

# The Compilation of an Online Corpus-Based Bilingual Collocations Dictionary: Motivations, Obstacles and Achievements

Adriane Orenha-Ottaiano

São Paulo State University (UNESP), São José do Rio Preto, Brazil  
E-mail: adriane@ibilce.unesp.br

## Abstract

This paper will discuss the motivations, obstacles and achievements of the building of an Online Bilingual Collocations Dictionary (English–Portuguese Collocations Dictionary and *Dicionário de Colocações Portugues–Inglês*). It is based on learner, parallel and online corpora and was designed for teachers and learners of English and Portuguese as a foreign language and learner and professional translators, among other users. With a view to fit in with the referred audience’s needs and profile, it focuses on all types of collocations (verbal, noun, adjectival and adverbial collocations), so that it may enable them to use collocations more accurately and productively, as well as boosting their collocational competence. The methodology firstly relied on the extraction and analysis of collocations from a Translation Learner Corpus and also of more collocational patterns extracted with the help of Sketch Engine (Kilgarriff et al., 2004) using a selection of the frequency lemma list (only content words, such as nouns, adjectives and verbs) from The Corpus of Contemporary American English (Davies 2008–2012). Being the first online bilingual collocations dictionary in the aforementioned language directions, we hope to incorporate collocational information more quantitatively and qualitatively as well as to achieve the challenge of meeting learners’ collocational needs in both languages.

**Keywords:** collocations dictionary; collocations; corpus-based dictionary; Portuguese as a foreign language; English as a foreign language.

## 1. Introduction

The aim of this paper is to report on the compilation of the *Online Bilingual Collocations Dictionary* (*English–Portuguese Collocations Dictionary* and *Dicionário de Colocações Português–Inglês*) as well as discussing its motivations, obstacles and achievements. The purpose of developing this collocations dictionary is a lexical pedagogical one. It is intended to help learners achieve collocational competence in the two focused languages, English and Portuguese. In this investigation, collocations are taken as pervasive, recurrent, arbitrary and conventionalized combinations, which are lexically and/or syntactically fixed to a certain degree and may have a more or less restricted collocational range.

The study of collocations has played a substantial role in Lexicology and Phraseology, and also in Lexicography or Phraseography, particularly in Pedagogical Lexicography

or Phraseography in the past few years. Innumerable work has advocated the building of dictionaries with a special focus on collocations or collocations dictionaries (Alonso-Ramos, 2006; Atkins & Rundell, 2008; Moon, 2008; Orenha-Ottaiano, 2013; 2016; Kilgarriff, 2015; etc.), owing to their relevance in learning and mastering a foreign language. As Fontenelle (2008: 12) noted “collocations are a hot topic in linguistics and in lexicography”. In addition, the combination of corpus linguistics, the use of corpora and corpus tools has enormously contributed to the investigation of collocations in Corpus Lexicography.

It seems to be a consensus among lexicographers that the quality of monolingual and bilingual dictionaries has improved significantly, due to the methodology provided by corpus linguistics, and the use of corpora has enabled us to identify and extract phraseological units more easily and effectively. However, even though corpus linguistics, and all the approaches and computational tools developed with it, have been of great help and interest to corpus and e-lexicographers, “they were not taken up as routine processes by lexicographers” (Kilgarriff, 2015: 83). According to Kilgarriff (2015: 83–84), “the first impression of a collocation list was a basket of earth with occasional glint of possible gems needing further exploration, and it took long to use them for every word”, and the solution to this problem was the creation of *Word Sketch*. Thus, considering that collocations are the central issue in this investigation, the use of Sketch Engine (Kilgarriff et al., 2004) and mainly of *Word Sketch* has become crucial to this investigation.

## 2. Motivations

The motivation for compiling an *Online Bilingual Collocations Dictionary* as proposed in this investigation lies in the fact that, as professors of a B.A. in Translation and a B.A. in the teaching of English Language, we readily recognize that collocations pose a serious problem, with regards to production (either oral or written), to foreign language learners as well as trainee and professional translators. The difficulties in combining words in order to produce the most frequently used collocations are clearly evident in these groups, mainly among learner translators. After having analyzed Stevick’s research (1989) on learner success in learning a foreign language, Wray (2002), sharing Pawley and Syder’s views (1983), concludes that the formulaic sequences used by native speakers are not easy for learners to identify and master, and that their absence greatly contributes to learners not sounding idiomatic. In that sense, having a collocations dictionary to provide them with collocational information may be of great help.

Additionally, we also have to take into account the scarcity of collocations dictionaries, especially bilingual ones. There are very good monolingual collocations dictionaries for learners of English as a foreign language, such as *Macmillan Collocations Dictionary for Learners of English* (Rundell, 2010), *Oxford Collocations Dictionary for Students of English* (Mcintosh et al., 2009), *LTP Dictionary of Selected Collocations* (Hill & Lewis,

1999) and *The BBI Combinatory Dictionary of English* (Benson et al., 1997). However, as for bilingual collocations dictionaries, specifically in the English–Portuguese or Portuguese–English directions, the proposed *Online Bilingual Collocations Dictionary* is, to my knowledge, the first.

Another motivation concerns the advantages of compiling an online collocations dictionary over a printed one. As Rundell (2013: 5) claimed “a dictionary accessed on a computer or a mobile device has great advantages over its analogue predecessors”. The author quoted Kilgarriff when he described Macmillan’s decision to stop printing dictionaries as “A day of liberation from the straitjacket of print”, as the researcher regards printed books as a not very efficient medium for reference materials. Rundell also predicted that the trend of online dictionaries was unstoppable, and mentioned that Macmillan’s focus only on digital dictionaries was “merely anticipating a move that all dictionary publishers will have to make eventually (and probably sooner than most people think)”. And he was right! Four years later, it has been “an even bigger game-changer than the arrival of corpora in the 1980s”! Hence, an Online English Collocations Dictionary will allow users to have access to a wider range of collocations, more examples and will offer the advantage of being constantly updated and revised, which would not be possible in a traditional and printed dictionary.

Besides practical issues regarding the need for compiling a dictionary of collocations, I also have a personal reason for having embarked on this phraseological enterprise. Being a learner of foreign languages, since I first came to know about collocations, I have always seen them as one of the greatest challenges in communicating in a foreign language and achieving fluency. Because of that, I have become a frequent and active user of collocations dictionaries. Nevertheless, I have always wished we had a bilingual collocations dictionary to help me with some of our tasks, especially one from Portuguese into English. The pleasure of compiling and having available online a Bilingual Collocations Dictionary is highly rewarding, mainly because I am sure we are selecting entries which indeed have a significant level of difficulty for Brazilian learners of English.

### **3. Methodology for extracting entries and collocations and The Corpus-based Online Bilingual Collocations Dictionary**

The *Online Bilingual Collocations Dictionary* is aimed at intermediate, upper intermediate, advanced and proficient learners of English and Portuguese as a foreign language (ranging from B1 to C2, according to Common European Framework of Reference for Languages).

We have been working on this project for over four years. So far, the dictionary contains more than 560 entries and more than 7,500 collocations, both in English and in Portuguese, all of them with their corresponding contexts (examples) (see Figure 3 below). Currently, more data are being extracted and inserted in the dictionary. It is

intended for the dictionary to contain at least 3,000 entries and more than 30,000 collocations by the time it is scheduled to be launched online in 2018, considering that we have now gained more expertise, have selected and managed a stronger team, not to mention the considerable help we have had from the Sketch Engine (Kilgarriff et al., 2004), which has definitely boosted our lexicographical work. We also have the goal of increasing this number from 7,000 to 10,000 in the following years, in order to meet the needs of more advanced learners.

As for the macro and microstructure of the dictionary, the entries (on the left side of Figures 1A and 1B) as well as the collocations (Figure 2) of the *Online Bilingual Collocations Dictionary* are displayed in alphabetical order:

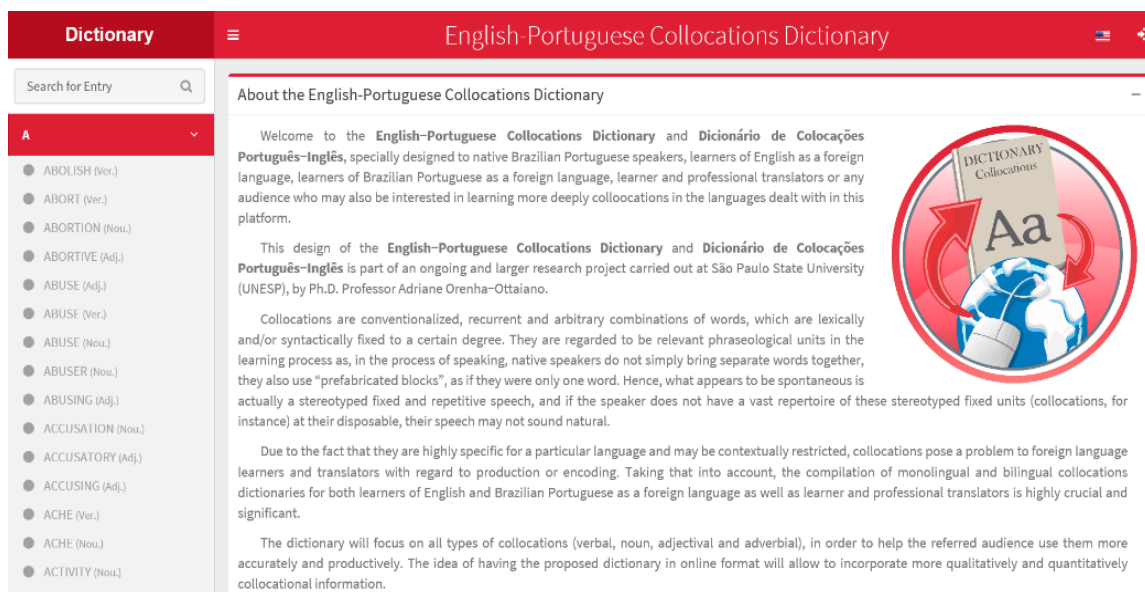


Figure 1A: Entries organized in alphabetical order (English-Portuguese direction)

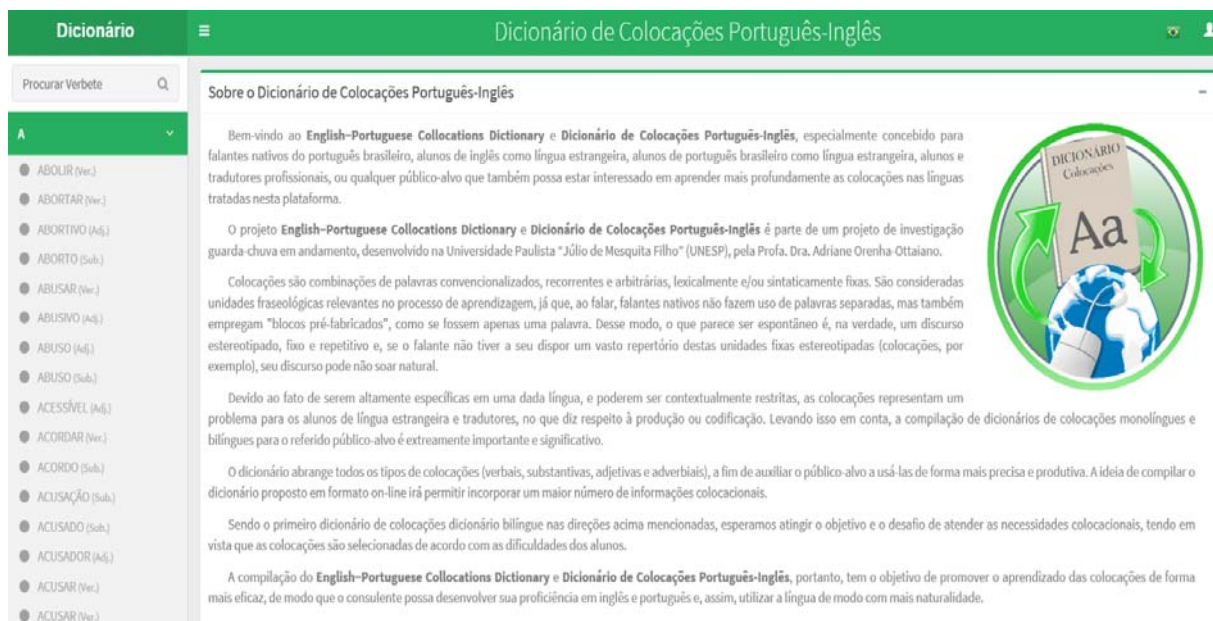


Figure 1B: Entries organized in alphabetical order (Portuguese-English direction)



Figure 2: Collocations in alphabetical order

The entries and collocations are being selected on the basis of frequency and were thus chosen:

- From the words with the highest keyness from a *Translation Learner Corpus*, of approximately 100,000 words. This corpus is made up of newspaper articles taken from newspapers and magazines written in Portuguese and translated into English by undergraduates<sup>1</sup> from a BA in Translation course, at São Paulo State University (UNESP), in Brazil. The decision to use a *Translation Learner Corpus* to extract entries and collocations lies in the fact that we wanted to make sure that collocations which are difficult to be produced by Brazilian learners of English as a foreign language would be included in the focused dictionary. The most frequent collocations produced by the translation learners and extracted from the keywords<sup>2</sup> were analyzed and those which were not suitably produced by the students were then replaced by the correct and most frequent ones, extracted with the use of Sketch Engine’s English Web 2013 (enTenTen13) corpus and were later included in the dictionary. It is worth mentioning that not only the collocations extracted from the *Translation Learner Corpus* were chosen for the dictionary, but also other patterns that were considered frequent in Sketch Engine’s enTenTen13 corpus<sup>3</sup>, with the use of Word Sketch<sup>4</sup>;

<sup>1</sup> Students’ level of English varied from B2 to C1 and their knowledge of language was identified according to the results of the *Oxford Placement Test* (Allan, 2004).

<sup>2</sup> At this point of the research, the computer tool used to generate keywords and concordance lines was WordSmith Tools (Scott, 2008). Currently, the Sketch Engine is used.


<sup>3</sup> The 2013 version of the enTenTen13 corpus contains almost 23 billion tokens.

<sup>4</sup> “The Word Sketch improves on standard collocation lists by using a grammar and a parser to find collocates in specific grammatical relations, and then producing one list of subjects, another of objects, etc, rather than a single grammatically blind list” (Kilgarriff & Tugwell, 2002: 25).

- From the most frequent words selected from *COCA*'s list (only content words). Again, the collocational patterns were extracted from the English Web 2013 corpus, with the use of Word Sketch. In due course, we intend to analyze the keywords from Sketch Engine's English and Brazilian Portuguese corpora;
- From the *New General Service List 1.01* (Browne, Culligan & Phillips, 2016), of approximately 2,800 words. This list was compared to *COCA*'s selected lemma list, so that more collocational patterns were extracted, using Sketch Engine, from the contrastive list, bearing in mind that, if students have access to collocations from the most common vocabulary for learners of English taken from the *New General Service List*, they may have more opportunities to improve collocational competence.

The aforementioned methodological steps to extract keywords and collocations are concerned with the compilation of the entries and collocations for the English–Portuguese direction of the *Online Bilingual Collocations Dictionary*. In what regards to the keywords and collocations of the Portuguese–English direction, the first measure was to translate all entries and collocations from English into Portuguese. At the moment, we are also comparing the translated entries to the keywords from Sketch Engine's Brazilian Portuguese Corpus (*Corpus Brasileiro*), with 1,133,416,757 words.

The organization of entries in a collocations dictionary can be done on two different concepts: either *node* and *collocate* (Sinclair, 1991) or *base* and *collocator* (Hausmann, 1985). In the *Online Bilingual Collocations Dictionary*, we followed the concept of *base* and *collocator*, which means that the lexical entries in this work are the *base*, taking into account that the *base* is usually what we already know and the *collocator* is the element we are looking for, that is to say, what the learners want to find out. We have chosen this approach as we consider it to be more user-friendly in the sense that, as Hausmann claims, users usually know the base and need to find out which words co-occur with it. For example, they may know the adjective *hot*, however, they may want to know which adverbs can be used with it when they mean it is, as Brazilian learners of English say “very, very, very hot”: native speakers of English would know they could use *boiling*, *scorching*, *stiflingly*, *unbearably*, etc. Having adjective entries is very productive in a collocations dictionary for Brazilian learners of English as they tend to use *very* for all adjectives, instead of more specific and straightforward ones.

As for the examples of the collocations displayed, if the user clicks on the collocation he or she is interested in learning, he or she will see the context in which this collocation is used. In case he or she wants to see the source of the examples for the selected collocation, the user may click on an icon , indicating the source, so that the dictionary gets visually cleaner, as shown in Figure 3, in the right:

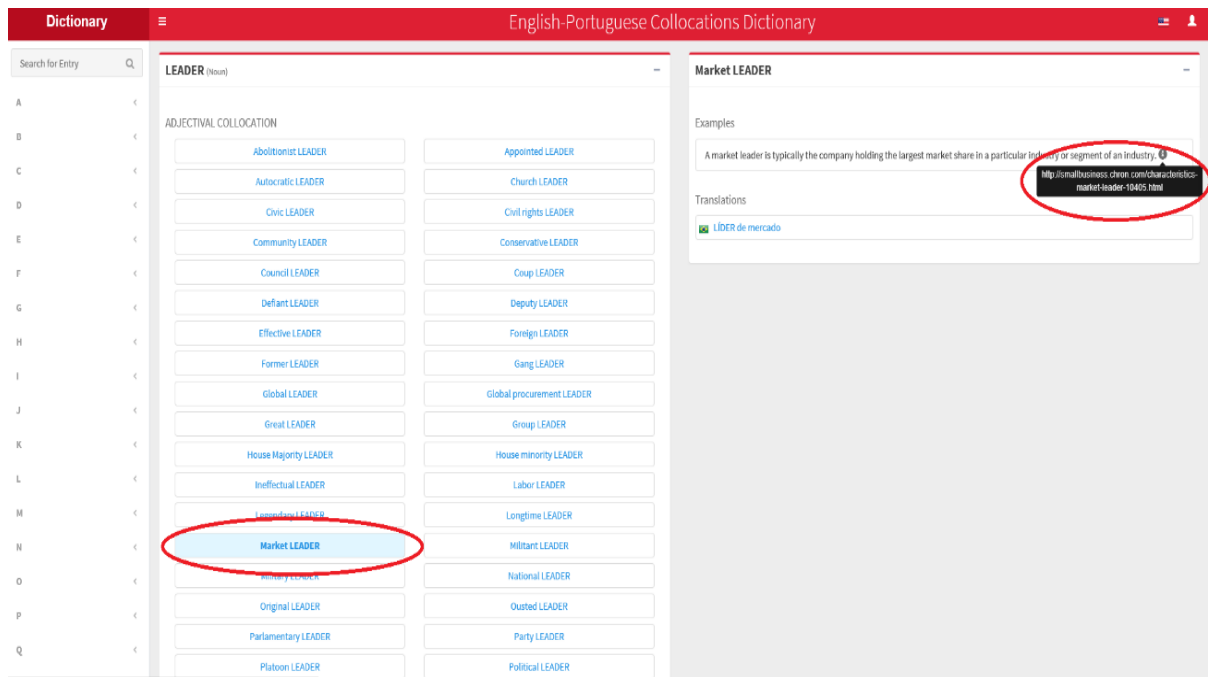


Figure 3: Icon indicating the source of the example

The screenshot from Figure 3 shows that the user clicked on one of the collocations (*market leader*), and an example of this collocation popped up on the right, as well as the equivalent collocation in Portuguese (“líder de mercado”).

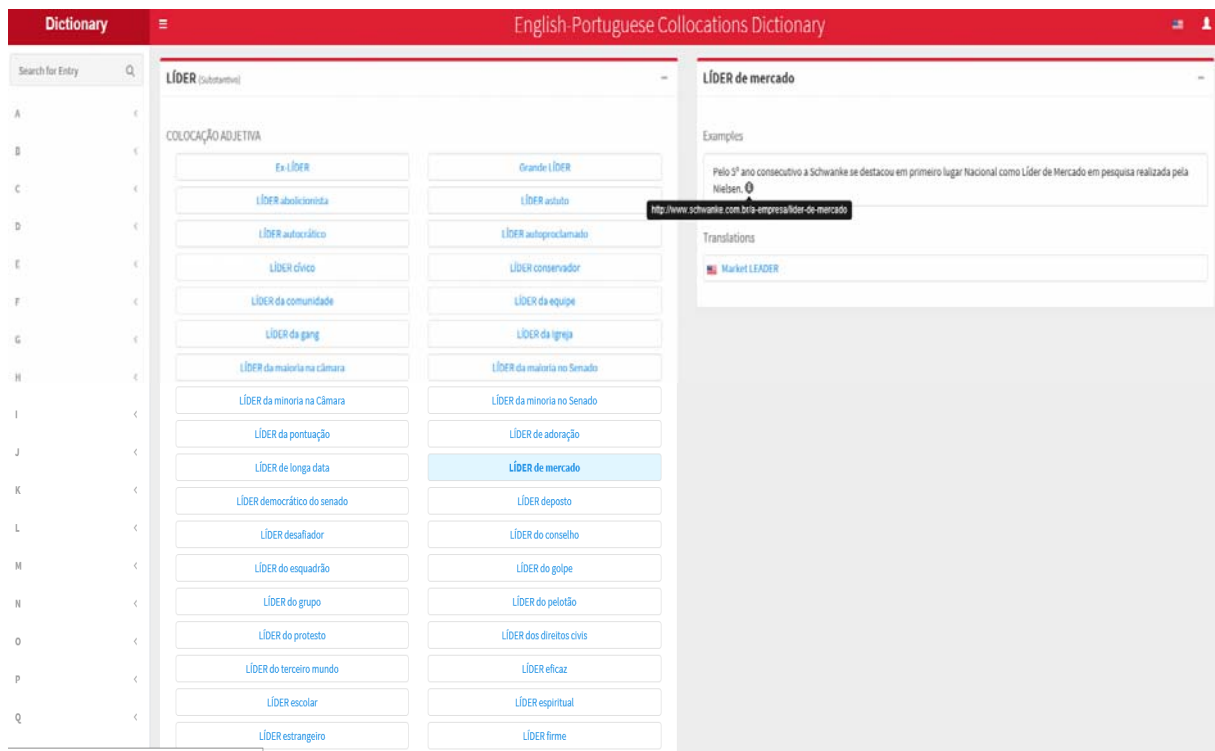


Figure 4: Screenshot of the equivalent collocation

If the user clicks on the translated collocation (“líder de mercado”), he or she will see the example of the equivalent collocation in Portuguese as well as all the possible collocations related to this entry in Portuguese and do the same search in this language:

The user will have the chance to choose the language from which he or she wishes to start, as shown in the top right position of Figure 5:

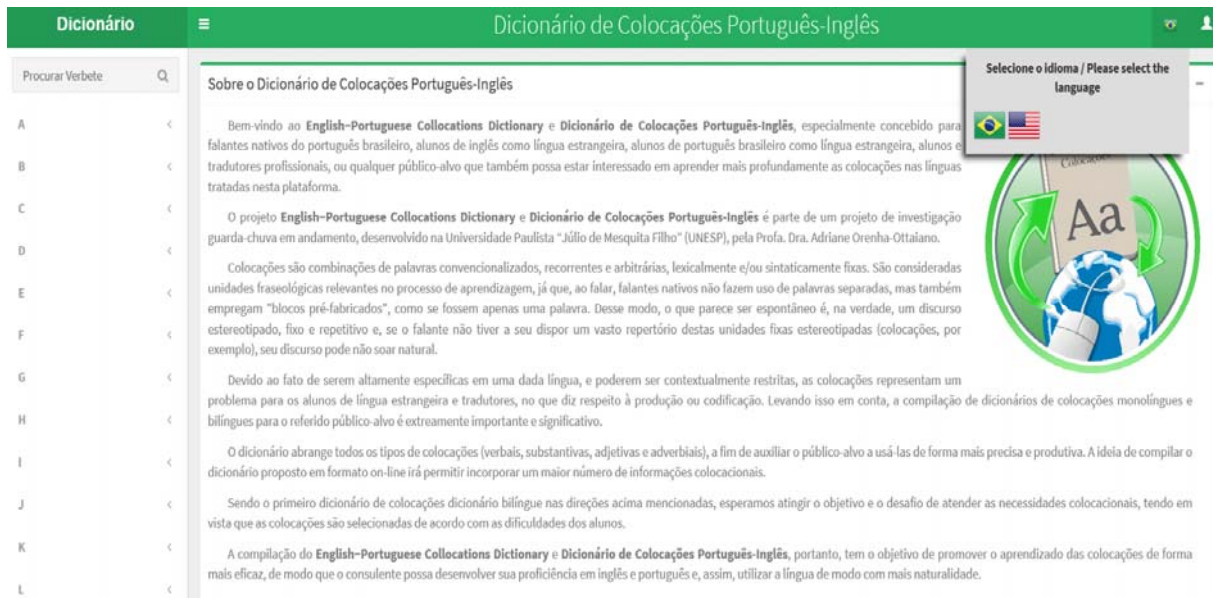


Figure 5: Language direction of the dictionary

As it can be seen from Figure 5, if you choose to use the Portuguese–English version, the screen will come up in green and blue. If the user chooses the English–Portuguese version, the color of the screen will appear in red and blue (as in Figure 6 below):



Figure 6: Screenshot of entry selected, types of collocations, and collocations extracted for the entry



According to Figure 6, one can also verify the types of collocations displayed from the entry interview: adjectival and verbal collocations, and the way they are arranged on the platform. As previously mentioned, the dictionary will bring all types of collocations: verbal, adjectival, nominal and adverbial, depending on the entry the user is focusing on. Here it follows the taxonomy of collocations we work with (Orenha-Ottaiano, 2004) expanded from Hausmann's classification:

**Verbal Collocations** – with four basic structures:

- Verb <sub>collocator</sub> + Noun <sub>base</sub>: acquire shares
- Noun <sub>base</sub> + Verb <sub>collocator</sub>: investments dropped
- Verb <sub>collocator</sub> + Preposition + Noun <sub>base</sub>: dispose of shares
- Verb <sub>collocator</sub> + Adverbial Particle + Noun <sub>base</sub>: set up a business
- Verb <sub>collocator</sub> + Adjective <sub>base</sub>: grow strong

**Nominal Collocations** – with two basic structures:

- Noun <sub>base</sub> + Noun <sub>collocator</sub>: share subscription
- Noun <sub>collocator</sub> + Preposition + Noun <sub>base</sub>: holder of shares

**Adjectival Collocations** – with one basic structure:

- Adjective <sub>collocator</sub> + Noun <sub>base</sub>: bearer shares

**Adverbial Collocations** – with three basic structures:

- Adverb <sub>collocator</sub> + Adjective <sub>base</sub>: fully eligible
- Verb <sub>base</sub> + Adverb <sub>collocator</sub>: drop dramatically
- Adverb <sub>collocator</sub> + Verb <sub>base</sub>: fully paid; duly appointed

Based on the macro and microstructure presented in this session and the screenshots of *The Online Bilingual Collocations Dictionary*, it is evident that the dictionary is thus working, but remains under test and is subject to changes.

## **4. Obstacles and achievements in the compilation of the Online Bilingual Collocations Dictionary**

*The Online Bilingual Collocations Dictionary* is the first bilingual English–Portuguese and Portuguese–English collocations dictionary, and this may be seen an achievement, particularly to the Brazilian Portuguese field of lexicography, considering that lexicographic work in Brazil still requires a lot of support, investment and recognition in research.

When we first proposed a bidirectional collocations dictionary, we were not aware of

the difficulties we would encounter. As we mentioned above, we made the decision to translate all the entries and collocations from English into Portuguese. When we finally had the online platform for the dictionary and inserted the translations, we realized we could have results which would conflict with the theoretical framework of a corpus-based work.

If we consider the direction English–Portuguese, there would not be any problem to translate all the entries and collocations from English into Portuguese, as that would be the natural procedure to carry out. However, if we take into account the Portuguese–English direction, as it is already displayed on the platform at this phase of the project, when we first visualize the entries and collocations in Portuguese, as researchers, we may question whether these are frequently used words and collocations in Portuguese, as they came from the translations.

To resolve this problem, we have compared the translated entries to the keywords from Sketch Engine’s Brazilian Portuguese Corpus (*Corpus Brasileiro*), as was the original idea, in the sense of having entries and collocations in Portuguese which are frequently used by native speakers of Brazilian Portuguese. By doing so, we intended to make sure learners could master a foreign language by having contact with the lexicon that is considered to be frequently spoken by Portuguese speakers. As for the translated collocations, we have been comparing them to the collocations displayed by Word Sketch from Sketch Engine’s Brazilian Portuguese Corpus (*Corpus Brasileiro*).

Having analyzed the online Portuguese–English direction dictionary, in comparison to the online English–Portuguese one, it seems as if they are two different dictionaries. If we delete the entries or collocations which are highly frequent in Portuguese or if we add up new entries or collocations that may be considered to be commonly used in Portuguese, according to the results of the Brazilian Portuguese Corpus, we would have to do the same with the equivalent ones in English. In this respect, we will have to take some measures, so as to better deal with this difficulty. The interface will have to be adjusted with a view to allowing access to an entry from a specific language without conflicting to the entries of the other one. And that is where the problem lies and obstacles have to be overcome. Another possible solution is dividing the dictionary and creating two bilingual dictionaries, in only one direction, instead of a one bidirectional dictionary, as presently proposed: An English–Portuguese Dictionary and a Portuguese–English Dictionary. This way, we would make sure users may have access to the most common collocations in each direction, also taking into account that the translations were carefully given, in the sense that we tried to find the most possible frequent equivalent in the target language.

Another obstacle we faced, mainly in the initial phase of the project, was not having a good team to rely on: and that is crucial when it comes to a phraseographical work. To carry out this project we needed to count on students with advanced or proficient knowledge of the language, so that they could identify collocations. Besides that, we

also needed the help of advanced or proficient learner translators to deal with the translation of the entries and collocations. The greatest problem was recruiting this team, as some students were not committed to research and, resultantly, we could not trust their work and ended up by doing everything ourselves. At the beginning of 2017, we started training a more committed team involving three master's students, one Ph.D. student, six undergraduate students (four of them from B.A in English Language and two from B.A. in Translation). As our aim is also to promote the teaching of collocations in public schools, two high school students from a public school will soon join the project, having applied for a scholarship. However, although we have now built a larger team, it is still necessary for the researcher in charge to undergo considerable revisions and reviewing of hundreds of collocations as well as their translations.

Another point worth mentioning regards the organization of the entries, as it has become a considerable difficulty when extracting collocations. As previously mentioned, the entries were compiled anchored in Hausmann's concept of *base* and *collocator*. When some members of the team were extracting collocational patterns, we realize we could have some problems. For instance, let us consider the keyword *vacancy*. When analyzing some of the collocations extracted, such as *job vacancy*, *senate vacancy*, *vacancy announcement*, and *vacancy rise*, we noticed that, based on Hausmann's concept, *vacancy announcement* and *vacancy rise* would not be inserted under the entry *vacancy*, but on the entries *announcement* and *rise*, as they are considered to be the base. Another very good example refers to the entry *blood*, from which the following collocations were extracted: *blood samples*, *blood evidence*, *blood pressure*, *blood stains*, *blood pool*, *blood trail*, *blood spatter*, *blood void*, *blood type*, *clotted blood*. If we apply Hausmann's theory, the only collocation for the entry *blood* would be *clotted blood*. The other collocations would have new entries: *samples*, *evidence*, *pressure*, *stains*, *pool*, *trail*, *spatter*, *void*, *type*.

Thus, we would have to analyse these new candidates for entries (*announcement*, *rise*, *samples*, *evidence*, *pressure*, *stains*, *pool*, *trail*, *spatter*, *void* and *type*), in order to check if they are frequently used, so that they can be included in the dictionary. That has become an obstacle as, besides having to extract collocations from the big list of keywords we have previously mentioned (the *Translation Learner Corpus keyword list*, and the contrastive list resulted from the comparison of COCA's selected lemma list and *New General Service List 1.01*), we would also have to spend a considerable amount of time extracting more collocations from these new entries and our work may end up by being endless. This way, it would be advisable to reconsider Sinclair's concepts if the change indeed favors the extraction of the most relevant collocations as well as addressing learner needs.

With respect to the translation process of collocations, some collocational types have posed a challenge in translating from English into Portuguese, mainly, and mostly, the adverbial collocations, both the Adverb + Adjective and Verb + Adverb construction.

Adverbial collocations are not as frequent in Brazilian Portuguese as they are in the English language and that is the reason why it turns out to be challenging to translate them or to find a correspondent collocation. Many adverbial collocations presented difficulties in translation, such as *hopelessly naive*, *increasingly political*, *fall sharply*, *rightly fear*, *hard-core unemployed*, etc., indicating that the translation of adverbial collocations may be a challenge in phraseographical work on collocations, in the English–Portuguese direction at least.

For instance, the collocation *increasingly irrelevant* was considered hard to translate into Portuguese. The translation options given were “extremamente irrelevante”, “completamente irrelevante”, “totalmente irrelevante” and “absolutamente irrelevante”. Although the four translation options proved to have a high co-occurrence frequency according to Google statistics, none of them describes the idea of *becoming larger (increasingly)*. An alternative translation is the use of three words before the noun, “cada vez mais irrelevante”, like the dictionarized meaning of *increasingly: more and more*, because in Portuguese we have not found a single adverb. Some translations therefore had to be carefully considered as the Word Sketch results for the adjective “irrelevante” were not satisfactory and did not suggest a more suitable equivalent in Portuguese, as shown in Figure 7:

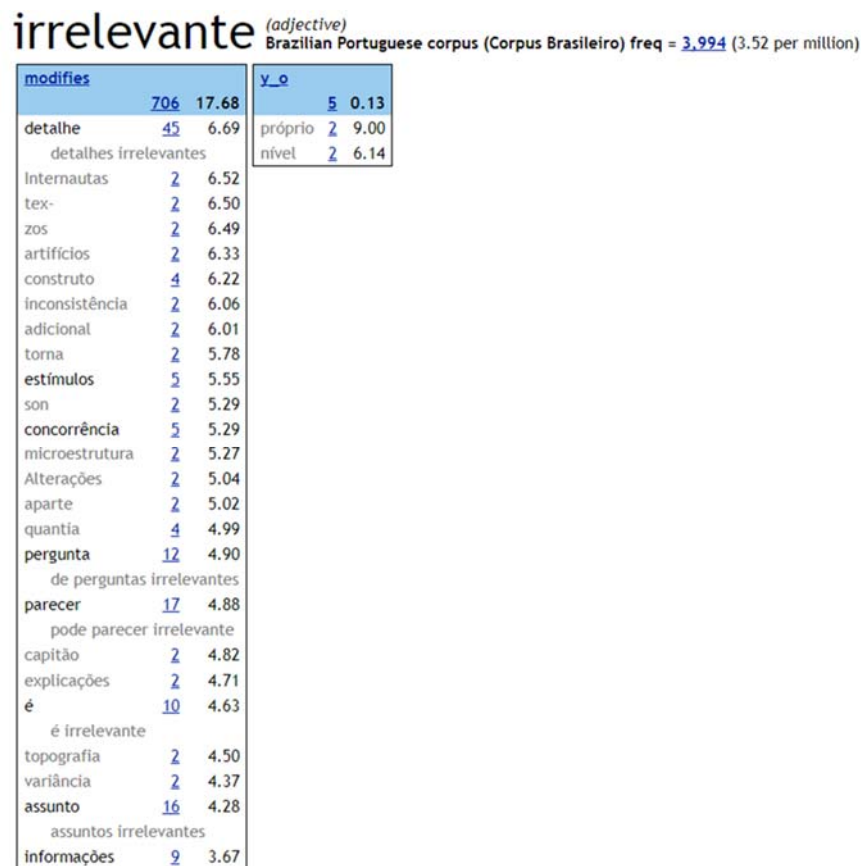


Figure 7: Word Sketch for the search word *irrelevante*

As Figure 7 shows, the result in Word Sketch for the search word *irrelevante* does not present any adverbial collocation patterns in Portuguese. One of the reasons for this may be that as the referred adjective is less frequent in Brazilian Portuguese (3,994 occurrences in the corpus, 3.52 per million) in relation to English (162,325 occurrences in the corpus, 7.14 per million), the number of collocational patterns is lower. Also, the search word “irrelevante” may not be frequently used in the adverb + adjective structure in Brazilian Portuguese and that may be why, out of 1,133,416,757 words which compose the Brazilian Portuguese Corpus (*Corpus Brasileiro*), no adverbial collocations were found. Nevertheless, these findings will have to be investigated more systematically in future studies. Another reason may be related to fact that the Brazilian Portuguese Corpus is not large enough to give the same results in comparison to the enTenTen13 corpus and thus, in order to have more reliable results or to provide more collocational options, it should be expanded. According to our experience in compiling the *Collocations Dictionary*, many search words in Portuguese did not have a satisfactory result in the Word Sketch with regards to collocational patterns, not only for adverbial collocations, but also for noun, adjectival and verbal collocations. Members of the team also reported that they have had difficulties in finding collocational patterns or have not found any successful results in the Brazilian Portuguese Corpus and, on account of that, we believe that expanding the corpus would be greatly welcome.

Another collocation considered hard to translate into an equivalent in Portuguese was *unflagging enthusiasm*. Although it is not an adverbial collocation, but a noun collocation, with the structure Adjective + Noun, it may be given as an example of translation problems. The translation options given, regarding the meaning of *unflagging* (not changing or becoming weaker) were “grande entusiasmo”, “crescente entusiasmo”, and “constante entusiasmo”, the latter being the best translation option. These translations were carefully proposed by the researcher’s team, taking into account that the Word Sketch results for the noun “entusiasmo” were also not satisfactory, as Figure 8 illustrates, even though some adjectival collocations for this search word were presented by the tool, differently from the previous search with the node *irrelevant*.

In order to achieve the translation options given, we also had to carry out a careful research using *Google*, checking whether they were frequently used in Portuguese. Besides that, three options were given and we had to decide on the best one to be included in the dictionary. Although some collocations and collocation types in English are hard to translate or may not have a ‘perfect’ equivalent in Portuguese, we have to take into account that part of the dictionary’s target audience is learner and professional translators. Therefore, we have made a great effort to offer them a translation equivalent, so that they may have the chance to enrich their translated texts, as well as easing their work.

**entusiasmo** *(noun)*  
Brazilian Portuguese corpus (Corpus Brasileiro) freq = 10,457 (9.22 per million)

object_of	2,266	21.67	subject_of	455	4.35	n_modifier	1,913	18.29	modifies	600	5.74	y_e	115	1.10
despertar	96	7.94	patriótico	5	7.67	contagiante	38	9.09	calafrio	5	7.98	trabalho	4	4.83
arrefecer	19	7.92	ingênuo	4	7.16	um entusiasmo contagiante			estremecimento	5	7.87	projeto	6	3.63
esfriar	10	6.77	durar	6	5.54	despertado	11	7.09	assomo	4	7.65			
manifestar	53	6.76	relação	19	4.32	entusiasmo despertado			fogueira	7	7.29			
susitar	22	6.71	o entusiasmo em relação			juvenil	24	6.61	cheia	6	6.79			
fingir	8	6.63	necessário	6	4.09	entusiasmo juvenil			auge	12	6.63			
esconder	31	6.55	crescer	6	3.87	incontido	6	6.52	no auge do entusiasmo					
esconde o entusiasmo			gerar	5	3.04	exagerado	14	6.41	grito	7	5.44			
demonstrar +	156	6.46	trazer	5	2.85	entusiasmo exagerado			surto	6	5.43			
moderar	5	5.75	fazer	8	0.95	Mestiço	4	6.08	clima	28	5.24			
idêntico	4	5.57	ser	16	0.66	moços	4	6.05	. O clima de entusiasmo					
apostar	4	5.51	o entusiasmo era			transbordante	4	6.01	onda	22	5.15			
inspirar	9	5.32	é	30	0.34	martinho	4	5.97	uma onda de entusiasmo					
compartilhar	13	5.30	, o entusiasmo é			estagiários	4	5.96	falta	93	4.92			
reinar	4	5.26	também	4	0.25	Edmundo	6	5.94	a falta de entusiasmo					
espantar	4	5.22	ir	24	0.15	milîtância	4	5.90	despeito	5	4.76			
transmitir	16	5.17	o entusiasmo foi			público	15	5.89	misto	4	4.52			
geração	4	5.15	\	19	0.14	o entusiasmo do público			excesso	14	4.43			
renovar	8	4.90	o entusiasmo \."			generoso	7	5.89	excesso de entusiasmo					
denotar	7	4.82				multidões	4	5.88	conduta	8	4.14			
mostrar +	110	4.78				investidor	12	5.83	tom	5	3.56			
economizar	4	4.74				o entusiasmo dos investidores			ausência	9	3.53			
testemunhar	4	4.70				manifestado	6	5.82	motivo	11	3.51			
notar	8	4.69				devoto	4	5.81	sentimento	9	3.51			
enfraquecer	4	4.69				redobrado	4	5.81	momento	21	2.85			
recuperar	15	4.55				contagioso	4	5.73	um momento de entusiasmo					
recuperar o entusiasmo						ardente	4	5.58	palavra	9	2.69			
						platéia	4	5.45	dose	6	2.60			
						marcial	5	5.37	exemplo	8	2.08			
						torcedor	6	5.34	fruto	4	1.95			
						suscitado	4	5.34	queda	4	1.65			

Figure 8: Word Sketch for the search word *entusiasmo*

It is worth mentioning that, having predicted the absence of a collocation in the target language, considering that some collocations in one language may correspond to a single lexical item in another, we created a symbol in the dictionary (☞) which indicates that the dictionary will inform the user whenever there is not a correspondent collocation in the target language.

All in all, compiling a corpus-based bilingual collocations dictionary is a time-consuming and long-term activity, demanding complete dedication and full commitment from the lexicographer. On the other hand, it is an enormously rewarding lexicographical enterprise. Regarding the Sketch Engine, even though it did not help us find a more accurate translation option in Portuguese, it has proved to be an invaluable tool for this work.

## 5. Conclusion

Although the compilation of a collocations dictionary may provide a lexicographer with a substantial amount of work, the result of it is an extremely useful tool for learners of English and Portuguese as foreign languages, as well as learner and professional translators. The pedagogical result of the lexicographical work may be of great help and value to the referred audience, and finally it can be regarded as a highly rewarding and extremely pleasant investigation and enterprise.

This paper reported on the motivations, as well as obstacles and achievements, in building the *Online Bilingual Collocations Dictionary*. Results have shown that even

though some difficulties and obstacles have arisen in the lexicographical process of compiling the collocations dictionary, the fact of having identified them may help us find more effective solutions so that we may produce a reliable source for the target audience.

Having shed light on the achievements and potential benefits of an online corpus-based bilingual collocations dictionary, and considering it to be the first collocations dictionary in the English–Portuguese and Portuguese–English directions, we hope the publication of the online version of the dictionary will be widely accessed and potentially useful.

## 6. Acknowledgements

I gratefully acknowledge the partial, but loyal, financial support provided by FAPERP (Fundação de Apoio à Pesquisa e Extensão de São José do Rio Preto).

## 7. References

- Allan, D. (2004). *Oxford Placement Tests 2*. Oxford: OUP.
- Alonso Ramos, M. (ed.) (2006). *Diccionarios y Fraseología*. A Coruña: Universidad A Coruña.
- Atkins, B. T. S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Benson, M., Benson, E. & Ilson, R. (1997). *The BBI Combinatory Dictionary of English*. Amsterdam/Philadelphia: John Benjamins.
- Browne, C., Culligan, B. & Phillips, J. (2016). The New General Service List. Available at: <http://www.newgeneralservicelist.org>. [4 Jan 2017].
- Kilgarriff, A., Rychly, P., Smrz, P. & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (eds.). *Proceedings of the 11th Euralex International Congress*. Lorient: Université de Bretagne-Sud, pp. 105-116.
- Davies, M. (2008-2012). *The Corpus of Contemporary American English: 425 million words, 1990-present*. [online] Available at: <http://corpus.byu.edu/coca>. [Nov 20, 2016].
- Fontenelle, T. (ed.) (2008). *Practical lexicography: A reader*. New York: Oxford University Press.
- Hausmann, F. J. (1985). ‘Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In H. Bergenholtz & J. Mugdan (eds.). *Lexikographie und grammatik*. Tübingen: Niemeyer.
- Hill, J. & Lewis, M. (1999). *LTP Dictionary of Selected Collocations*. Hove, London: Language Teaching Publications.
- Kilgarriff, A. (2015). Using corpora as data sources for dictionaries. In H. Jackson (ed) *The Bloomsbury companion to Lexicography*. London: Bloomsbury, pp 77-96.
- Kilgarriff, A. & Tugwell, D. (2002). Sketching words. In *Lexicography and Natural Language Processing: a Festschrift in Honour of B. T. S. Atkins*. Marie-Hélène

- Corréard (ed.). EURALEX, pp. 125-137.
- Mcintosh, C., Francis, B. & Poole, R. (eds.) (2009). *Oxford Collocations Dictionary for Students of English*. 2nd ed. Oxford: Oxford University Press.
- Moon, R. (2008). Dictionaries and collocations. In F. Meunier and S. Granger (eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, pp. 247-252.
- New General Service List 1.01 (2016). Available at: <http://www.newgeneralservicelist.org>. [25 jan. 2017].
- Orenha-Ottaiano, A. (2004). *A compilação de um glossário bilíngue de colocações, na área de jornalismo de Negócios, baseado em corpus comparável*. MA diss., University of São Paulo (USP), Brazil.
- Orenha-Ottaiano, A. (2013). The proposal of an electronic bilingual collocational dictionary based on corpora. In *X International School on Lexicography Proceedings*, 2013. Florence, Italy. Life Beyond Dictionaries. Ivanovo: University of Ivanovo, pp. 405-408.
- Orenha-Ottaiano, A. (2016). The compilation of an online corpus-based bilingual Collocations Dictionary. In Gloria Corpas Pastor. (ed.). *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Editions Tradulex, 2016, pp. 486-493.
- Pawley, A. & Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J Richards & R. Schmidt (eds.) *Language and Communication*. London: Longman.
- Rundell, M. (2010) *Macmillan Collocations Dictionary for Learners of English*. Oxford: Macmillan Publishers Ltd.
- Rundell, M. (2013). Redefining the dictionary: From print to digital. In *Kernerman Dictionary News 21*. Available at: <http://kictionaries.com/kdn/kdn21.pdf>. [4 Jan 2017].
- Scott, M. (2008). *WordSmith Tools*, version 5.0. Liverpool: Lexical Analysis Software Ltd.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stevick, E. W. (1989). *Success with foreign languages*. Hemel Hempstead: Prentice Hall.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

