diafásicas e distráticas.

Quanto à definição de colocação, Corpas Pastor (1996) defende que a restrição de uma colocação depende da relação semântica entre os seus itens, considerando a norma linguística. Sendo assim, a autora vê a palavra-chave como o item lexical que interpreta o colocativo. Por conseguinte, Corpas Pastor (1996) enuncia que,

El colocado autónomo semánticamente (la base) no sólo determina la elección del colocativo, sino que, además, selecciona en éste una acepción especial, frecuentemente de carácter abstracto o figurativo. (p. 66)

Ao longo dos seus estudos lexicográficos, para além de criticar autores que não defendem a diferença de estatuto/hierarquia entre palavra-chave e colocativo, o significado transparente das combinações ou até mesmo a interpretação abstrata da frequência de colocações, a autora contribui com uma distinção pertinente entre colocação e locução. Para ela, uma colocação possui um significado literal enquanto a locução possui um significado figurado.

Continuamente, Corpas Pastor (1996) dá ênfase à esfera normativa das colocações. É através da norma de uso das colocações que se encontram distinções, pois, de acordo com a preferência ou restrição de uso, entra a questão da semânticidade e do fenómeno lexical das colocações lexicais restritas.

Para a autora, é através da normativa linguística que um aprendiz de língua estrangeira consegue interpretar as colocações. Pois são as colocações que estruturam a compreensão textual e a conseqüente produção de texto, “Cuando se trata de comprender un texto, se han de tener en cuenta las colocaciones de las palabras para captar sus sutilezas semánticas.” (Corpas Pastor, 2003, p. 171).

Através de uma adequada metodologia, a aprendizagem de colocações possibilita uma maior habilidade na receção e no desenvolvimento de uma língua para qualquer aprendiz. E, para Corpas Pastor, é essencial uma representação sistemática e eficiente das colocações nos dicionários para facilitar a descrição, a compreensão e a consulta das mesmas a favor do usuário.

2.7. Porzig

Porzig é um nome com grande influência no estudo da Semântica Teórica.

No estudo das colocações, Porzig (1970) vê estas como combinações lexicais pertencentes ao eixo sintagmático e paradigmático. As colocações situam-se em campos verbais na qual os seus acompanhantes definem a sua estruturação na língua. O autor introduz o conceito de “parataxe” de forma a identificar a correspondência de vocábulos nestes campos verbais. Um dos exemplos ilustrativos oferecidos pelo autor é o campo verbal das estações do ano – onde se consegue caracterizar facilmente as ramificações paradigmáticas associadas aos substantivos “Primavera”, “Verão”, “Outono” e “Inverno”.

Em *Wesenhafte Bedeutungsbeziehungen* (Porzig, 1973), o eixo sintagmático dos campos verbais apresenta delimitações quanto à combinação de palavras. É neste eixo que, segundo Porzig, as colocações se inserem. O autor acreditava que, como tudo no mundo natural, cada palavra possui o seu próprio campo magnético onde estabelecia ligações com outras palavras.

Em *Der Wunder der Sprache* (1970), Porzig sublinha a restrição léxica dos elementos combinatórios consoante a esfera de cada língua. Como exemplo, o autor fala na restrição entre as palavras alemãs “Haar” e “blond” (Porzig, 1970, citado por Beneduzi, 2008, p. 26) que demonstra não só uma restrição de associação, mas também um unilateralismo – a palavra “loiro” apenas se associa a “cabelo”, no entanto “cabelo” pode se coligar a outras cores ou tipos de cabelo. A restrição léxica pode ser também variável não só de língua para língua, mas também em termos de significados. Porzig (1970, p. 132), dá como exemplo, no alemão, a palavra “Schnee”, e a dicotomia entre os adjetivos “weiß” e “schmutzig” – aqui a palavra “neve” pode ser interpretada de forma variável consoante o adjetivo adquirido. Seguindo o exemplo, ora pode ser “neve branca” ou “neve suja”.

Porzig interpreta também a linguagem como um objeto de pré-fabricação. A sua visão heterogénea sobre as colocações apresenta uma acentuação no significado de cada item lexical em oposição à relação colocacional das palavras. Para o autor, a sua noção de campos verbais estabelece uma maior compreensão em termos combinatórios entre palavras. Porzig incorpora assim a teoria do estruturalismo lexical na sua defesa dos campos verbais das colocações.

2.8. Coseriu

Coseriu (1977) apresenta uma grande contribuição para a investigação de colocações em relação à sua estruturação ao nível do significado, sendo esta a distinção entre as relações sintagmáticas e paradigmáticas das palavras.

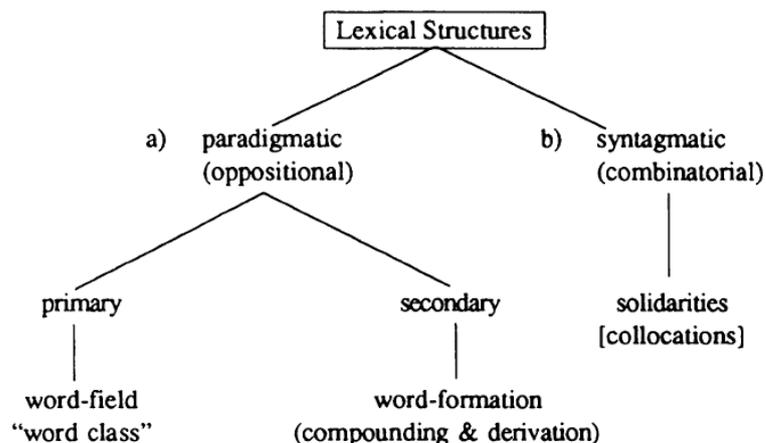
Sendo um ávido seguidor de Porzig, Coseriu (1977) apresentou a noção de “solidariedade léxica” na qual elabora este conceito como uma

determinación semántica de una palabra por medio de una clase, un archilexema o un lexema, en el sentido de que una clase determinada o un determinado archilexema o un determinado lexema funcionan como rasgo distintivo de la palabra considerada. (p. 148)

A figura 9 explicita a estrutura lexical das combinações, onde as colocações consideradas como “solidarities”, ou seja, as colocações, conferem “the relationships between words which are systematically and conventionally combined in a sentence” (Lipka, 1992, p.11).

Figura 9

Esquema ilustrativo das estruturas lexicais



Nota: Adaptado de Coseriu (1970), retirado de Lipka (1992, p. 11).

Com as colocações, Coseriu (1973) defende o espaço geográfico (a variedade diatópica), as diferenças socioculturais (a variedade diastrática) e a diversidade de expressão dos itens lexicais em diferentes situações comunicativas (a variedade diafásica).

Assim sendo, Coseriu (1973) baseia-se na concepção de que a linguagem se ramifica entre as estruturas universal, histórica e individual,

el sistema es un sistema de posibilidades, de coordenadas que indican caminos abiertos y caminos cerrados. (...) Lo que, en realidad, se impone al individuo, limitando su libertad expresiva y comprimiendo las posibilidades ofrecidas por el sistema dentro del marco fijado por las realizaciones tradicionales, es la norma. La norma es, en efecto, un sistema de realizaciones obligadas, de imposiciones sociales y culturales, y varía según la comunidad. (p. 98)

Também semelhante a Porzig, Coseriu defende a autonomia da palavra-chave e acredita que as restrições lexicais se integram nas relações de semântica dentro de uma colocação. Isto acaba por demonstrar uma abordagem ontológica da língua, no sentido em que a análise e a interpretação das colocações se verificam pelo significado das palavras.

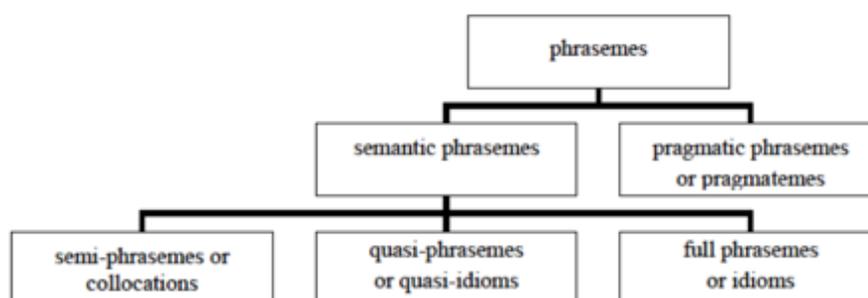
2.9. Mel'čuk

Mel'čuk é uma figura imperativa para o estudo das colocações ao nível linguístico. A sua visão lexicográfica para com as colocações foi das que mais contribuiu para o desenvolvimento deste ramo.

Seguindo a figura 10, Mel'čuk (1998) aprecia as colocações como semi-frasemas semânticos a favor da codificação. Estes semi-frasemas, segundo o autor, são combinações formadas por dois ou mais lexemas parcialmente fixos, onde a palavra-chave possui o significado essencial da colocação. O autor explora assim a noção de assimetria entre os constituintes de uma colocação.

Figura 10

Classificação das combinações por Mel'čuk (1998)



Nota: Retirado de Pereira da Costa (2017, p. 39).

Em 1981, Mel'čuk introduz a noção de funções lexicais. Estas funções demonstravam correlações de sinonímia, antonímia e hiperonímia nas colocações. Mel'čuk (1984) define as funções lexicais como “un sens assez abstrait [...] ou un certain rôle sématico-syntaxique [...] tel que son expression linguistique dépend du lèxeme auquel cette FL s'applique” (p. 6).

Para o autor, a perceção das colocações tinha origem na pré-fabricação da linguagem, pois as colocações são fruto do uso frequente de determinados vocábulos. Por conseguinte, Mel'čuk (2001) defendia o “idiom principle” pois,

People speak in set phrases, rather than in separate words [...]. A phraseme is a lexical unit; and, more crucially, it is the numerically predominant lexical unit: in any language – i.e. in its lexicon – phrasemes outnumber words roughly ten to one. (p. 24)

A maior contribuição do autor para a representação das colocações foi a sua obra *Dictionnaire Explicatif et Combinatoire du français contemporain* (DECFC)⁶.

⁶ O *Dictionnaire Explicatif et Combinatoire du français contemporain*, inicialmente proposto nos anos 60, por Aleksandr Žolkovski e Igor Mel'čuk, consiste numa ferramenta monolíngue capaz de modelar a linguagem natural e apresentar um registo do léxico. A sua metodologia foi estendida a versões em outras línguas como em Russo e em Espanhol (o famoso DICE).

O DECFC (Mel'čuk *et al.*, 1984), contendo inicialmente cerca de 300 lemas do francês, aborda combinações da perspectiva do seu significado e ao nível sintático de cada lema (palavra-chave).

É no DECFC que Mel'čuk explora as funções lexicais, apoiando-se no conceito da função de **f(X)** = **Y**. Nesta equação, **X** equivale à palavra-chave e **Y** corresponde ao colocativo da colocação **f(X)**. Mel'čuk aborda no total 50 funções lexicais que descrevem combinações recorrentes, de elevada frequência estatística, na língua francesa. As funções lexicais seguintes são alguns exemplos em inglês ilustrativos do método utilizado no DECFC⁷, que tomamos de Fontenelle (1994, p. 7):

- Son (dog) = bark
Son (horse) = neigh
Son (elephant) = trumpet
- Oper (pressure) = exert
Oper (attention) = pay
Oper (suicide) = commit
- Liqu (file) = delete/erase
Liqu (law) = abrogate/abolish
Liqu (marriage) = annul
- Mult (fish) = school/shoal
Mult (bee) = swarm
Mult (dog) = pack
- Sing (glass) = splinter
Sing (advice) = piece
Sing (dust) = speck

Este tipo de representação faz com que o DECFC seja uma ferramenta extensa em termos do seu conteúdo. A sua consulta exige um alto nível de conhecimento para qualquer usuário conseguir interpretar a sua informação – particularmente no entendimento das funções lexicais.

É de realçar que esta contribuição para o estudo das colocações não se limita apenas ao aspeto lexicográfico. De facto, Mel'čuk proporciona uma representação de enfoque semântico.

O autor contesta os problemas sobre a representação das colocações em dicionários, sendo que o objetivo do seu estudo foi o de proporcionar uma visão mais ampla ao nível do léxico. O modelo Sentido-Texto é o que justifica esta visão.

Em 1978, Mel'čuk desenvolve assim o modelo Sentido-Texto e revela que o mesmo é um “système de règles formelles assurant la traduction de tout sens donné dans tous les textes

⁷ “**Son** (French, to denote the typical sound or cry made by something), **Oper** (to denote the ‘empty’ transitive verb which takes the keyword as object) or **Liqu** (to refer to the destruction of something) to **Mult** (which denotes the multitude or aggregate of something) or **Sing** (the converse function which denotes a single entity).” (Fontenelle, 1994, p. 7).

correspondants et vice versa” (p. 273). O modelo Sentido-Texto garante um sistema lógico para a representação da metalinguagem lexicográfica, como as unidades lexicais, os lexemas, as colocações, e afins.

Este modelo faz do DECFC uma ferramenta com potencial ao nível da síntese linguística, desviando-se dos aspetos extralinguísticos do léxico como a compreensão da língua. Existe então uma correlação mecânica entre a semântica e o *corpus* usado, onde a ferramenta não se preocupa em descrever o significado de cada unidade lexical, mas sim procura em oferecer o uso contextual de cada colocação.

Posteriormente, Mel’čuk (1995) idealiza a multiplicidade de significados que o léxico pode ter, quando analisando a comunicação pessoal de indivíduos ou então presente em conjuntos limitados de textos especializados.

Em síntese, para o autor, as colocações possuem um significado transparente sendo que este depende da relação semântica restrita entre a palavra-chave e o colocativo. Para Mel’čuk (1995) existe então uma diferença de estatuto entre a palavra-chave e o colocativo, e o autor sublinha ainda que as colocações pertencem à norma linguística e que a sua frequência de uso é o que interpreta a relevância do seu estudo.

Por conseguinte, o DECFC, e a sua versão noutros idiomas, contém um “formalized semantically-based lexicon designed to be part of a linguistic model of natural language” (Mel’čuk 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 1). Este dicionário por si só engloba assim um *outlook* similar ao dos dicionários recorrentes e, por essa razão, não poderia de deixar de exibir o uso de abreviações e de simbolismos, visível na Figura 11.

Figura 11

Abreviações e anotações da simbologia usada na versão do DECFC para o léxico inglês (ECD).

DSyntA : Deep-Syntactic Actant	SemA : semantic actant
ECD : <i>Explanatory Combinatorial Dictionary</i>	SemR : Semantic Representation
GP : Government Pattern	‘X’ : the meaning of X
L : a particular language	$\underline{L_1 \dots L_2}$: the idiom $L_1 \dots L_2$
L : a particular lexical unit	\oplus : the operation of linguistic union (p. 00)
LF : Lexical Function	‘X Y’ : X is the presuppositional part of a meaning (p. 00)
LU : Lexical Unit	*X : ungrammatical expression X
MTM : Meaning-Text Model	?X : grammatically dubious expression X
MTT : Meaning-Text Theory	#X : semantically and grammatically correct, but pragmatically unacceptable expression X

Writing conventions

Italics : linguistic examples, i.e., cited linguistic forms

Boldface : highlighting (including titles)

Courrier New : technical terms on their first occurrence

CAPITAL LETTERS (10 points) : lexical expressions

Nota: Retiradas de Mel’čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 3).

Sendo uma ferramenta orientada a favor da semântica do léxico, Mel'čuk procura descrever, no DECFC, a língua na sua esfera lexical e gramatical.

Em termos lexicais, o autor representa elementos essenciais ao entendimento das unidades do léxico. E, gramaticamente, Mel'čuk interpreta elementos básicos de inflexão e derivação – “affixes, apophonies, conversion” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 3), apoiando-se nas regras semânticas, sintáticas, morfológicas e fonológicas do francês.

Consecutivamente, o autor descreve que o léxico de uma língua particular (**L**) “describes **L**'s individual lexical signs, and the grammar of **L** covers a) **L**'s individual grammatical signs and b) the behavior of sets of **L**'s signs.” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 3).

Sendo assim, o léxico apresentado no DECFC combina todos os aspetos elementares da língua, interpelando os fatores da semântica e da sintática fazendo com que a ferramenta seja um “depository of all lexical data” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 4).

Enquanto o convencionalismo dos dicionários é o de satisfazer as especificações de um determinado usuário, o DECFC procura estudar o lado científico da linguagem (Mel'čuk, 1984). Por esta razão, a ferramenta não se apresenta como um dicionário convencional pois procura representar o conhecimento lexical inculcado nos falantes do léxico. O DECFC não se atenta em seguir as regras da indústria lexicográfica e não se delimita pela sua potencialidade comercial ou pedagógica.

O DECFC é então “developed for the sake of linguistics and should satisfy the normal requirements for scientific descriptions” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 6). Este dicionário é uma ferramenta formal, com ênfase na explicitação e consistência do léxico ao ponto de se caracterizar como uma ferramenta apropriada à esfera computacional.

Quanto às unidades lexicais usadas, o DECFC procura descrever detalhadamente cada lema. Por esta razão, o seu conteúdo não se estende a todas as unidades lexicais presentes numa língua, mas sim importa-se com os fragmentos do léxico mais pertinentes para a comunidade linguista. Estes fragmentos são então estudados de uma forma exaustiva de maneira que “it should contain all of the information needed to use the head LU [lexical units] successfully in all possible contexts” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 8).

Quanto à microestrutura, a versão em papel do DECFC usa referências e determinados códigos para estabelecer ligações entre as unidades lexicais (Mel'čuk, 1984). Na sua obra, as unidades lexicais são divididas entre lexemas, elementos compostos e expressões idiomáticas listados em ordem alfabética. Os lexemas e as expressões idiomáticas partilham praticamente a mesma estrutura ao nível de entrada assim como o mesmo tipo de informação. As colocações (semi-frasemas de terceiro tipo,

segundo Mel'čuk) são descritas em função da sua palavra-chave. Por conseguinte, o DEFCF é uma ferramenta de cariz frásico.

As definições desta ferramenta são assim “verbal and linear” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 47). O que Mel'čuk realça é que as definições do DEFCF passaram por um processamento de forma que os aspetos da linguagem natural se relacionem com a macro- e microestruturas do dicionário. Este processamento visa assim no respeito da metalinguagem – “A description of a lexical meaning in the form of a network” (Mel'čuk, 1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 48).

Em termos de representação de colocações, Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006) vê elas da seguinte forma,

A phrase **AB** = $\langle 'S' ; /A/ \oplus /B/ ; \Sigma_{AB} \rangle$ of **L** is called a collocation if it satisfies

simultaneously the following three conditions:

1. The signified of **AB** includes the signified of **A** as its semantic pivot: 'A' is the argument of the difference 'AB' – 'A' = 'C'. [Formally: 'S' = 'A' \oplus 'C' & 'C'('A').]
2. **A** is selected by the speaker freely, i.e., independently of **B**—for its own signified 'A'.
3. **B** is not selected freely—it is selected for its signified 'C' restrictedly, i.e., as a function of **A**.

(p. 56)

Sendo **A** a palavra-chave e o **B** o colocativo, este último pode ter o sentido de **C** em combinação com **A**, mas **B** encontra-se restrito pois não pode ser substituído por sinónimos, e afins, sem afetar ou perder o seu sentido inicial com **A**. Contudo, **B** sozinho não possui o sentido de **C** mas sim só possui significado apenas numa possível combinação com **A**.

Esta lógica de Mel'čuk ofereceu uma maior contribuição para a lexicologia, particularmente na formulação de ferramentas monolíngues com foco nos fraseologismos. Demonstrando assim que as colocações podem ser uma excelente fonte de estudo e de pedagogia da semanticidade do léxico. Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006) acrescenta que,

High school students are not supposed to be afraid of trigonometry, chemical formulas and computers; why should they be of analytical lexicographic definitions, syntactic patterns and collocation descriptions? (p. 118)

Na versão inglesa do DEFCF, Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 9). dá o exemplo da palavra “dog”, onde a sua descrição ou qualquer variante da palavra inclui as seguintes referências:

- the term of any particular breed of dogs—BORZOI, BULLDOG, GREAT DANE, GREYHOUND, etc.;

- the term for a dog of no definable breed—MONGREL, MUTT; CUR;
- the form of the sign warning of an attack dog's presence—BEWARE OF THE DOG;
- the terms for various human activities related to dogs—KEEP [a dog], WALK [a dog], HUNT [with a dog], MUSH, etc.;
- the names of different actions typical of dogs—BARK, SNARL, HOWL, WHINE, YAP, SNIFF, WAG [the tail], BITE, MAUL, etc.;
- the names of such artifacts used for dogs as COLLAR, MUZZLE, LEASH, DOGHOUSE, BASKET, etc.

No lema “dog”, uma colocação provável a ser associada como “walk a dog” é então identificada e descrita na entrada de “dog,” como uma subentrada contendo a sua informação lexicográfica.

Figura 12

Subentrada de “walk a dog”

take the D. out so that
it can get exercise
and relieve itself : [to] walk [ART ~]

Nota: Retirada da versão do léxico inglês (ECD) do DECFC Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006, p. 9).

O tipo de subentrada, demonstrado na Figura 12, oferece uma riqueza de conteúdo linguístico relativamente a cada lema do dicionário, que não se compara com um dicionário convencional. Por esta razão, o DECFC, ou qualquer das suas versões noutras línguas, através do modelo Sentido-Texto, demonstra ser um dicionário ativo, pois ajuda na consulta de fraseologismos consoante o significado das mesmas.

Categoricamente, o que Mel'čuk acaba por defender através do DECFC é a implementação de um estudo mais crítico da língua na pedagogia contemporânea. O autor acredita que o ensino da semântica e da sintática são importantes para o desenvolvimento linguístico de qualquer aprendiz de língua segunda. O DECFC, com a sua estruturação, favorece características diferentes a serem exploradas em futuros dicionários. Assim, Mel'čuk (1984, 1988, 1992, 1999, citado por SICA (Ed.), 2006) dá ênfase à criação de dicionários monolíngues a favor da aprendizagem integral de uma língua,

The creation of really scientific, reasonably formalized monolingual dictionaries should go hand in hand with new ways of teaching the mother tongue in schools— with the emphasis on formal description of the lexicon and the grammar rather than simply on spelling. (p. 118)

2.10. Benson

Benson e os seus contributos lexicográficos são indispensáveis para a representação das colocações em dicionários especializados.

No seu projeto de 1986, o famoso *The BBI Dictionary of English Word Combinations*, o autor representa as colocações lexicais sobre um sistema de sete tipos de classificações, variando desde L1 a L7m adicionando ainda a representação de colocações gramaticais, mas com oito tipos de classificação, entre G1 e G8.

Para Benson (1986), as colocações lexicais são combinações sintagmáticas de lexemas de conteúdo na qual, na maioria dos casos, não contêm preposições ou infinitivos. Estas combinações sintagmáticas contêm então nomes, adjetivos, verbos e/ou advérbios. Já as colocações gramaticais, estas são combinações entre lexemas de conteúdo em combinação com lexemas gramaticais.

Considerando as combinações lexicais as mais pertinentes para a nossa pesquisa, resumimos a sua caracterização de acordo com Benson *et al.* (2010):

- **L1** – verbo + nome, “make mistakes” / verbo + pronome, “come to an agreement”

As combinações L1 são a junção de um verbo transitivo com um nome, pronome ou frase preposicional. Segundo Benson *et al.* (2010), o verbo admite a conotação de “creation and/or activation” (p. 32) seguido por um nome. Esta fixidez, ora de criação ou ativação, pode ser encontrada em exemplos como “compose music” ou “set an alarm”, respetivamente. Estas duas vertentes do verbo podem estar presentes também num só verbo, como por exemplo em “display bravery”. As colocações L1 são então denominadas como “CA collocations”.

Este modelo de colocação não coloca desafios em termos de compreensão tanto para alunos da língua como para nativos. Benson *et al.* (2010) revela que alunos de inglês não são capazes de distinguir determinadas formas que, embora sejam semelhantes, estejam erradas como, por exemplo, com o uso de “commit treachery” em vez de “commit treason”. O autor também refere que os próprios nativos se podem confundir com colocações deste tipo pois há a vertente dos diferentes tipos de inglês, como o Inglês Britânico e o Inglês Americano, onde dialeticamente o posterior usa “to have a bath” enquanto o anterior diz “take a bath”. O que o BBI tenta mostrar é então estas diferenças de forma a esclarecer o usuário.

- **L2** – verbo + nome, “reject an appeal”

Com as colocações de nível L2, o verbo refere-se a “eradication and/or nullification”, acompanhado por um nome. Por conseguinte, estas colocações são chamadas de “EN collocations”. Exemplos de tal podem ser “break a code”, “revoke a license” ou “exterminate vermin”.

- **L3** – adjetivo + nome, “heavy rain”

As colocações L3 possuem um adjetivo e um nome. Este tipo de colocação pode também abranger colocações de cariz de linguagem técnica como “land reform” ou “aptitude test”. Neste caso, a sua localização no BBI é fixada sobre o segundo nome da colocação, neste caso “reform” e “test”. Já as colocações L3 que possuem a fórmula de nome + nome como “cabinet reshuffle”, estas ficam então identificadas sobre o primeiro nome da colocação, como “cabinet”.

- **L4** – nome + verbo, “bees buzz”

Nas colocações L4, o verbo relaciona-se com uma pessoa ou objeto como, por exemplo, em “alarms go off” ou “blood circulates”. Contudo, o BBI não se preocupa em representar colocações L4 que sejam demasiado previsíveis como “bakers bake”.

- **L5** – nome + *of* + nome, “colony of bees”

As colocações L5 possuem dois nomes separados por “of”. São assim os exemplos de “a pack of dogs” ou “an act of violence”.

- **L6** – advérbio + adjetivo, “deeply absorbed”

Nas colocações L6 é identificada a combinatória entre um advérbio e um adjetivo como, por exemplo, em “anchor firmly” ou “appreciate sincerely”.

- **L7** – verbo + advérbio, “laugh merrily”

Finalmente, as colocações L7 são o último tipo de colocações lexicais representadas no BBI. Nas colocações L7, encontra-se a junção de um verbo com um advérbio, como acontece em “argue heatedly”.

Quanto à sua microestrutura, o BBI identifica as colocações pela sua palavra-chave, apesar de que, para Benson, esta não possui hierarquia sobre o colocativo. Benson defende a apresentação dos lemas pela palavra-chave das colocações lexicais, pois, para o autor, é através da palavra-chave que o usuário encontra mais facilmente o que pretende, em oposição à distribuição dos lemas pelos seus elementos verbais.

O que o BBI não apresenta são as combinações livres, definidas por Benson *et al.* (2010) como “elements that are joined in accordance with the general rules of English syntax and freely allow substitution” (p. xix). Desta forma, as combinações livres não apresentam uma ordem fixa de

interpretação devido à sua fluidez combinatória e por essa razão não são aludidas na metodologia do BBI.

O *BBI Dictionary of English Word Combinations* (2010) foi dos primeiros dicionários a ser eficazmente compilado como uma ferramenta de combinatória. A sua edição de 1997 inclui 18,000 entradas e cerca de 90,000 colocações. Sendo que a sua revisão de 2010 (Benson *et al.*, 2010), incorpora combinações pertencentes à esfera da tecnologia e da internet e também apresenta exemplos mais extensos e notas de uso com mais detalhe.

O BBI possui uma estrutura baseada no modelo Sentido-Texto de Mel'čuk. No entanto, o seu conteúdo não se baseia unicamente num *corpus*, mas é sim "corpus-refined" (McGee, 2012, p. 330). O facto de o BBI ser refinado por um *corpus* demonstra que o seu conteúdo é dependente da subjetividade do autor, pois o *corpus* é apenas utilizado como referência e não como o meio imperativo de apuramento de colocações. Benson não nega a frequência da coocorrência das palavras, mas também não fundamenta o conteúdo do BBI nesse método de extração. Inclusive, Benson *et al.* (2010) acrescentam que,

Nowadays, our task is eased not only by the availability of corpuses of contemporary English (such as the British National Corpus) but also by the amazing resource of the Internet itself, which enables us to search in it for a word and find superb examples of that word in context. Nor should it be forgotten that an important source of new information in BBI 3 is, paradoxically, BBI 2, now that the computer allows material from an entry in BBI 2 to be added to other entries in BBI 3 when such material is appropriate. (p. viii)

Sendo um dicionário sintático, como o DECFC de Mel'čuk, o BBI apresenta entradas com informação extensa, tanto ao nível verbal, nominal e adjetival. Por esta razão, o BBI é denominado explicitamente como um dicionário de combinatória. Benson *et al.* (2010) explicam que,

Traditionally, the combination of words into grammatical patterns has been called colligation or complementation or construction (though in BBI it is called collocation, too) and its result has been called valency. A dictionary that provides both phraseology and valency is a dictionary of word combinations; or, in the terminology of Igor Mel'čuk, whose work has inspired us, a combinatory or combinatorial dictionary. BBI is a combinatory dictionary. (p. vii)

Partindo para a análise detalhada da microestrutura do BBI, usaremos a entrada de "assignment" como exemplificada na Figura 13.

Figura 13

Artigo do lema *assignment*

assignment *n.* ["task"] ["mission"] 1. to give smb. an ~ 2. to accept, take on; refuse an ~ 3. to carry out, complete an ~ (the ambassador carried out her ~ brilliantly) 4. a dangerous; difficult, rough, tough ~ 5. an easy ~ 6. an overseas ~ 7. a rush; special ~ 8. an ~ to + inf. (an ~ to guard the president) 9. on ~ ("on a mission") (the correspondent was on ~ in the Far East) ["homework"] (esp. AE) 10. to give, hand out an ~ 11. to do; hand in an ~ (the pupils did their ~(s)) 12. a difficult; easy ~ 13. an ~ about, on ["appointment"] (esp. AE) 14. an ~ to (an ~ to a new job)

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 22).

A estruturação do lema "assignment", semelhante a todas as outras entradas do BBI, apresenta duas definições do nome, entre aspas e dentro de parêntesis retos ["task"] ["mission"] cujas explicações se estendem entre as aceções 1. e 9. A partir daqui, até à definição 13., são apresentadas aceções correspondentes a ["homework"]. Sendo que o artigo termina com a aceção 14., equivalente a ["appointment"].

Nas definições 1. a 8., é possível encontrar as colocações lexicais, e na definição 9. a expressão idiomática ("on a mission") é identificada com aspas dentro de parêntesis circulares. As expressões idiomáticas não são o foque do BBI, no entanto, Benson *et al.* (2010) procuram exemplificar aquelas cujo se assemelham a colocações, sendo que expressam um "simile" (p. xxxiv).

Quanto às colocações gramaticais, estas são identificadas por um tilde ~ ("the swung dash") como, por exemplo, na definição 8. com "an ~ to + inf". É importante referir que no BBI as colocações lexicais precedem sempre às colocações gramaticais.

As frases ilustrativas são dadas entre parêntesis, como é possível encontrar na definição 3. com "(the ambassador carried out her ~ brilliantly)".

A organização de cada definição é feita em forma de "strings", separadas por vírgulas, pois este método ajuda no aproveitamento dos espaços brancos.

Nesta entrada é possível ainda localizar a identificação de termos usados em Inglês Americano, com o uso de "(esp. AE)" nas definições 9. e 13. Ao longo do BBI, também é podemos encontrar a identificação de colocações em Inglês Britânico com a etiqueta de "(BE)", como no lema "about-turn", da Figura 14.

Figura 14

Artigo do lema *about-turn*

about-turn *n.* (BE) see **about-face**

Nota: Retirado do BBI (Benson *et al.*, 2010 p. 1).

As notas de uso, que providenciam informação adicional ao lema, não se encontram presentes especificamente nas figuras 13 e 14, no entanto no BBI estas são identificadas pela etiqueta “USAGE NOTE”, como na entrada do lema “able”, na figura 15.

Figura 15

Artigo do lema *able*

able *adj.* 1. perfectly ~ 2. ~ to + inf. (I’m perfectly ~ to cope on my own; she tried to reach him but she was not ~ to (reach him)) USAGE NOTE: In passive constructions, *able to* is replaced by *can* – X isn’t always able to replace Y = Y can’t always be replaced by X.

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 1).

No BBI também são apresentados verbos frasais como “act up”, ilustrado na figura 16.

Figura 16

Artigo do lema *act up*

act up *v.* (D; intr.) (“to function badly”) to ~ on (my leg has been ~ing up on me)

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 5).

Outro ponto importante da microestrutura é o uso da marca “(misc.)” – esta etiqueta lexicográfica identifica expressões idiomáticas consideradas pertinentes para a lógica do artigo de dicionário. Como exemplo, na figura 17, no artigo do lema “bush”, podemos encontrar na definição 3. a expressão idiomática “to beat about/around the”.

Figura 17

Artigo do lema *bush*

bush *n.* 1. to prune, trim a ~ 2. a clump of ~es 3. (misc.) to beat about/(AE) around the ~ (“to speak indirectly”)

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 52).

Quanto à informação sobre a pronúncia, o BBI não sugere nenhuma indicação. Somente no caso dos homógrafos é que o BBI suporta uma breve transcrição fonética para ajudar na distinção das

palavras. Como podemos ver na Figura 18, com os lemas “bow I”, “bow III” e “bow down”. Adicionalmente, o dicionário inclui uma listagem de “Simplified Transcription (ST) and IPA Transcription”⁸.

Figura 18

Artigos dos lemas *bow I*, *bow II*, *bow III*, e *bow down*

bow I /bau/ *n.* [“bending of the head or body”] 1. to give, make; take a ~ (the envoy made a ~ on entering the throne room; the actor took his ~) 2. a courtly; deep, low ~ 3. a ~ to [“debut”] 4. to make one’s ~ (he made his ~ as Hamlet) (compare **curtsy**)

bow II *v.* 1. (“to bend the head or body”) to ~ politely 2. (D; intr.) (“to bend the head or body”) to ~ before; to (to ~ before an emperor; to ~ politely to one’s host; to ~ to the inevitable) 3. (d; intr.) to ~ out of (“to abandon”) (to ~ out of politics) 4. (misc.) to ~ and scrape (“to be obsequious”) (compare **curtsy**)

bow III /bou/ *n.* [“device for shooting arrows”] 1. to draw a ~ (in order to shoot an arrow); to release a ~ [“decorative ribbon”] 2. to wear a ~ (in one’s hair) [“knot”] 3. to tie a ~; to tie in a ~

bow down /bau/ *v.* (D; intr.) to ~ before, to

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 45).

Podemos afirmar assim que Benson apoia a criação de ferramentas da combinatória porque, apesar de serem arbitrárias, as colocações evidenciam os diferentes aspetos sociais e culturais de uma língua.

Por esta razão, vemos que o conhecimento das colocações é vital para alunos de língua segunda. Apesar de existirem divergências em termos de potencialidade e representação das colocações no contexto lexicográfico, Benson *et al.* (2010) procuram apresentar uma ferramenta concisa, de cariz educativo, pois para Benson a dimensão idiossincrática das colocações merece ser estudada. E é através do BBI que o autor procura fornecer uma ferramenta monolíngue de combinatória lexical útil à produção do léxico.

⁸ Consultar Anexo 1.

3. O ensino e tratamento das colocações. Metodologia para um dicionário monolíngue de colocações portuguesas

3.1. A importância das colocações para a didática do Português Língua Estrangeira

A potencialidade das colocações para a didática de uma nova língua continua a ser uma vertente pouco explorada no ensino das línguas estrangeiras. Ainda assim, as colocações demonstram não só a liberdade de expressão do vocabulário, mas também refletem aspetos socioculturais inerentes ao léxico.

Segundo Thornbury (2002), o vocabulário de uma língua estrangeira apresenta determinados desafios que precisam de ser minuciosamente entendidos de forma a conter o léxico. Estes desafios compreendem-se através da exploração da forma e do significado das palavras. Para proceder à produção do léxico, Thornbury (2002, p. 2) revela que o aluno de língua estrangeira precisa de,

- acquire a critical mass of words for use in both understanding and producing language
- remember words over time, and be able to recall them readily
- develop strategies for coping with gaps in word knowledge, including coping with unknown words, or unfamiliar uses of known words

No caso do ensino de Português Língua Estrangeira (PLE), a combinatória das palavras mostra-se vital para o seu entendimento intrínseco. As colocações aspiram assim à integração dos alunos de PLE na sociedade portuguesa de uma forma inclusiva e homogénea.

As chamadas “lexical chunks” (Thornbury, 2002, p. 6) são onde se inserem os fraseologismos a que chamamos de colocações. Estas são os elementos mais desafiantes para um aluno identificar e absorver durante a aprendizagem. Thornbury (2002) dá o exemplo de “A record number of 54 teams will be competing in three sections as the Bryants Carpets Intermediate Snooker League gets underway this week.” (p. 7), onde as combinações recorrentes “record number” e “this week” demonstram a potencial dificuldade em compreender a complexidade lexical, como, por exemplo, no entendimento que na mudança dos fraseologismos referidos para “a record lot” e “that week”, respetivamente, a combinatória sofreria alterações em termos de significado e de contextualização.

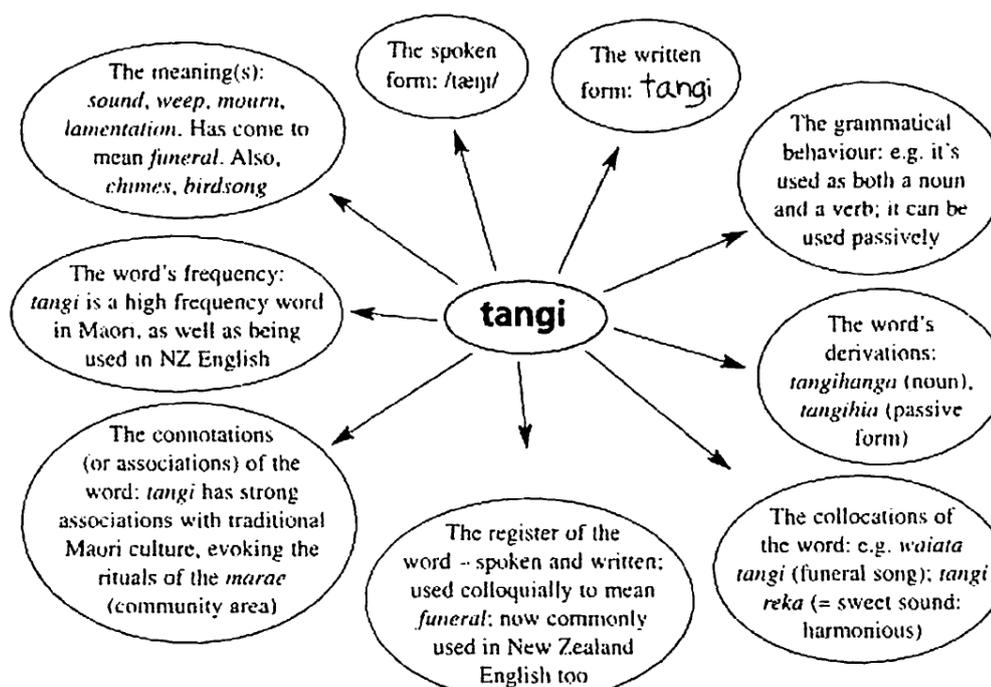
Averiguamos, através da nossa pesquisa, que na didática de PLE é oferecida pouca instrução aos professores em como ensinar as colocações ou como conduzir os alunos ao entendimento das mesmas. Desta forma, acreditamos que o que pode auxiliar ambos os lados da experiência educativa são as ferramentas especificamente construídas para tal, particularmente um dicionário monolíngue sobre a combinatória lexical portuguesa. Com o auxílio desta ferramenta lexicográfica, prevemos que o

aluno poderia aprender os conteúdos semânticos das colocações, pelo que estas possuem “semantic opaqueness and uniqueness of meaning” (Nation, 2001, p. 325). Sendo assim, consideramos deveras importante que os alunos compreendam os campos lexicais das palavras de forma a reter a temática do léxico.

O gráfico apresentado na figura 19 sugere a importância de *networking* entre itens lexicais e a sua semântica. Por intermédio desta prática, Thornbury (2002) defende a construção de um “mental lexicon” (p. 16). Para o autor, a forma mais eficiente em aprender uma língua nova é a de etiquetar mentalmente o vocabulário através da construção de *networks* léxicas. Por conseguinte, este tipo de exercício mental pode ser assistido por um dicionário de colocações que já saliente este tipo de relação entre palavras e conotações, expondo o usuário à versatilidade do vocabulário.

Figura 19

Gráfico temático relativo à palavra *tangi* do léxico maori



Nota: Retirado de Thornbury (2002, p. 16).

Apesar de defendermos a eficiência das ferramentas lexicográficas e a sua consequente “dictionary culture” em ambiente pedagógico (Hausmann 1989, p,13), o papel do professor como orientador de estudo não pode ser substituído pelas funções do dicionário. Juntamente com o auxílio de um dicionário e de exercícios educativos, o professor tem a possibilidade de proporcionar uma orientação que incentive à assimilação de novos conceitos e combinações. Acreditamos assim que a atmosfera de

uma sala de aula é o ambiente mais favorável à aprendizagem e ao estímulo da comunicação interpessoal usando nova terminologia.

Adicionalmente, pensamos que instruir professores a usarem a sua experiência pessoal com a análise de textos e consulta de dicionários é vital para o ensino de uma língua estrangeira. Thornbury (2002) sugere, como método de ensino, o uso de pequenos textos como o objeto ideal de estudo, pois estes “can be subjected to intensive grammatical and lexical study, without overtaxing learners’ attention or memory” (p. 53). Este tipo de textos especializados oferece uma padronização propícia ao conhecimento textual e lexical, que acaba por treinar os alunos em como futuramente interpretar textos maiores e mais complexos.

Ao encorajar a instrução dos alunos através da análise de textos, desenvolve-se a curiosidade do aluno em conhecer melhor a padronização da língua. Nesta fase de exploração, são as ferramentas de consulta que podem oferecer um esclarecimento acrescido. Idealmente, segundo Thornbury (2002, p. 81), este tipo de ferramentas ajuda em,

- providing an example situation
- giving several example sentences
- giving synonyms, antonyms, or superordinate terms
- giving a full definition

Em suma, um dicionário deste cariz é, por norma, organizado de forma simples e esclarecedora, sendo que, na nossa opinião, deve de se fundamentar num corpus contemporâneo. Pois, “High-quality dictionaries and successful consultation is a sine qua non for a proper dictionary culture and enthusiasm” (Tarp, 2018, p. 95).

No entanto, a consulta de um dicionário especializado em colocações depende da prática do aluno com ferramentas lexicográficas. Para Thornbury (2002), a aprendizagem não se pode fundamentar apenas na gramática, “Simply teaching learners a variety of structures, such as the *present simple* or the *second conditional*, is no guarantee that their communicative needs will be met” (p. 112), sendo que inclusivamente são as colocações que oferecem a proficiência lexical para um aluno comunicar como um nativo. Posteriormente, Wray (2000) refere que “gaining full command of a new language requires the learner to become more sensitive to the native’s preferences” (p. 463).

Na nossa opinião, o principal objetivo de um dicionário de colocações para alunos de PLE seria, essencialmente, a representação das combinações que beneficiam à fluência da língua. Este tipo de conteúdo estabelece o uso da combinatória com base na pré-fabricação da língua, ao que Cowie (2009) chama de “ready-made” (p. 49). Igualmente, Thornbury (2002) elabora as “lexical chunks” como elementos da pré-fabricação linguística,

A number of researchers have noticed that a lot of early language learning takes the form of chunks (such as *this-is-mine*, *give-me*, and *leave-me-alone*). These are acquired as single, unanalysed units. The capacity to use these chunks in conversational exchange seems to be an important factor in developing fluency. Using ‘pre-fabricated’ language, rather than using grammar rules to fabricate language from scratch, saves valuable processing time. These chunks are then stored away and only at a later stage of development are they analysed into their component parts. So, *this-is-mine* is eventually broken down into: determiner (*this/that*, etc.) + to be + possessive pronoun (*mine*, *yours*, etc.) (p. 144)

Seguindo a pré-existência da combinatória, o aluno acaba por criar o tal “mental lexicon” referido anteriormente, baseado numa “grammaticalised lexis” (Thornbury, 2002, p. 114). As chamadas *lexical chunks* de Thornbury (2002) devem, portanto, de ser alvo da didática pois nelas se inserem as colocações que exprimem a fluidez comunicativa.

As colocações são, segundo Pawley e Syder (1983), “the normal building blocks of fluent spoken discourse” (p. 208), que ajudam à naturalização do aluno na língua em estudo. Para além disto, a temática das colocações oferece um processamento intelectual menos exaustivo em relação, por exemplo, ao estudo concentrado na gramática.

Estamos convictos de que, com as colocações, o aluno desenvolve uma consciência semântica e contextual e é com a utilização das ferramentas lexicográficas integradas em atividades letivas⁹ que o ensino institucional de PLE pode ser potencializado. Seguindo as palavras de Nattinger (1988) acreditamos igualmente que “students will not have to go about reconstructing the language each time they want to say something” (p. 75) se integrarem o conhecimento das colocações na sua aprendizagem pois as colocações proporcionam continuamente “a marriage contract between words” (McCarthy 1990, p. 12).

Thornbury (2002, p. 122) refere então que o papel do professor pode ser formalizado pelas seguintes instruções,

- Become more aware of phrases and collocations yourself.
- Make your students aware of phrases and collocations.
- Keep an eye on usefulness and be aware of overloading the students.
- Feed in phrases on a ‘little but often’ basis.
- Introduce phrases in context, but drill them as short chunks.
- Point out patterns in phrases.
- Be ready to answer students’ questions briefly.

⁹ Consultar Anexo 2.

- Keep written records of phrases as phrases.
- Reinforce and recycle the phrases as much as you can.

Outra estratégia didática adicional poderia ser o uso de “lexical notebooks” como encorajamento à autodidática (Lewis, 2002, p. 49). A compilação destes cadernos de anotações do léxico pretende encorajar o aluno a organizar o seu próprio vocabulário incluindo a representação de “words, strong collocations, and fully fixed expressions with L1 equivalents” (Lewis, 2002, p. 76) que sejam relevantes para o seu contexto de aprendizagem.

Porém, a anotação de amostras do léxico possui determinados critérios que podem variar entre a frequência das combinações (Nation, 2001) e a coerência e/ou restrição das mesmas (Nesselhauf, 2003). Existe também a limitação causada pela tradução dos equivalentes das colocações, ao que Schmitt (2000) sugere que “we limit instruction to nontransferable collocates” (p. 89), pois “where collocations are similar between the first and second language, the learning burden will be lighter” (Nation, 2001, p. 56). Deste modo, a anotação do léxico em forma de cadernos lexicais pode não ser a melhor via de aprendizagem caso o aluno não possua conhecimento suficiente ou perspicácia para o entendimento destes critérios de anotação.

De qualquer forma, defendemos que um ponto importante a reter da aprendizagem, através das colocações, é a exposição do aluno às *chunks* da língua. Embora este particular processo seja laborioso, Wray (2005) insiste na exposição contínua dos alunos às colocações pois constrói uma “consciousness-raising” quanto ao léxico (Lewis, 2002, p. 121). Estas *chunks*, de preferência elaboradas a partir de textos especializados em diferentes contextos da língua, auxiliam na compreensão do léxico e no desenvolvimento do ensino prático-didático de PLE.

Desta maneira, um professor deve enfatizar a autoridade das ferramentas lexicográficas com variedade lexical. Consequentemente, as ferramentas lexicográficas deverão exibir uma boa representação geral da língua em estudo para se tornarem úteis para a consolidação da competência lexical dos aprendentes dessa língua. Isto ajuda o professor a cativar os alunos pelas diversas temáticas sociais e culturais, ao mesmo tempo que leva à exploração intelectual e individual de cada turma. Consideramos que uma turma de PLE deve de se inserir num ambiente de cooperação extra e interlinguística, com objetivos não só lexicais, mas também multiculturais.

Independentemente do seu nível de proficiência linguística, os alunos de PLE (e de qualquer outra língua estrangeira) apresentam maiores dificuldades não na compreensão das colocações, mas sim na produção das mesmas (Nesselhauf, 2003). Pois, segundo Corpas Pastor (2003), “Cuando se trata de comprender un texto, se han de tener en cuenta las colocaciones de las palabras para captar

sus sutilezas semánticas” (p. 171). Reconhecemos assim que, com exercícios didáticos juntamente com a análise contínua de uma ferramenta monolíngue de colocações portuguesas, os alunos de PLE conseguem passar menos tempo focados no aperfeiçoamento gramatical e empregar o que já conhecem da língua através comunicação por colocações, pois potencializa-se assim uma maior amplitude ao diálogo empírico em português.

3.2. Conceções gerais da metodologia

O cuidado e atenção prestados na compilação de um dicionário são dois elementos-chave neste tipo de trabalho de investigação. A compilação e a estruturação desta ferramenta exigem não só análise crítica, mas também uma perceção especializada ao nível da representação do léxico, como, por exemplo, referente a problemas associados à teoria lexicográfica atual, à contextualização da ferramenta em si e à sua viabilidade no mercado comercial. Como mencionado por Tarp (2018),

‘Language knowledge’ (conocimiento de la lengua) is a dual concept as it can both refer to language skills and learned knowledge of the language. This distinction is very important from a lexicographical point of view because dictionaries, on the one hand, may transmit learned knowledge, e.g. knowledge of the origin and grammar of a specific language, directly to their users while they, on the other hand, can only indirectly assist the development of language skills which necessarily have to be mediated by communication, [...] (p. 97)

De igual forma, durante a construção de um dicionário, é importante que seja criado um ambiente de trabalho onde a equipa lexicográfica tenha em consideração todas as ramificações da representação do léxico perante a comunidade didática, como potencial usuário. É, assim, imperativo estabelecer uma metodologia e uma planificação monitorizada de forma a colmatar todas as possibilidades de criação de um dicionário especializado.

A planificação caracterizada ao longo deste segmento não é vista como a única solução à problemática lexicográfica em questão, mas fundamenta-se na prioridade de uma metodologia multidimensional que pode ser usada para futuros projetos pedagógicos na esfera de PLE. Enfatizamos que o rigor desta metodologia visa à produção e não à receção da língua, pois, seguindo as palavras de Cop (1988), acreditamos que “Only the specialized, monolingual or passive bilingual collocation dictionary can deal with collocations for text production and for learning in a satisfactory manner.” (p. 44).

Primeiramente, é importante estabelecer o tipo de usuário do dicionário e em que ambiente de aprendizagem a ferramenta pode inserir-se. A viabilidade do dicionário no mercado comercial influenciará também o tipo de estruturação da ferramenta. Nesta fase, coloca-se então a questão de quem irá comprar e utilizar um dicionário monolíngue de colocações portuguesas? Tendo em conta axiomas da didática da língua portuguesa, reconhecemos que um dicionário deste tipo seria uma excelente fonte de conhecimento para alunos de português língua estrangeira.

Por conseguinte, este dicionário pretende focar-se no que é relevante ao nível das colocações lexicais portuguesas, especificamente para o perfil dos alunos de PLE. Analisaremos o tipo de colocações que devem de ser apresentadas, o seu contexto pedagógico e o tipo de representação nos artigos do dicionário.

O processo lexicográfico de compilação do material passa, segundo Zgusta (1971), pela coleção do próprio material seguida pela seleção dos lemas, pela construção dos artigos lexicográficos correspondentes e pela sua consequente organização macro e microestrutural. Cada um destes passos, igualmente defendidos por Landau (1984, 2001), enquadram-se na fase da planificação, da escrita e da produção do dicionário.

Para a formatação que a nossa metodologia segue, é irrelevante o formato do produto lexicográfico final, seja este em dicionário de papel ou eletrónico. Por esta razão, não estabelecemos filosofias quanto à sua viabilidade no mercado em consequência do seu formato final. O objetivo da nossa metodologia é que esta seja adaptável para qualquer tipo de formatação desejada, e que, acima de tudo, forneça uma boa representação das colocações portuguesas para alunos de PLE.

Contudo, quanto ao método de trabalho, acreditamos, como Wiegand (1999), que independentemente do formato final da ferramenta, o apoio da computerização é vital para a sua compilação. Inclusive, apesar de não defendermos uma específica formatação final do produto, gostaríamos de realçar que um dicionário eletrónico possui uma mutabilidade de acesso e de consulta que o dicionário em papel não suporta, seja o dicionário eletrónico do tipo “on a standalone Computer” ou “on a networked computer (intranet dictionaries and internet dictionaries)”¹⁰ (Klosa, 2013, p. 517).

Uma ferramenta eletrónica possui “an organic, changing database” (Prinsloo, 2001, p. 141), sendo assim um instrumento facetado e dinâmico. O seu concorrente em papel, é um projeto mais estático, pois, por consequência do seu formato, após a sua publicação, é mais dispendioso o processo logístico de avançar com novas edições do material. Por conseguinte, Klosa (2013) oferece a ideia de que “Online dictionaries under construction can be first published in print, then digitalized, and

¹⁰ Consultar Anexo 3.

continually enlarged (e.g. OED, DRW), or first published online (e.g. DDN, ELEXIKO).” (p. 519). Este tipo de fusão de formatos confere um processo colaborativo que incentiva à contínua evolução da ferramenta especializada.

Svensén (2009, p. 5) sugere a seguinte tipologia de planeamento do processo lexicográfico:

- Planning:
 - identifying the demand
 - preliminary project planning
 - pilot study
 - revising the project plan
 - preparing a specimen section
 - establishment of a final plan
- Implementation:
 - collection and selection of data
 - producing draft manuscript
 - revising draft manuscript
 - production

Este modelo de planificação e implementação fica submetido à organização do trabalho de equipa, que, idealmente, deve de ser composta por lexicógrafos, especialistas em linguagem computacional, linguistas e editores. A equipa deve, assim, possuir habilidades ao nível da apreciação do léxico e do entendimento didático da ferramenta pois,

Dictionaries being made the way they are, the lexicographer must also have a good analytical mind and a particular gift for being systematic. In the age of corpus lexicography, these abilities can be said to have become even more important, because lexicographic work now also requires the ability to survey and interpret the enormous quantities of linguistic material available. (Svensén, 2009, p. 8)

Consideramos assim que uma ferramenta de colocações, de apoio à aprendizagem de PLE, tem de ser capaz de cobrir linguagem contemporânea ao mesmo tempo que lidar com a descrição concisa do léxico, de maneira especial do ponto de vista da produção ou codificação linguística. Detalhadamente, esta ferramenta deve de proporcionar quais os elementos frásicos que se podem juntar a um determinado lema, de forma a demonstrar o binarismo entre palavra-chave e colocativo das combinações restritas, seguindo o modelo Sentido-Texto de Mel’čuk (1981). Adicionalmente, as colocações referentes a cada lema devem de apresentar um nível de especificidade relevante para, neste caso, alunos de PLE. Hausmann (1985) dá o exemplo da representação da colocação “fumador inveterado” que, por não ser uma colocação comum, o autor sugere que esta deve de ser representada

sobre o lema do substantivo “fumador” exatamente pela sua especificidade na língua. Desta forma, apoiamos que tal noção deve de ser inculcada na exploração da nossa ferramenta a favor da temática de PLE. Por esta razão, Svensén (2009) confere que “A good lexicographer must also be well informed generally, be interested in most things and have a great deal of curiosity.” (p. 8).

No entanto, um dicionário de colocações não deve de ser somente uma ferramenta de contextualização do léxico nem um espaço de apresentação de sinonímia pois,

This practice is dangerous in that the student will tend to transfer the typical contextual partners of a word to its synonyms. E.g. **redress** is a synonym of **remedy** — you can remedy a SITUATION but you cannot redress it, and **rectify** is a synonym of correct —you can correct a PERSON but you cannot rectify him. (Cop, 1988, p. 39)

Acreditamos que, para este tipo de função, a sinonímia deve de ser representada em forma de glossário, dependendo da sua relevância para o usuário.

Tendo em conta que estamos a lidar com uma ferramenta monolíngue, a acessibilidade do dicionário deve de ser associada a uma organização onomasiológica. Assim sendo, as palavras-chave de cada colocações são o que devem de formar a lemata do dicionário, pois o objetivo da ferramenta é o de responder à pergunta “O que pode ser feito com X?”, sendo X a palavra-chave de cada colocação, de forma a influenciar a produção da comunicação oral e textual portuguesa, pois o que a palavra-chave representa é fundamentalmente “what a person is writing or talking about” (Cop, 1988, p. 40). Deste modo, a equipa lexicográfica deve de possuir,

A very good knowledge of one's native language and a sound feeling for language in general is absolutely necessary. However, this ability must not only be a passive one: it is necessary also, particularly when working on monolingual dictionaries, to have no difficulty in expressing oneself clearly, concisely and correctly in writing. (Svensén, 2009, p. 8)

Em termos de contexto, existe assim uma “lexical-semantic constant” (Cop, 1988, p. 41), através da interpretação do léxico pela palavra-chave de cada colocação.

Um dicionário monolíngue especializado em colocações pode ser uma ferramenta de referência em conjunto com exercícios didáticos à base de *corpora*, pois textos empíricos de uma língua são o que exprimem melhor a atualidade do léxico, tanto a nível figurativo como a nível fraseológico. Desta forma, segundo Rundell (1998, citado por Jackson 2002),

The description of a language that a dictionary provides corresponds more closely to reliable empirical evidence regarding the way in which that language is actually used; the presentation

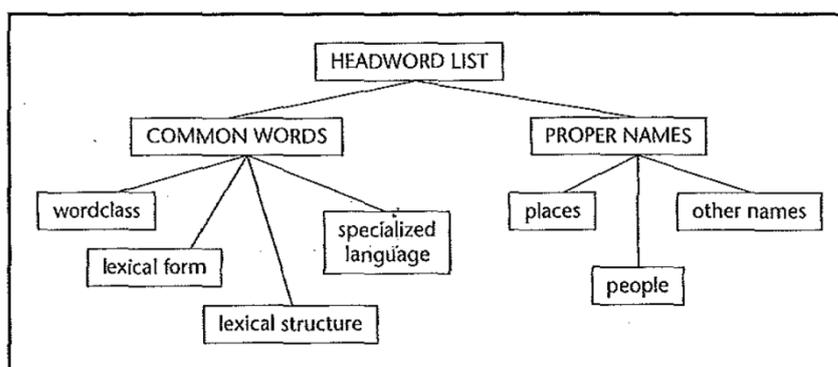
of this description corresponds more closely to what we know about the reference needs and reference skills of the target user. (p. 177)

Mas como se deve de proceder à anotação de *corpora* para a formulação de um dicionário? Esta etapa é uma tarefa que exige a especialização teórica e empírica da equipa lexicográfica. A anotação pode ser feita manualmente ou automaticamente, dependendo do tipo de extração mais viável para a ferramenta em causa. Neste caso, com a representação das colocações para alunos de PLE, a extração deve de ser feita e vigiada por nativos do português europeu em conjunto com especialistas em pedagogia de PLE. Neste processo consideramos elementar o balanço entre a frequência estatística do uso da combinatória e a sua contextualização cultural, que, a nosso ver, deve de se preocupar com níveis superiores a A2 do Português sendo que é a partir deste nível da língua que os alunos mais se interessam em adquirir conhecimento avançado do léxico.

Através do esquema da Figura 20, podemos argumentar que a extração das colocações de um *corpus* deve de ser realizada conforme a importância das mesmas para o público a que se destina. E por isso, a escolha dos lemas para o dicionário deve de ser feita a partir da sua relevância a favor da comunicação nativa do português.

Figura 20

Esquema ilustrativo dos fatores a serem considerados para a escolha de lemas



Nota: Retirado de Atkins (2008, p. 178).

Após a compilação dos dados, é importante verificar a anotação de erros. Estes erros podem ter origem na localização da colocação, caso a colocação seja incoerente quanto à sua semanticidade, ou então erros mais seletivos como na paradigmática da colocação (relativamente à palavra-chave), ou ao nível lexical sintagmático (no colocativo) – esta análise intrínseca de *corpora* é denominada por taxonomia, e o seu processamento pode variar dependendo do especialista que realiza a extração. Consideramos significativo nesta fase que o lexicógrafo esteja atento à análise da combinatória inerente nos *corpora*, e que faça um julgamento benéfico das colocações a serem representadas. Este processo

de extração resume-se então na identificação das colocações, na sua análise gramatical e lexical e na sua compilação pela perspetiva didática de PLE.

3.2.1. Corpora e software de extração

Um bom exemplo de um *corpus* de extração para a nossa metodologia poderia ser o *Corpus de Produções Escritas de Aprendentes de PL2* (PEAPL2)¹¹. Este projeto, iniciado em 2008, pelo Centro de Estudos de Linguística Geral e Aplicada da Universidade de Coimbra, é uma iniciativa que contém “um acervo de produções escritas” executadas por 458 alunos de PLE, desde os níveis de A1 a C1, faseadas primeiramente entre Maio de 2009 e de 2010, terminando em Maio de 2011. Deste projeto de *corpora*, resultou a representação de aspetos textuais empíricos por alunos de PLE que divulgam a sua conceptualização da língua portuguesa. O processo de extração de informação por nove estímulos contextuais a favor da produção do léxico.

Com a apresentação dos estímulos da Figura 21, os alunos puderam demonstrar a sua posição intelectual acerca da temática no léxico português, particularmente tópicos sobre o indivíduo, a sociedade, e o meio ambiente.

Figura 21

Tabela dos nove estímulos utilizados para favorecer a produção textual de alunos de PLE

Estímulo	Código do estímulo
O indivíduo	
Escreva um texto em que se apresente, em que fale das suas características físicas, da sua vida familiar, da sua casa, dos seus gostos e dos seus desejos. Se não quiser falar de si, pode inventar!	1.1A
Escreva uma carta a um amigo que não vê há muito tempo. Recorde momentos passados em conjunto e fale-lhe da sua vida pessoal e profissional actuais.	6.1B
Fale daquilo que gosta de fazer nos tempos livres.	33.1J
A sociedade	
Todos os países são diferentes a nível cultural e geográfico. Descreva o seu país, observando as particularidades das suas regiões, os principais monumentos e saliente alguns dos hábitos mais frequentes da sua cultura.	50.2L
Certamente já teve oportunidade de contactar com pessoas de cultura diferente da sua. Fale de um episódio que lhe recorde esse momento, das dificuldades sentidas, das diferenças e semelhanças encontradas entre as duas culturas e das experiências que partilharam.	52.2L
Há, certamente, comidas de que gosta muito e há outras que detesta. Fale disto e daquilo que pensam os seus familiares e amigos sobre o assunto.	55.2M
O meio ambiente	
Gosta de viver na cidade? Acha que, se pudesse, gostaria mais de vir no campo? Pense em vantagens e desvantagens de viver na cidade ou no campo. Escreva sobre isso.	69.3Q
Fale de meios de transporte. Fale daqueles em que já viajou e daqueles em que gostaria de viajar. Se quiser, pode contar uma viagem que tenha feito.	75.3S
Fale do bairro onde mora. Diga se gosta dele e se acha que há coisas que podiam mudar para que fosse mais agradável lá viver.	77.3T

Nota: Retirada de <https://www.uc.pt/fluc/rcpl2/metodologia> (Universidade de Coimbra, 2008).

¹¹ Disponível em <https://www.uc.pt/fluc/rcpl2>

Com um *corpus* como o PEAPL2, a equipa lexicográfica do dicionário poderia então proceder à recolha de dados da combinatória consoante o nível de PLE pretendido.

Em combinação com o PEAPL2, outro *corpus* prolífico que poderia ser utilizado nesta metodologia, seria o *Corpus de Referência do Português Contemporâneo* (CRPC)¹². Este banco de palavras, em expansão desde 1988, abrange uma esfera vasta de variantes do português contemporâneo de todo o mundo, e explora, atualmente, 311,4 milhões de palavras, combinadas entre textos de diferentes géneros, contendo ainda materiais de comunicação oral de diferentes registos.

Como defendemos a utilização de software e programas de extração automática, é pertinente nomear algumas das ferramentas que podem ser pertinentes nesta etapa. Nomeadamente ferramentas como o Sketch Engine¹³, ou então a consulta de listas de frequência monitorizadas online, como as disponíveis na Linguateca¹⁴. Porém, tendo em conta a exclusividade das colocações no estudo do léxico português, uma extração não pode depender apenas de software e/ou de listas de frequência já publicadas. A conjugação destas com uma extração manual e subjetiva por parte da equipa lexicográfica é o que consideramos ser o método mais viável a ser seguido.

Para Halliday (2004),

It is not difficult to pick out instances of text in which a given word occurs and to display these in the form of a concordance, showing their collocations on either side; the collocations can then be treated as single items and investigated quantitatively in their turn. All this information feeds in to the making of a dictionary. (p. 69)

Apesar de não ser um trabalho intelectualmente difícil de se realizar, a extração de colocações a partir de um *corpus* extenso é uma tarefa bastante dispendiosa em termos de tempo e também vulnerável ao (bom) trabalho do lexicógrafo.

O trabalho anotado resultante da extração do *corpus* deve de ser analisado a pensar na perspetiva dos alunos de PLE, pois o usuário é a referência principal na construção de qualquer dicionário. As colocações anotadas devem de ser assim interessantes para o contexto em que os estudantes se inserem. Assim sendo, neste passo, a teoria estudada deve de ser tida em conta, principalmente os diferentes tipos de definição de colocação assim como as suas classificações sintáticas. Após estabelecida a teoria e a anotação final das colocações, a equipa do projeto consegue então proceder ao planeamento macro e microestrutural do dicionário.

¹² Projeto de corpora contemporânea referente não só ao Português Europeu como a outras variedades da língua, como o Português Brasileiro e o falado em países PALOP. Disponível em: <http://ciul.ulisboa.pt/en/projeto/crpc-referencia-corpus-contemporary-portuguese>

¹³ Disponível em: <https://www.sketchengine.eu/>

¹⁴ Disponível em: <https://www.linguateca.pt/>

3.2.2. Macroestrutura

Um dicionário é um projeto em constante correção pois a língua encontra-se continuamente em evolução e crescimento. Posto isto, um dicionário deve de saber capturar o presente estado do léxico, mas também estar disposto a ser objeto de alterações futuras. O que influencia estas alterações futuras, neste caso, são os alunos de PLE pois eles são a bússola que orienta o nosso planeamento. Assim, a formatação do conteúdo e da apresentação da ferramenta é fundamental para a sua acessibilidade futura. Svensén (2009), baseando-se em Klosa (2001), discute que,

The production of electronic dictionaries will inevitably confront lexicographers with new kinds of demands. It will be necessary to rid oneself of several of the presentational limitations imposed by print-dictionary publishing, developing instead, for instance, a feeling for what words should be explained by means of pictures, animation sequences, video clips or audio recordings, what words should be explained by a combination of these media and what words are best explained by traditional means, i.e. by other words. This development will also require the acquisition of pedagogical and technical knowledge of a new kind. (p. 404)

A inteligibilidade de um dicionário é o que vai defender o seu uso. Achamos assim que os lemas devem de proporcionar uma visão esclarecedora e conceptualizada sobre as colocações portuguesas pois,

Because words constitute the largest section of the lexicon of many languages it is quite natural that the majority of macrostructural items should also be words, included as lexical lemmata. The extent of the selection of lexical items to be included in any given dictionary is determined by the typological criteria. (Hausmann e Wiegand, 1989, p. 330)

Assim sendo, é na introdução do nosso dicionário que deve de ser incluída toda a informação explícita quanto ao uso e organização do projeto. Igualmente, a simbologia lexicográfica deve ser corretamente explicada e a organização da ferramenta deve também estar justificada na introdução de forma que o aluno saiba navegar pelo dicionário, “In the front matter of a dictionary the lexicographer has to give a clear explanation of the macrostructural ordering of that specific dictionary” (Hausmann e Wiegand, 1989, p. 336).

Considerando as palavras de Hausmann e de Wiegand (1989), os dicionários apresentam com regularidade uma macroestrutura alfabeticamente ordenada, que se reflete igualmente na organização da microestrutura. Acreditamos que a alfabetização dos lemas, associados à temática de PLE, possibilita o acesso direto à estrutura do dicionário principalmente quando se trata de uma lemata restrita, como

é o caso das colocações. Este tipo de organização também favorece às práticas de “nitching” e de “nesting”, sendo que este modelo de apresentação é algo que defendemos pelo facto de enriquecer a coerência do dicionário.

Uma obra que pensamos que defende este tipo de apresentação é o dicionário BBI, de Benson *et al.* (2010). Na figura 22 é possível visualizar o tipo de formatação da sua lemata, na qual defendemos para a estruturação da nossa metodologia.

Figura 22

Excerto que exemplifica a alfabetização, o *nitching* e o *nesting* dos artigos de dicionário

A

aback *adv.* taken ~ (“startled”) (I was taken ~ at/by his angry words = I was taken ~ when he spoke to me angrily = It took me ~ when he spoke to me angrily)

abacus *n.* 1. to operate, use an ~ 2. with an ~

abandon I *n.* 1. great, reckless, wild ~ 2. with ~ (with reckless ~)

abandon II *v.* (D; tr.) to ~ to (they ~ed us to our fate)

abbreviate *v.* (D; tr.) to ~ to (*Esquire* can be ~d to *Esq.*)

ABC *n.* as easy as ~

abdication *n.* ~ from

abduct *v.* (D; tr.) to ~ from (to ~ a child from its home)

aberration *n.* a mental; sexual; statistical; momentary; temporary ~

abet *v.* (formal) (D; tr.) to ~ in (to ~ smb. in doing smt.; to aid and ~ smb. in doing smt.)

abeyance *n.* in; into ~ (to hold in ~; to fall into ~)

abhor *v.* 1. (D; tr.) to ~ for (she ~s him for smoking) 2. (G) she ~s smoking 3. (K) she ~s his smoking

abhorrence *n.* 1. to show ~ (she showed her ~ of smoking) 2. in, with ~ (of) (she holds smoking in ~ = she regards smoking with ~)

abhorrent *adj.* (formal) 1. to find smt. ~ (she finds smoking ~) 2. ~ to (smoking is ~ to her)

abide *v.* (d; intr.) 1. to ~ by (“to agree to, obey”) (we must ~ by her decision) 2. (obsol. and formal) (d; intr.) (“to stay”) to ~ with smb. (“In life, in death, O Lord, ~ with me” – H.F. Lyte (1793–1847))

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 1).

Hausmann e Wiegand (1989) analisam este tipo de organização da seguinte forma:

This kind of nested lemmatisation does not only have a space-saving function but it increases the ability of the user to interpret the presented information correctly. This is typical of a user-driven lexicography and illustrates an application of metalexigraphic developments. (p. 337)

Porém, esta estruturação é produto da subjetividade do lexicógrafo. Independentemente do estilo adotado pelo(s) autor(es) do dicionário, a sua organização deve de ser adequadamente explicada na introdução, como salientamos anteriormente. Em suma, o que traz sucesso a um dicionário é a sua praticabilidade, pois

One of the major challenges for the designers of dictionaries is to facilitate the users' quick and easy access to the relevant lexicographical data, cf. Tarp (2009a). Quick and easy access is one of the basic criteria for user-friendliness and quality. If it is too difficult and takes too long to find the needed data – and even worse, if the user gets lost *en route* – then the dictionary does not fulfil its purpose as an information tool. (Tarp, 2018, p. 102)

Sendo que a aglomeração dos artigos do dicionário demonstra a representação da macroestrutura da ferramenta, a microestrutura de um dicionário é então elaborada no interior de cada artigo representado. A nossa metodologia preocupa-se assim com uma macroestrutura alfabetizada, com aproveitamento de espaços brancos e de simples acesso para aprendentes de PLE, igualando ao método utilizado no BBI de Benson *et al.* (2010).

3.2.3. Microestrutura

Existem diversos componentes lexicográficos que podem beneficiar a microestrutura do dicionário.

A microestrutura que propomos não vai oferecer tradução dos lemas, visto que estamos a lidar com um dicionário monolíngue. Pretendemos explorar cada lema e as suas colocações associadas somente ao seu nível lexical e sintático.

Procedemos então à explicação dos componentes estruturais que acompanham a microestrutura da nossa metodologia. Estes elementos são imperativos pois identificam e organizam todo o conteúdo associado a cada lema e permitem a condensação da informação lexicográfica. Os indicadores a serem utilizados neste projeto podem ser múltiplos e por isso aconselhamos a sua listagem e explicação na introdução do dicionário, de forma a melhorar a interpretação da ferramenta. Estes componentes estruturais são chamados de indicadores tipográficos e não-tipográficos.

Segundo Gouws e Prinsloo (2005),

Typographical structural indicators are the different typefaces, e.g. bold, italic, roman, and the use of capitals, small caps, etc. in a dictionary. The function of these indicators is to mark specific search fields or data categories. In a monolingual dictionary one often finds the lemma sign to be presented in bold, the paraphrase of meaning in roman and illustrative examples in italics. (p. 2)

Por outro lado, os indicadores não-tipográficos são

[...] symbols and signs used to mark the beginning of a certain search field or data category and they play an important role in the inner access structure of a dictionary. Dictionaries employ different types of non-typographical structural markers, e.g. diamonds, triangles, squares, brief headings, etc. (Gouws e Prinsloo, 2005, p. 3)

Estes indicadores estruturam cada artigo na sua condição micro, de forma a potencializar a eficácia da própria macroestrutura e *design* do dicionário. Sendo assim, cada artigo da ferramenta pode conter características ao nível do comentário de forma e do comentário semântico.

O comentário relativo a questões formais distingue-se entre a ortografia, a pronúncia, a morfologia e a categoria gramatical do lema. O primeiro componente na construção do artigo é a ortografia do lema. Este comentário permite o usuário guiar-se corretamente pela ferramenta pois assegura não só a soletração do lema, mas também “holds the lexical form of the headword, showing how it is written, whether in a single word, a hyphenated word, or in several words” (Atkins, 2008, p. 204).

Relativamente à sua tipografia, defendemos a caracterização do lema por letras minúsculas e a negrito sem qualquer tipo de “wordbreaks” (Atkins, 2008, p. 204), ou seja, como referido anteriormente por Gouws e Prinsloo (2005), o uso de símbolos como “diamonds, triangles, squares” (p. 3) na separação das sílabas.

Quanto à pronúncia, consideramos facultativo recolher este comentário, tendo em conta que a nossa metodologia se direciona para a representação de colocações. Entendemos que o usuário que consulte o nosso dicionário já possua conhecimento sobre da pronúncia das palavras portuguesas, na perspetiva da receção do léxico. Salientamos que a representação IFA (o Alfabético Fonético Internacional) dos lemas é uma característica essencial que ajuda à comunicação oral direta, coisa que ultrapassa os limites do nosso projeto de dicionário especializado em colocações. A opção que sugeríamos, no entanto, seria a de representar somente os sons fonéticos na introdução do dicionário,

One possibility, found in many dictionaries, is to present the IPA in a front matter text along with a transcription of typical examples taken from the language treated in the dictionary. Some dictionaries also repeat the key to the IPA (or whatever pronunciation system is used) as a footer on every page of the dictionary. Thorough consultation should precede decisions regarding the presentation of pronunciation data. (Gouws e Prinsloo, 2005, p. 6)

Quanto à apresentação da informação morfológica de cada lema, como, por exemplo, a apresentação do plural e/ou de sufixos diminutivos, a nossa metodologia não considera tal comentário para o tipo de dicionário em causa.

Em contrapartida, a representação da categoria gramatical de cada lema é um elemento fundamental para qualquer dicionário. Todos os lemas da nossa ferramenta devem ser acompanhados pela sua categoria gramatical, seja ela substantivo, verbo ou adjetivo, assim como o seu género, se forem substantivos. A forma mais conhecida de recolher a informação sobre a categoria gramatical é a sua abreviação em itálico logo após a identificação do lema a negrito, como se mostra na Figura 23, nos lemas “acquaintanceship”, “acquainted” e “acquiesce” do BBI.

Figura 23

Artigos dos lemas *acquaintanceship*, *acquainted* e *acquiesce*

acquaintanceship *n.* 1. to strike up an ~ with 2. a casual; close, intimate ~ 3. an ~ with
acquainted *adj.* 1. casually; closely, intimately, thoroughly; personally ~ 2. ~ with (he got/became ~ with the situation; are you ~ with him?; I introduced her to a lawyer I was ~ with)
acquiesce *v.* (D; intr.) to ~ in, to (they ~d in the decision)

Nota: Retirados do BBI (Benson *et al.*, 2010, p. 4).

A microestrutura apresentada na Figura 23 revela uma familiaridade em relação à maioria dos dicionários no mercado, tornando-se acessível em termos de interpretação. Este tipo de estruturação familiar e comum é o modelo que defendemos para a nossa metodologia.

Terminando os comentários de forma, e continuando a utilizar o BBI como exemplo demonstrativo da nossa visão, na Figura 24, podemos ver que os comentários de variação de forma da palavra são assinalados por (BE) e por (AE), correspondentes a British English e a American English respetivamente.

Figura 24

Artigo do lema *ad*

ad *n.* (colloq.) 1. to place an ~ 2. to answer an ~ 3.
to follow, read the ~s 4. a classified, small (BE);
help-wanted (esp. AE); lonely-hearts; want (AE)
~ (see also **advertisement**)

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 5).

Tendo em conta que a nossa metodologia sugere uma ferramenta monolíngue em português, de preferência europeu, a informação sobre possíveis variações linguísticas não será considerada. Todavia, acentuar as colocações pelo seu âmbito formal ou informal, ou seja, o registo linguístico das combinações, seria um aspeto interessante a adicionar ao nosso planeamento visto que seria útil para a comunicação do usuário do nosso dicionário.

Partindo para os comentários semânticos, estes representam a informação de cada lema ao nível semântico e pragmático. Atendendo ao planeamento do nosso dicionário especializado, é significativo sublinhar que

One definition, in a monolingual dictionary, or one translation equivalent, with possible synonym equivalents where applicable, will constitute the dominant entries in the respective comments on semantics. However, where a lemma sign represents a polysemous word all the polysemous senses should fall within the scope of the treatment. Negotiating the different polysemous senses should be done in such a way that the user gets a clear indication of the variety of senses a given word has. (Gouws e Prinsloo, 2005, p. 11)

É assim, desta forma, que surgem os subcomentários semânticos que informam sobre a polissemia das palavras. Esta polissemia é normalmente separada por aceções numéricas que oferecem definições do lema e, no caso das colocações, o colocativo da colocação.

No exemplo da figura 25, o lema “acumen” apresenta três subcomentários semânticos, correspondente à polissemia do lema, representada pelos marcadores “1.”, “2.” e “3.”. Vemos este tipo de organização benéfica para o usuário, pois explicita a representação da polissemia. Ainda na mesma figura, podemos iluminar a distinção entre o seu comentário de forma e semântico pois, o comentário de forma consiste na palavra “acumen” a negrito e na sua classe gramatical “*n.*” em itálico.

Já o comentário semântico é simbolizado pelas aceções numéricas e por toda a restante informação lexicográfica que complementa o artigo. Sendo igualmente esta a forma como visualizamos o nosso dicionário.

Figura 25

Artigo do lema *acumen*

acumen *n.* 1. to demonstrate, display, show ~ 2. business, financial; legal; political ~ 3. the ~ to + inf. (she had enough ~ to see through the scheme)

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 5).

Outro ponto que enfatizamos é a contextualização dos lemas, através de “domain labels” (Atkins, 2008, p. 215). Sendo um dicionário de colocações a favor da produção da língua, vemos os marcadores que oferecem contexto às colocações significativos quanto ao nível da pragmática de PLE pois, “The *context* of a given word can be regarded as the pragmatic environment in which it is typically used.” (Gouws e Prinsloo, 2005, p. 13). A representação dos mesmos facilita a perceção das colocações, nomeadamente em que ambiente comunicativo se inserem. Mais uma vez, estas etiquetas lexicográficas (como as abreviações) devem ser clarificadas na introdução da nossa ferramenta lexicográfica¹⁵ e pressupomos que a seleção dos domínios no nosso dicionário deve estar relacionada a campos didáticos e temáticos do estudo de PLE, como, por exemplo, a literatura, as ciências, a política e a história do país.

Na figura 26 vemos que a marca lexicográfica “(mus.)” refere-se ao contexto musical o que auxilia o usuário a perceber que “to sing a canon” faz parte de uma linguagem especializada.

Figura 26

Artigo do lema *canon*

canon *n.* [“dogma”] 1. to establish, lay down a ~ [“round”] (mus.) 2. to sing a ~

Nota: Retirado do BBI (Benson *et al.*, 2010, p. 56).

Assim sendo, este tipo de representação lexicográfica é relevante para a nossa metodologia, pois auxilia no estudo do léxico português como língua segunda. Defendido por Gouws e Prinsloo (2005), Context entries like these assist the user in both text reception and text production but for text production purposes it is even better when a dictionary also offers cotext guidance by giving typical illustrative

¹⁵ Consultar Anexo 4.

collocations, phrases and example sentences to illustrate the way in which the specific word functions within the linguistic system of the language. (p. 14)

Os comentários semânticos, sejam em termos de contexto, estilística ou de cronologia, oferecem informação pragmática que faz a ligação entre a teoria e o uso empírico do léxico português.

Qualquer outro tipo de indicação lexicográfica, como a informação etimológica, as traduções equivalentes do lema ou até mesmo ilustrações pictóricas, são marcadores inoportunos para o nosso dicionário. Apesar de entendermos o quão importante é o desenvolvimento detalhado do comentário semântico para aumentar a eficácia da microestrutura, valorizamos que “the typological classification of the dictionary, the users, their needs and reference skills, the situation of dictionary use and the functions of the dictionary should influence the decisions.” (Gouws e Prinsloo, 2005, p. 21).

Por último, de forma a potencializar todos os elementos tipográficos e não-tipográficos e os diversos tipos de comentários de forma e semânticos, destacamos o desenvolvimento de uma detalhada introdução do dicionário. A nossa introdução necessita de conter um glossário que reflita o significado de todas as abreviações e elementos usados em cada artigo. Inclusivamente, seguindo os axiomas de Atkins (2008), o uso particular de um glossário pragmático seria uma boa adição para a interpretação das colocações, visto que

When inserting a pragmatic force gloss you must make it clear from the wording that it is not a simple explanation of meaning (i.e. a gloss proper) but an explanation of how the phrase is used to convey much more than its surface meaning. (p. 210)

Em suma, de forma a potencializar a microestrutura do dicionário, seguimos a implementação de uma introdução que consista num prefácio contendo toda a informação relativa à elaboração do dicionário, as suas normas de uso e de acesso assim como um glossário ou uma lista adequada de abreviaturas que explicita toda a simbologia presente na obra.

3.3. Considerações finais

Sendo a consulta de um dicionário uma forma de transmissão de conhecimento especializado, este tipo de ferramenta potencializa a compreensão não só da didática da língua, mas também do seu carácter socio-comunicativo.

Uma das obras mais exploradas ao longo da nossa pesquisa tornou-se numa referência importante para a construção da nossa metodologia. Referimo-nos ao trabalho de Benson *et al.* (2010), o *The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar*. Esta obra é a

que se assemelha mais ao que consideramos um planeamento ideal para um dicionário especializado em colocações lexicais portuguesas.

Em conjunto com a aplicação da teoria lexicográfica analisada, particularmente sobre o modelo Sentido-Texto de Mel'čuk e a formatação escolhida por Benson *et al.*, salientamos a nossa observação sobre dois *corpora* pertinentes à construção da metodologia lexicográfica em causa – o *Corpus de Produções Escritas de Aprendentes de PL2* (PEAPL2) e o *Corpus de Referência do Português Contemporâneo* (CRPC). Sendo estes dois *corpora* baseados em linguagem real característica aos aspetos atuais do léxico, entendemos cada um deles como significativos para o contexto da lematização em causa.

Considerando o nosso planeamento lexicográfico a favor de uma macroestrutura alfabeticamente ordenada em associação com uma microestrutura concisa, sem negligenciar a implementação de um prefácio instrutivo, o produto final da nossa metodologia poderia espelhar um bom caso de compilação e de formatação de um dicionário especializado em colocações portuguesas. Deste modo, a nossa ferramenta poderia ser alvo de práticas didáticas em ambiente de sala de aula de PLE.

Acreditamos que, para além do mapeamento da sua temática, a essência do conteúdo de um dicionário fundamenta-se nos seguintes ideais,

A dictionary article is an accessible dictionary entry characterized by the following three definitive features: (a) It shows at least one external access text element. (b) It is a constituent of a wordlist. (c) It consists of lexicographic data made accessible, including at least one such part, by means of which the dictionary user can unlock lexicographic information mentioned by the access text element. (Wiegand, Feinauer e Gouws, 2013, p. 317)

De forma a ilustrar a nossa trajetória metodológica, formulamos os seguintes exemplos da microestrutura que se enquadram na usabilidade do nosso dicionário como um veículo de estudo.

Exemplo 1:

amor *n. m.* ["sentimento de afeto ou atração"] 1. expressar ~ 2. cego; (lit.) cortês; fraternal; incondicional; livre; platónico; romântico; 3. ~ de, por ~, ter ~ (amor de mãe, por amor à arte, ter amor ao dinheiro, não ter amor à vida) 4. ~ à primeira vista 5. ~ entre (o amor incondicional entre eles) 6. com ~ (saudação final "Com amor, Maria") ["atividade sexual"] 7. fazer ~

Exemplo 2:

maduro *adj.* 1. emocionalmente; mentalmente; fisicamente; sexualmente ~ 2. ~ para (ela é ~ para a sua idade)

Exemplo 3:

marchar *v.* [“andar a pé”] 1. (intr.) ~ contra; por; em apoio (~ contra a guerra; ~ pela paz; ~ em apoio dos direitos das mulheres) 2. (intr., tr.) de ~ para; ~ em direção (o batalhão ~ ou desde a igreja até ao castelo) 3. (intr., tr.) ~ ao longo; ~ fora (~ ao longo do rio; ~ fora da sala de aula)

Com os referidos exemplos procuramos desenvolver a amplitude e as expectativas da ferramenta lexicográfica. Ao empregar o dicionário como um produto de massificação do conhecimento do léxico, defendemos que uma metodologia especializada como a nossa é capaz de normalizar questões multilinguísticas e de referência intrínsecas à pedagogia do português como língua segunda.

Conclusão

A representação sistemática e valorizada das combinações lexicais em ferramentas lexicográficas continua a ser uma aceção pouco explorada nos mais diversos léxicos, incluindo no português. De igual modo, o próprio estudo linguístico das colocações apresenta desafios quanto ao esclarecimento da diversidade da combinatória, nomeadamente nos campos da linguística, da metalexigrafia e da linguística informática.

Com o desenvolvimento desta dissertação, tornou-se clara a importância das colocações para o entendimento de qualquer língua e das suas ramificações socioculturais. A literatura estudada ao longo do nosso trabalho proporcionou uma visão mais intrínseca deste ramo da lexicografia, amplificando a potencialidade educativa das colocações lexicais.

Seguindo as palavras de Firth (1957), o autor refere que “you shall know a word by the company it keeps” (p. 11). Desta forma, Firth argumenta a ideia de que as relações entre palavras são indispensáveis para o conhecimento oral e escrito da língua. Por conseguinte, é do nosso entendimento que os dicionários representam um meio para elaborar a produção do léxico, explorando as suas características sintáticas e lexicais.

Uma ferramenta lexicográfica torna-se assim substancial na aprendizagem do léxico, pois um dicionário não é meramente um “commercial product compiled in a scissors-and-paste manner without any linguistic theory or at least theoretical backing coming from the realm of linguistics.” (Kleparski, G. e Włodarczyk-Stachurska, A., 2008, p. 85). Um dicionário abre as portas à competência didática e reflete como interpretamos a língua no nosso quotidiano.

Acreditamos que seria crucial para o nosso projeto apresentar uma metodologia que formulasse uma melhor representação das colocações no espectro dos dicionários portugueses especializados. A disciplina da Lexicografia contempla não só a teoria, mas também o lado prático da compilação de dados. A contextualização destas ferramentas é uma vertente na qual nos concentramos ao longo da nossa pesquisa, particularmente no aprofundamento dos axiomas didáticos do PLE. Sendo os dicionários ferramentas de utilidade linguística, consideramos que a compilação de um dicionário a favor de um determinado público confirma por si só a relevância do produto lexicográfico.

Hoje em dia, o ramo computacional permite a extração de grandes excertos de *corpora* do léxico, pois, com novo software, é possível uma compilação mais exata e menos intuitiva do léxico, de modo a favorecer uma melhor representação das colocações em dicionários. No entanto, é da nossa opinião que a colaboração da máquina com o conhecimento profissional e teórico do lexicógrafo e da sua equipa é

deveras importante para equilibrar a representação das colocações. Foi assim crucial para o nosso projeto apresentar meios que favoreçam uma compilação favorável à aprendizagem das colocações.

Apesar de admitirmos que existem mais e melhores correções a serem feitas para a metodologia descrita nesta dissertação, a finalidade da mesma preocupa-se em demonstrar as potencialidades insurgentes das colocações como elementos de estudo do léxico. Seguindo a normativa das colocações em conjunto com a indústria da produção de dicionários apuramos que é possível criar um espaço na didática atual que empregue o incentivo à *dictionary culture* (Hausmann, 1989) e que implemente o dicionário como uma mais-valia para o estudo de PLE e para o desenvolvimento extralinguístico.

Bibliografía

Aguilar-Amat, A. (1993). *Las colocaciones de nombre y adjetivo. Un paso hacia una teoría léxico-semántica de la traducción* (tese de doutoramento). Barcelona: Universitat Autònoma de Barcelona, Servei de Publicacions.

Atkins, S. e Rundell, M. (2008) *The Oxford Guide to Practical Lexicography*. Oxford University Press.

Barnbrook, G., Mason, O. e Krishnamurthy, R. (2013). *Collocation: Applications and Implications*. Reino Unido: Palgrave Macmillan.

Bartsch, S., Evert, S. (2014). *Towards a Firthian Notion of Collocation*. Mannheim: Institut für Deutsche Sprache.

Beneduzi, R. (2008). *Colocações substantivo + adjetivo: Propostas para sua identificação e tratamento lexicográfico em dicionários ativos Português-Espanhol* (tese de mestrado). Porto Alegre: Universidade Federal do Rio Grande do Sul.

Benson, M. (1986). Lexical Combinability. *Papers in Linguistics*. Edmonton, 18(1), pp. 3-15.

Benson, M., Benson, E. e Ilson, R. (2010). *The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar* (3ª edição). Amesterdão/Filadélfia: John Benjamins Publishing.

Cop, M. (1988). *The Function of Collocations in Dictionaries*. Retirado de Magay, T. e Zigány, J. (Eds.) *BudaLEX 88 Proceedings: Papers from the Euralex Third International Congress*. Budapeste: Akadémiai Kiadó, pp. 35-46.

Cop, M. (1991). *Collocations in the Bilingual Dictionary*. Berlim/Nova Iorque: De Gruyter.

Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.

Corpas Pastor, G. (2003). *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.

- Coseriu, E. (1973). *Sistema, norma e habla*. Madrid: Gredos.
- Coseriu, E. (1977, 1986). *Principios de semántica estructural*. Madrid: Gredos.
- Cowie, A. P. (1981, 1988). *The Treatment of Collocations and Idioms in Learner's Dictionaries*. Oxford University Press.
- Cowie, A. P. (2002). *English Dictionaries for Foreign Learners: A History*. Oxford University Press.
- Cowie, A. P. (2009). *Semantics*. Oxford: Oxford University Press.
- Cowie, A. P. (2009). *The Oxford History of English Lexicography*. Oxford: Clarendon Press.
- Cowie, A. P., Mackin, R. (1975). *Oxford Dictionary of Current Idiomatic English*. Oxford University Press.
- Ding, J. (2018). *Linguistic Prefabrication: A Discourse Analysis Approach*. Singapura: Springer Nature.
- Firth, J. R. (1935). *The Technique of Semantics*. Transactions of the Philological Society.
- Firth, J. R. (1951). *Modes of Meaning*. Indianápolis: Bobbs-Merrill.
- Firth, J. R. (1957). *Papers in linguistics: 1934–1951*. Oxford University Press.
- Firth, J. R. (1968). *A Synopsis of Linguistic Theory, 1930-55*. Retirado de Palmer, F. R. (Ed.), Selected Papers of J. R. Firth (1952-59). London: Longmans, pp. 168-205.
- Fontenelle, T. (1994). *What on Earth are Collocations? An assessment of the ways in which certain words co-occur and other do not*. Retirado de English Today, 10(4).
- Fraser, B. (1970). *Idioms within a transformational grammar*. Foundations of Language, pp. 22-42.

- Gouws, R. H. e Prinsloo, D. J. (2005). *Principles and practice of South African lexicography*. Stellenbosch: African Sun Media.
- Halliday, M. A. K. e Matthiessen, C. M. I. M. (1985, 1994, 2004, 2014). *Halliday's Introduction to Functional Grammar* (4ª edição). Londres/Nova Iorque: Routledge.
- Halliday, M. A. K., e Hasan, R. (1976). *Cohesion in English*. *English Language Series*. Londres: Longman.
- Hanks, P. (2008). *The Lexicographical Legacy of John Sinclair*. Oxford University Press.
- Hausmann, F. J. (1989). *Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart*. Retirado de Hausmann, F. J./Reichmann, O./Wiegand H. E./Zgusta, L. (Eds.), *Wörterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography I*. Berlim/Nova Iorque: De Gruyter, pp. 1-19.
- Hausmann, F. J. (1989). *Le dictionnaire de collocations*. Berlim/Nova Iorque: De Gruyter.
- Hausmann, F. J. (2004). *Was sind eigentlich Kollokationen?* Retirado de Steyer, K. (Ed.) *Wortverbindungen – mehr oder weniger fest. Institut für Deutsche Sprache Jahrbuch*, pp. 309-334.
- Hausmann, F. J. (2007). *Die Kollokationen Im Rahmen der Phraseologie – Systematische und Historische Darstellung*. Retirado de ZAA 55(3), pp. 217-234.
- Hausmann, F. J. e Wiegand, H. E. (1989). *Theory of Monolingual Lexicography I: Components and Structures of Dictionaries*. Berlim/Nova Iorque: De Gruyter.
- Hausmann, F. J. (2007). *Lexicographie française et phraséologie*. Retirado de *Collocations, phraséologie, lexicographie*. Aachen: Shaker Verlag, pp. 121-153.
- Iriarte Sanromán, Á. (2001). *A Unidade Lexicográfica. Palavras, Colocações, Frasemas, Pragmatemas*. Braga: Centro de Estudos Humanísticos-Universidade do Minho.

- Jackson, H. (2002). *Lexicography: An introduction*. Londres/Nova Iorque: Routledge.
- Jones, S., Sinclair, J. M. (1974). *English lexical collocations: A study in computational linguistics*. Paris: Cahiers de Lexicologie.
- Kleparski, G. e Włodarczyk-Stachurska, A. (2008). *Towards the Main Highlights in the History of Modern Lexicography*. Rzeszów: Wydawn. Uniwersidade de Rzeszów.
- Klosa, A. (2013). *The lexicographical process (with special focus on online dictionaries)*. Retirado de Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlim/Nova Iorque: De Gruyter.
- Landau, S. L. (1984, 2001). *Dictionaries: The Art and Craft of Lexicography* (2ª edição). Cambridge University Press.
- Léon, J. (2007). *Meaning by collocation. The Firthian filiation of Corpus Linguistics*. Retirado de Kibbee, D. (Ed.) ICHoLS X, 10th International Conference on the History of Language Sciences. Amsterdão/Filadélfia: John Benjamins Publishing.
- Lewis, M. (2002). *Implementing the lexical approach: putting theory into practice*. Boston: Thomson Learning Inc.
- Lipka, L. (1992). *An Outline of English Lexicology: Lexical Structure, Word Semantics, and Word-Formation* (2ª edição). Tübingen: Niemeyer.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- McGee, I. (2012). *Collocation Dictionaries as Inductive Learning Resources in Data-driven Learning: an Analysis and Evaluation*. International Journal of Lexicography 25(3), pp. 319-361.

McIntosh, A., Halliday, M. A. K. (1966). *Patterns of Language: Papers in general, descriptive and applied linguistics*. Londres: Longman, Green & Co. Ltd.

Mel'čuk, I. (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain*. Recherches lexico-sémantiques I-IV. Montreal: Les Presses de l'Université de Montréal.
Retirado de Giandomenico SICA (Ed.) (2006) *Open problems in Linguistic and lexicography*. Monza (Itália): Polimetrica.

Mel'čuk, I. (1978). *Théorie de langage, théorie de traduction*. Meta. Montreal, 23(4), pp. 271-302.

Mel'čuk, I. (1995). *Phraseemes in Language and Phraseology in Linguistics*. Retirado de Everaert M. *et al.* (Eds.), *Idioms: Structural and Psychological perspectives*, pp. 167–232. Hillsdale, NJ: Lawrence Erlbaum.

Mel'čuk, I. (1998). *Collocations and lexical functions*. Retirado de Cowie, A. P. (Ed.) *Theory, Analysis, and Applications*. Oxford University Press.

Nadais da Gama, B. S. (2009). *O léxico em aulas de PLE: Um contributo para o ensino das colocações* (tese de mestrado). Porto: Faculdade de Letras, Universidade do Porto.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Nova Iorque: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nattinger, J. R. e DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press.

Nesselhauf, N. (2003). *The use of collocations by advanced learners of English and some implications for teaching*. Retirado de *Applied Linguistics*, 24(2), p. 223-242.

Pereira da Costa, A. M. (2017). *As colocações em corpora de PL2: identificação, classificação e análise de erros* (tese de doutoramento). Lisboa: Faculdade de Ciências Sociais e Humanas- Universidade Nova de Lisboa.

Porzig, W. (1934, 1973). *Wesenhafte Bedeutungsbeziehungen*. Retirado de Schmidt, L. (Ed.) (1973) *Zur Geschichte und Theorie des sprachlichen Feldes*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Porzig, W. (1971). *Der Wunder der Sprache*. Tübingen: Francke Verlag.

Prinsloo, D. J. (2001). *The Compilation of Electronic Dictionaries for the African Languages*. Retirado de Lexikos 11, pp. 139-159.

Sinclair, J. (1966). *Beginning of the Study of Lexis*. Londres: Longman.

Sinclair, J. (1987). *Collins COBUILD English Language Dictionary*. Glasgow: Harper Collins.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sterkenburg, P. G. J. van (2003). *A Practical Guide to Lexicography*. Amesterdão/Filadélfia: John Benjamins Publishing.

Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge University Press.

Tarp, S. (2012). *Do We Need a (New) Theory of Lexicography?* Retirado de Lexikos 22 (AFRILEX-reeks/series 22), pp. 321-332.

Tarp, S. (2018). *New Experiences in Pedagogical Lexicography: Two Cuban School Dictionaries* Retirado de Miyares, L. R. (Ed.), *Estudios de Lexicología y Lexicografía: Homenaje a Eloina Miyares Bermudez*. Santiago de Cuba: Ediciones Centro de Linguística Aplicada.

Thornbury, S. (2002). *How to teach vocabulary*. Reino Unido: Pearson Education Limited.

Vygotsky, L. S. (1987). *Thinking and speech*. Retirado de Rieber, R. W. e Carton, A. S. (Eds.); Minick, N. (Trans.) *The collected works of L.S. Vygotsky*. Nova Iorque: Plenum Press.

Wiegand, H. E. (1999). *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin/Nova Iorque: De Gruyter.

Wiegand, H. E., Feinauer, I. e Gouws, R. H. (2013). *Types of Dictionary Articles in Printed Dictionaries*, Retirado de Gouws R. H. *et al.* (Eds.) *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (2013). Berlin/Nova Iorque: De Gruyter, pp. 314–366.

Wray, A. (2000). *Formulaic sequences in second language teaching: principle and practice*. *Applied Linguistics*, 21(4), pp. 463-489.

Wulff, S. (2008, 2010). *Rethinking Idiomaticity: A Usage-based Approach*. Londres/Nova Iorque: Continuum International Publishing Group.

Zgusta, L. (1971). *Manual of lexicography*. Paris: Mouton.

Webgrafia

- Corpus de Produções Escritas de Aprendentes de PL2 (PEAPL2), disponível em <https://www.uc.pt/fluc/rcpl2>
- Corpus de Referência do Português Contemporâneo (CRPC), disponível em <http://clul.ulisboa.pt/en/projeto/crpc-reference-corpus-contemporary-portuguese>
- Linguateca, disponível em <https://www.linguateca.pt/>
- Sketch Engine, disponível em <https://www.sketchengine.eu/>

Anexos

Anexo 1 – Comparative table of Simplified Transcription (ST) and IPA transcription

<i>Vowels</i>			<i>Consonants</i>		
<i>as in</i>	<i>ST</i>	<i>IPA</i>	<i>as in</i>	<i>ST</i>	<i>IPA</i>
beet	/biyt/	/bi:t/	pet	/pet/	/pet/
bit	/bit/	/bɪt/	bet	/bet/	/bet/
bet	/bet/	/bet/	tell	/tel/	/tel/
bat	/baet/	/bæt/	debt	/det/	/det/
balm	/bam/	/bɑ:m/, /bɑm/	cat	/kaet/	/kæt/
bob	/bab/; /ð/	/bʊb/, /bɑb/	get	/get/	/get/
boss	/bos/; /ð/	/bʊs/, /bɔs/	check	/ček/	/tʃek/
bought	/bot/	/bɔ:t/	jet	/džet/	/dʒet/
bull	/bul/	/bʊl/	fat	/faet/	/fæt/
boot	/bu:t/	/bu:t/	vat	/vaet/	/væt/
butt	/bæt/	/bʌt/	theft	/theft/	/θeft/
bird	/bə(r)d/	/bɜ:d/, /bɜrd/	then	/th:en/	/ðen/
ago	/ə'gou/	/ə'gəʊ/	self	/self/	/self/
bay	/bey/	/beɪ/	zoo	/zu:/	/zu:/
boat	/bout/	/bəʊt/	shoe	/šu:/	/ʃu:/
bite	/bayt/	/baɪt/	vision	/'viʒən/	/'viʒən/
bout	/baut/	/baʊt/	help	/help/	/help/
boy	/boy/	/bɔɪ/	men	/men/	/men/
beer	/biy(r)/	/bɪə/*	net	/net/	/net/
bear	/bey(r)/	/beə/*	ring	/rɪŋ/	/rɪŋ/
boor	/bu:(r)/	/buə/*	led	/led/	/led/
			yes	/jes/	/jes/
			wet	/wet/	/wet/

Nota: Tabela retirada do BBI (Benson *et al.*, 2010, p. xli).

Anexo 2 – Métodos e atividades pedagógicas em como ensinar colocações

Here are some more ideas for teaching **collocation**:

- ✎ Learners sort words on cards into their collocational pairs (e.g. *warm + welcome, slim + chance, golden + opportunity, lucky + break, mixed + reception*, etc). Use the same cards to play pelmanism (see page 97). Or they sort them into **binomial pairs** (pairs of words that follow a fixed sequence and often have idiomatic meaning such as *hot and cold, to and fro, out and about, sick and tired*). Or into groups, according to whether they collocate with particular 'headwords': e.g. *trip (business, day, round, return, boat), holiday (summer, family, public, one month, working)* and *weekend (long, every, last, next, holiday)*. Follow up by asking learners to write sentences using these combinations.
- ✎ Read out a list of words: learners in groups think of as many collocations or related expressions as they can. Set a time limit – the group with the most collocations wins a point. Good words for this include parts of the body (*face, head, back, foot, hand*), colours (*red, green, blue, black*, etc.) and opposites, such as *weak/strong, narrow/wide, safe/dangerous, old/young*, etc.
- ✎ Fill in a collocational grid, using dictionaries, to show common collocations. For example, here's a very simple (and completed) one for *wide* and *broad*:

wide	broad	
•		door
•	•	street
•	•	river
	•	smile
	•	shoulders
	•	nose
•		gap
	•	accent
•		world
•	•	range
•		variety
•		apart
•		awake

 In preparation for writing or speaking activities, learners can spend some time searching databases for useful collocations. Ask them first to brainstorm any nouns and verbs they are likely to need, and then to check for common collocates, using a concordance program (see Chapter 5) such as the COBUILD corpus on the Internet or a collocation dictionary (such as the *LTP Dictionary of Selected Collocations*), or simply a good learners' dictionary. Here, for example, are collocates and compound words for keywords selected in preparation for a composition on the subject of *flying*. They were all found using entries in the *Longman Dictionary of Contemporary English*:

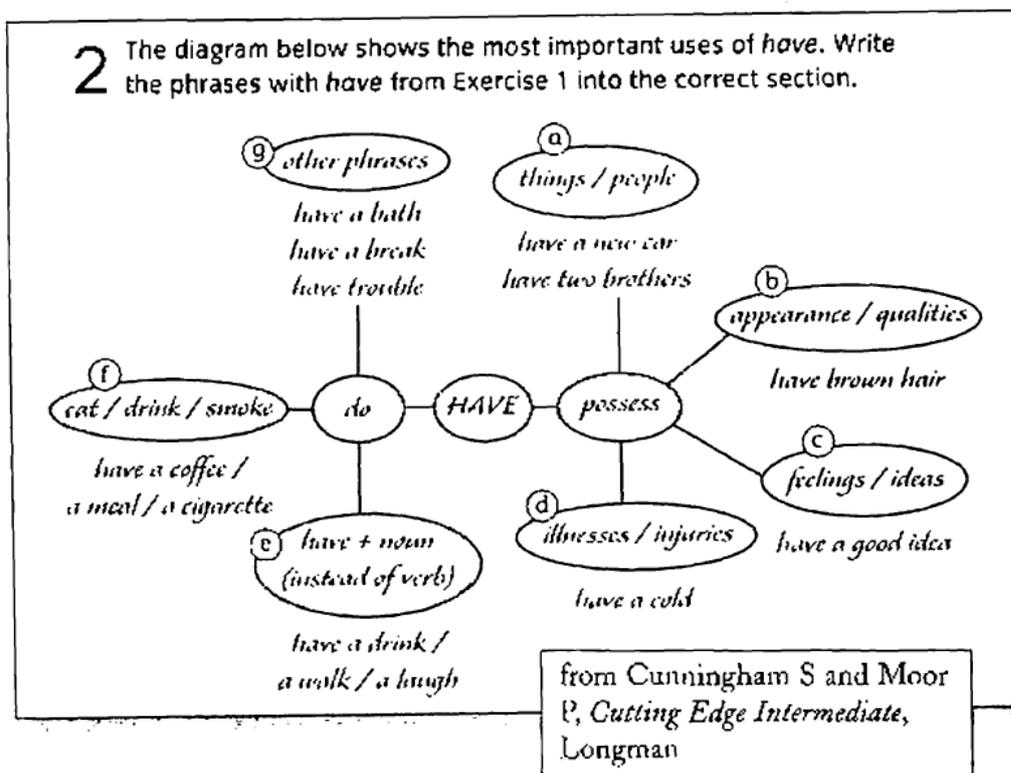
fly: fly direct, fly on to, fly economy class, fear of flying

flight: an hour's flight, my flight's been called, charter flight, flight attendant, flight path, flight recorder

air: by air, airborne, airbus, aircraft, aircrew, airfare, air hostess, airline, airplane, airport, airsick, air traffic controller

travel: travel by train, car etc, travel widely, travel around, travel light, travel the world; well-travelled, widely travelled

 Ask learners to prepare 'collocation maps' of high frequency words and their collocates. Words like *have*, *take*, *give*, *make* and *get* lend themselves to this kind of treatment. They are often used in combination with nouns to form an expression which has a meaning of its own, as in *have a look*, *take a break*, *give advice*, *make an appointment*, so that the verb itself has little or no independent meaning. For this reason, they are called **delexical verbs**. Here, for example, is a collocation map for *have*, which shows its range of collocations organised into meaning categories:



Learners can either create their own maps using dictionaries (or concordance programs – see page 70), or add to an existing map, as this task (also from *Cutting Edge Intermediate*) suggests:

3 a) Add the phrases below to the correct section of the diagram.

have a broken leg have a party have fun have a lot of energy
have a holiday have a meeting have a strange feeling have a wash

b) With which uses can you also use *have got*? What do you notice?

Teachers can exploit the fact that many film and book titles and names of pop groups are common collocations. Think of *Fatal Attraction*, *Desperate Measures*, *Deep Impact*, *The Usual Suspects* (all films) or *Dire Straits*, *Take That*, *Primal Scream* and *Public Enemy* (all names of groups). A search of the Internet quickly revealed that all the following collocations of *last* are names of bands: *Last Call*, *The Last Dance*, *Last Free Exit*, *Last in Line*, *Last Laugh*, *The Last Resort*, *Last Supper* and *Last Tuesday*. Learners can do the same – search for band names and check, using a dictionary, if they are common collocations or not. Alternatively, they could consult a dictionary, or a dictionary of collocations, in order to invent band names or film titles of their own.

Because of the two-part nature of collocations, any matching activities lend themselves to work on them (see page 97). Similarly, odd one out tasks are useful. For example:

What is the one word in each row that does *not* usually go with the word on the left?

win match war salary election race lottery

earn money degree living salary interest place

gain weight advantage access support wages experience

But there is a limit to the number of collocations that can be dealt with in activities like the ones above. The amount of time spent on targeting particular isolated collocations has to be balanced against time spent engaged in real language use, such as reading and speaking. It may, in fact, be in the context of real language use that the best learning opportunities will occur. A lot of work on collocation (and vocabulary generally) may happen in response to learners' errors. This reactive approach is described by Morgan Lewis:

Imagine a student produces *He's a strong smoker*. You could simply supply the student with the standard collocate – *heavy* – and move on. But an

ideal opportunity to activate language on the edge of the student's lexicon has been missed. It requires very little extra time or explaining to add: *occasional, chain* and *non* as more collocates of *smoker*.

(from Morgan Lewis in *Teaching Collocation*, LTP)

Finally, as a general approach to the teaching of lexical phrases and collocation, the following advice is sound:

- Become more aware of phrases and collocations yourself.
- Make your students aware of phrases and collocations.
- Keep an eye on usefulness and be aware of overloading students.
- Feed in phrases on a 'little but often' basis.
- Introduce phrases in context, but drill them as short chunks.
- Point out patterns in phrases.
- Be ready to answer students' questions briefly.
- Keep written records of phrases as phrases.
- Reinforce and recycle the phrases as much as you can.

(from *Cutting Edge Intermediate Teachers' Book*, Longman)

Nota: Retirado de Thornbury (2002, pp. 119-122).

Anexo 3 – Classificações de dicionários eletrônicos e de dicionários online

ELECTRONIC DICTIONARY			
Dictionary on a stand-alone computer (one user at a time)		Dictionary on a network computer (many users at a time)	
Handheld dictionary	Robust-machine dictionary	Intranet dictionary	Online dictionary
One user uses a palmtop to access a dictionary stored on a small disk	One user uses a laptop/desktop to access a dictionary stored on a CD-ROM	A group of users use laptops/desktops to access a dictionary stored on a local mainframe	Users worldwide use laptops/desktops/netbooks/mobile phones to access a dictionary stored on an online server

Tab. 26.1: Classification of electronic dictionaries (based on de Schryver 2003: 151).

CHARACTERISTIC FEATURES	DICTIONARY TYPES
Original form of publication	First published as printed dictionary
	First published as electronic offline dictionary
	First published as online dictionary
Completeness	Completed dictionary
	Dictionary under construction
Hypertextualization	Hypertextualized dictionary
	Dictionary without hypertextualization
Interaction with the user	Dictionary with interaction
	Dictionary without interaction
Multimedia	Dictionary with text, illustrations, charts, or diagrams
	Dictionary with text and audio files
	Dictionary with text, illustrations, charts, diagrams, and audio files
	Dictionary without multimedia
Dictionary access	Dictionary with access through scrolling in an entry list
	Dictionary with access via hyperlinked list of headwords
	Dictionary with access via search options
	Dictionary with combined access
	etc.

Tab. 26.2: Classification of online dictionaries (based on Storrer/Freese 1996, Storrer 1998, and Storrer 2001).

Nota: Retirado de Klosa (2013, p. 518).

Anexo 4 – Lista de abreviações significativas de etiquetas de informação lexicográfica

Abbreviations

<i>adj.</i>	adjective	<i>med.</i>	medicine, medical
<i>adv.</i>	adverb	<i>mil.</i>	military
AE	American English	<i>misc.</i>	miscellaneous
Am.	American	<i>mus.</i>	music
<i>anat.</i>	anatomical	<i>n.</i>	noun
BE	British English	<i>neg.</i>	negative
Br.	British	<i>obso.</i>	obsolete
CA	creation and/or activation	<i>occ.</i>	occasionally
CE	Common English	<i>pol.</i>	politics, political
<i>cf.</i>	compare	<i>pred.</i>	predicative
<i>colloq.</i>	colloquial	<i>prep.</i>	preposition
<i>comm.</i>	commercial	<i>refl.</i>	reflexive
<i>derog.</i>	derogatory	<i>rel.</i>	religion, religious
EN	eradication and/or nullification	RP	Received Pronunciation
<i>esp.</i>	especially	<i>smb.</i>	somebody
<i>fig.</i>	figurative	<i>smt.</i>	something
GA	General American	<i>subj.</i>	subjunctive
GB	Great Britain	(T)	Trademark
<i>imper.</i>	imperative	<i>tr.</i>	transitive
<i>inf.</i>	infinitive	<i>usu.</i>	usually
<i>intr.</i>	intransitive	US	United States
<i>ling.</i>	linguistics	<i>v.</i>	verb
<i>lit.</i>	literary	*	incorrect English
<i>math.</i>	mathematics		

Nota: Retirado do BBI (Benson *et al.*, 2010, p. xv).