
The Compilation of a Printed and Online Corpus-based Bilingual Collocations Dictionary

Adriane Orenha-Ottaiano

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)

e-mail: adriane@ibilce.unesp.br

Abstract

Collocations are conventionalized, recurrent and arbitrary lexical combinations. Due to the fact that they are highly specific for a particular language and may be contextually restricted, collocations pose a problem to foreign language learners and translators with regard to production or encoding. Taking that into account, the compilation of monolingual and bilingual collocations dictionaries for both learners of English and Brazilian Portuguese as a foreign language and professional translators is highly crucial and significant. The aim of this paper is to give an account of the compilation of a printed and online *Corpus-based Bilingual Collocations Dictionary*, in the English-Portuguese and Portuguese-English directions. The dictionary will focus on all types of collocations (verbal, nominal, adjectival and adverbial collocations), in order to help the referred audience use them more accurately and productively. The idea of having the proposed dictionary in online format will allow to incorporate more qualitatively and quantitatively collocational information. Furthermore, more examples can be included, different from conventional printed collocations dictionaries. Being the first bilingual collocations dictionary in the aforementioned directions, it is hoped to achieve the challenge of meeting learners' collocational needs as the collocations will be selected according to Brazilian learners' difficulties.

Keywords: corpus-based collocations dictionary; collocations; bilingual collocations dictionary; collocational patterns

1 Introduction

There is a general consensus among researchers about the great pervasiveness of collocation in the language as already pointed out by Sinclair (1987) and agreed by Wray (2002:52) when the latter restated “the much greater pervasiveness of collocation within the fabric of language”.

When discussing vocabulary learning in another language, Nation (2001) mentions three important aspects: (1) language knowledge is collocational knowledge; (2) all fluent and appropriate language use requires collocational knowledge; and (3) many words are used in a limited set of collocations and knowing these is part of what is involved in knowing the words. All these aspects give evidence that collocational knowledge plays a key role when learning another language and show the need of teaching collocations both implicitly and explicitly. By mapping out strategies to expand their collocational knowledge, especially when learners are feeling linguistically limited, we may also contribute to most learners' goal as well as wish of sounding like others in a specific linguistic community, thus serving to signal membership and assert identity within this community (Wray 2002, Orenha-Ottaiano 2015).

Another issue which corroborates the relevance of developing collocational competence in learners is concerned with achieving fluency in a foreign or second language as, in the process of writing or speaking, native speakers do not simply bring separate words together in linear succession, ‘they also use ‘ready-made’ units, prefabricated blocks (i.e. phraseological units) that already exist in language as a global whole and function in speech as one word’ (Ter-Minassova 1992:535). Hence, what appears to be spontaneous is actually a stereotyped fixed and repetitive speech, and if the speaker does not have a vast repertoire of these stereotyped fixed units at their disposal, their speech may not sound natural. Moreover, according to Nesselhauf (2005), “psycholinguistic evidence points out that the human brain is much better equipped for memorizing than for

processing, and that the availability of large numbers of prefabricated units reduces the processing effort and thus makes fluent language possible”. And, to develop this fluency, it is worth counting on a wide repertoire of collocations.

For the fact of being conventionalized, recurrent and arbitrary lexical combinations and also for being highly specific for a particular language and contextually restricted, collocations pose a problem to foreign language learners and translators with regard to production or encoding. Easily understood collocations may not cause comprehension problems when a learner is reading a text or listening to a fluent speaker of a language. This happens because they encounter many literal collocations, and if the learners know the meaning of the individual components of the collocation they will not pay attention to them. For this reason, collocations tend to be neglected by learners of a second or foreign language and, as Wray (2002:) points out, “learners do not pay attention to collocational relationships”.

In a study carried out by Granger (1998), whose hypothesis was that learners would make less use of prefabricated language, such as collocations, than native speakers, it was concluded that learners use words “more as building bricks than as parts of prefabricated sections” and also that they use fewer prefabs than their native speaker counterparts. Besides that, the researcher believes there is evidence that the collocations used by the learners are for the most part congruent and may thus result from transfer from LI. By considering this context, it can be implied that more pedagogical material on collocations, as noted by Nesselhauf (2005) and Orenha-Ottaiano (2014; 2015), as well as monolingual and bilingual collocations dictionaries (Orenha-Ottaiano 2016), claimed by Atkins and Rundell (2008), Kilgariff *et al.* (2010), Granger and Paquot (2012), are not only desirable but also indispensable.

The contribution of corpora to Lexicography as well as Phraseography is evident and has been argued by many researchers (Halliday *et al.* 2004; Teubert 2004, 2007; Burger *et al.* 2007; Sinclair 2007; Atkins & Rundell 2008; Moon 2008; Granger 2012; Jackson 2013; Kilgariff 2013; Orenha-Ottaiano 2013, 2016, among others). Atkins and Rundell (2008), when talking about the central role of corpora to Lexicography, argue that in order to compile a reliable dictionary, one has to count on objective evidence of language in use and for that a lexicographer needs very large volumes of data so that he or she can observe them to take place frequently and recurrently in different types of texts. According to the authors, a researcher “can only confidently distinguish what is conventional from what is idiosyncratic if we have plenty of data at our disposal”.

Moon (2008) stated that a corpus is “the great facilitator for the description of phraseology, and its use in dictionary-making has heavily influenced the ways in which lexicographical attitudes towards phraseology have developed over the last twenty-five years”. It is also true that the quality of monolingual and bilingual dictionaries has improved, due to the methodology provided by Corpus Linguistics. The use of corpora has enabled us to identify and extract phraseological units more easily and quickly, especially collocations and contributes to improving collocational information in dictionaries.

Rundell (2012) and Kilgariff (2013) mention one aspect that cannot be left out when pointing out the benefits of electronic corpora to Lexicography: the corpus query system, such as the *Sketch Engine* (Kilgariff *et al.* 2004). Kilgariff (2013) adds that taking into account that we are dealing with more and more quantities of data, larger and larger corpora, faster and faster computers, with progress in corpus and linguistic software, the field of Lexicography is changing all the time. Because of that, lexicographers should keep up with the latest relevant and innovative technologies and should always have in mind that “by the time any practice is standard and widely accepted, it will be well behind the latest developments” (Kilgariff 2013:78).

The growing relevance of collocations to Phraseography has also led to improvements in general language monolingual learners’ dictionaries thanks to the use of corpus and corpus-querying software. Dictionaries like *MacMillan English Dictionary*, *Collins Cobuild Advanced Learners’ Dictionary*, *Oxford Advanced Learners’ Dictionary* and *Longman Advanced Learners’ Dictionary* were compiled, having included, even though more implicitly, a lot of collocational information.

Nevertheless, even the best dictionaries like the ones mentioned cannot provide the range of information that a good collocations dictionary can.

In 1986, Benson, Benson and Ilson having already and accurately foreseen the relevance of collocations for encoding, published the first monolingual English collocations dictionary, *The BBI Combinatory Dictionary of English*. More than a decade later, *LTP Dictionary of Selected Collocations* (Hill; Lewis 1999) was launched. However, neither dictionaries were corpus-based and, due to the advances of Corpus Linguistics, access to gigantic corpora as well as use of corpus-querying software have become crucial to Lexicography and Phraseography.

The need for more reliable and accurate dictionaries has arisen and, as a natural consequence, in 2002 we witnessed the launching of *Oxford Collocations Dictionary for Students of English* (McIntosh; Francis; Poole [2002], 2009), the first corpus-based monolingual English collocations dictionary, thus reducing collocational gaps. After that, Rundell (2010) published *Macmillan Collocations Dictionary for Learners of English*, another corpus-based monolingual English collocations dictionary, this time counting on the help of the *Sketch Engine* (Kilgarriff et al. 2004) and, three years later, *Longman Collocations Dictionary and Thesaurus* (Mayor 2013) was published as printed and online version.¹

Along with the birth of monolingual collocations dictionaries, there can also be seen a driving need for bilingual collocations dictionaries (Benson & Benson 1993; Benson et al 1995; Konecny & Autelli 2014 etc.), with the aim of encouraging learners and translators more actively use collocations and incorporate them into their mental lexicon, also drawing attention to the relevance of compiling bilingual dictionaries. Adamska-Sałaciak (2013) shows the significance of bilingual dictionaries in general, which can also be applied to bilingual collocations dictionary:

While it is hard at present to imagine a world without dictionaries, especially bilingual ones, no one really knows how the world of reference science is going to be transformed by the technological advances we are witnessing. What does seem certain is that people will always need information about interlingual lexical correspondences and that, no matter in what form and through what medium that information is conveyed to them, it will capitalize upon the work which bilingual lexicographers have done over the past centuries. (Adamska-Sałaciak 2013:228).

On the scenario presented above, considering that collocations are increasingly acknowledged as one of the main aspects of language, the rationale for the compilation of a *Corpus-based Bilingual English Collocations Dictionary*, with all types of collocations (verbal, noun, adjectival and adverbial collocations) is, first of all, the scarcity of corpus-based printed or online collocations dictionaries in the pair of languages investigated in this research (English-Portuguese and Portuguese-English directions) and, secondly, the complexity of learning collocations and developing collocational competence and fluency for learners of English as a second or foreign language, learners of Brazilian Portuguese as a foreign language as well as translators.

Hence, the proposed dictionary has the aim of bridging a collocational gap in what concerns providing students and translators with an equivalent collocation in a specific target language, in order to respond to their needs. Moreover, being a Web-based dictionary, we share Rundell's view in the sense this type of work provides "a fuller more systematic account of how language works". Rundell (2012: 16) adds that a printed book would not store all the information we want to convey: "it is only through digital media that we can make the fruits of our language analysis available to users at a level of details that was never previously possible".

2 Corpora and Methodology Used for the English-Portuguese Direction

The first phase of this investigation with a view to compiling a *Corpus-based Bilingual Collocations Dictionary* was extraction and analysis of collocational patterns from a *Translation Learner Corpus*, a parallel corpus, with almost 100,000 tokens. It is made up of university students'

¹ It is also worth mentioning a very complete work on Spanish collocations compiled by Bosque (2004, 2006).

translations in the Portuguese-English direction, whose level of English varies from B2 to C1 - students' knowledge of language was identified according to the results of the *Oxford Placement Test* (Allan 2004). The original texts that comprise the corpus are newspaper articles taken from well-known Brazilian newspapers and magazines and the typology of the texts are related to current world news and the topics selected were *One year after Tsunami in Japan; Financial crises in Greece and in Europe; Unemployment; Elections in the US; Bullying; Marijuana Legalization* among others. The referred corpus was compiled at *Universidade Estadual Paulista (UNESP)*, in Brazil. The rationale for compiling and analyzing a *Translation Learner Corpus* was that we intended to investigate translation learners' difficulties regarding the use of collocations and further incorporate them into the dictionary.

The collocations were then extracted with the help of *WordSmith Tools* (Scott 2008), which enabled us to raise the most frequent collocational patterns and the most/least used types of collocations used by the translation learners in comparison to the original texts, the influence of the mother tongue on their choices, the difficulties they had in encoding collocational patterns, among other aspects.

For example, among the keywords from the *Translation Learner Corpus*, the node *primaries* was analyzed, more specifically the translational options for the collocation *realizar primárias* (hold primaries). Below you can find students' options for *realizar primárias* and our search in COCA and *Sketch Engine*, in order to check whether these possibilities were frequently used by native speakers of English:

PRIMARIES – TLC	PRIMARIES (COCA)	PRIMARIES (<i>Sketch Engine</i>)
achieve primaries	0	2
realize primaries	0	0
make primaries	0	11
do primaries	0	2
perform primaries	0	0
hold primaries	22	191
carry out primaries	0	0
set up primaries	0	1

Table 1: List of learners' possible collocates for the node *Primaries*.

As it can be seen from the table above, among all the translational options given by students for *realizar primárias*, the most frequent and recurrent collocation is *hold primaries*. The other options given were not frequently used and conventionalized and thus may not be considered good translation options. In the case of *realize primaries*, students probably used the verb *realize* because this verb can be, in most contexts, a false cognate in Portuguese. The same happened to the combination *to make primaries*. They were used in a combination such as *to make primaries + an adjective: to make primaries mandatory* etc., different from the meaning of the investigated collocation.

This result may suggest that even students at higher proficiency levels are not quite aware of the conventionalized and collocational aspects of the language, regarding that they had access to the Internet and dictionaries (printed and online) to do the translations. The analysis of the collocations extracted from the *Translation Learner Corpus* has helped us decide which collocations could be included in the *Collocations Dictionary*, besides the other collocational patterns, which have already been selected from COCA and *Sketch Engine*.

In a second stage, based on the keywords and collocations from the *Translation Learner Corpus*, more collocational patterns were extracted firstly using *COCA, The Corpus of Contemporary American English* (Davies 2008-2012) and, later on, just counting on the *Sketch Engine* (Kilgarriff *et al.* 2004). From then on, only the *Sketch Engine* has been used as the software tool which gives us the word sketches,² that is, it shows us a wide range of collocational patterns and so it made our work much easier and even more reliable, mainly because it was not necessary to check the frequency of collocations anymore, as we had to do after we had extracted them from COCA.

It is worth emphasizing that the collocations to be included in the dictionary are not only the ones generated from the keywords of the *Translation Learner Corpus*, otherwise it would compromise the usefulness of the dictionary, as most frequently used lexicon and collocations in the English language would not be fully represented in the work. In order to include a wider range of collocational patterns as well as to ensure dictionary users will have access to the most frequent and recurrent collocations, we also used *COCA's* most frequent lemma list (100,000 words), based on the 500 million words that comprise the referred corpus, with the purpose of extracting more patterns out of the content words from this list. Currently, we are also analyzing the most frequent content words from the wordlist of enTenTen13, the largest corpus available via *Sketch Engine*, with nearly 20 billion tokens, from which more collocational patterns are extracted by using *Word Sketch*. It is important to note that we have to analyze the most frequent words as we do not necessarily have to include them in the entries of the collocations dictionary. We decided to include the collocations which pose collocational problems to our target audience as well as the ones with a good range of strong collocates, as Adamska-Sałaciak (2013:220) argues, 'the wordlists should be based on corpus frequency, but not only on this, as usefulness to a target audience must also be considered'.

The target audience aimed at the printed and online *Corpus-based Bilingual Collocations Dictionary's* English-Portuguese direction version is mainly Brazilian learner and professional translators and learners of Brazilian Portuguese as a foreign language whose first language is English and also Brazilian learners of English as a foreign language.

3 Corpora and Methodology Used for the Portuguese-English Direction

Based on the difficulties Brazilian learner translators had in the use of collocations according to the analysis of the *Translation Learner Corpus* above mentioned, a lot of collocations were also selected to be included in the Portuguese-English direction version of the *Corpus-based Bilingual Collocations Dictionary*.

Besides that, the Portuguese TenTen corpus via *Sketch Engine*, with 4,626,584,246 tokens has also been used. We are analyzing the most frequent content words from the wordlist of Portuguese TenTen, from which more collocational patterns are extracted by using *Word Sketch*.

The same rule used for the English-Portuguese version is applied to the Portuguese-English one in what concerns the most frequent words of the Portuguese TenTen corpus: we analyze them, check the collocational patterns each word offers and do not necessarily all frequent words raised from the referred corpus in the entries of the collocations dictionary. We again just insert the collocations which pose collocational problems to our target audience as well as the ones with a good range of strong collocates.

As the proposed dictionary is designed for encoding purposes, the target audience aimed at the printed and online *Corpus-based Bilingual Collocations Dictionary's* Portuguese-English direction version is mainly Brazilian learners of English as a foreign language, Brazilian learner and professional translators and also learners of Brazilian Portuguese as a foreign language whose first language is English.

² A *Word Sketch* is a "summary of a word's grammatical and collocational behaviour produced automatically, from a large corpus, for a lexicographer" (Kilgarriff; Tugwell 2002:125).

4 Macro and Microstructure of the Bilingual Collocations Dictionary

The macro and microstructure of the dictionary, both printed and online version, is based on the methodology proposed by Orenha-Ottaiano (2004, 2009) for the compilation of corpus-based specialized collocations glossaries, anchored in Hausmann’s theory (1985) and Heid *et al*’s point of view (1991) in that the basis is usually what we already know and the collocate is the element we are looking for.

On that account, and according to the principles of Corpus Linguistics, a basis, displayed in alphabetical order, will become an entry as long as: 1) it is frequent (for the English-Portuguese direction – according to *COCA*’s most frequent lemma list and the most frequent content words from the wordlist of the enTenTen13; and for the Portuguese-English one – the most frequent content words from the Portuguese TenTen corpus); 2) it poses collocational problems to our target audience when combined with their collocates; and 3) it provides with a good range of strong collocates. The collocates are organized right below the basis, also in alphabetical order, as shown in the table below:

PRIMARY (n.)
Verbal – Verb + PRIMARY
Conduct a ~ I took The Chronicle’s advice and called an election official in New York to see how they conduct September primaries .
Enter a ~ If you had decided to run, what do you think your chances would have been entering the Republican primaries and going through the primary process?
Hold a ~ Because those states held primaries , both won by Hillary Clinton, in defiance of party rules, their delegates will not be seated at the Democratic convention.

Table 2: Collocations extracted from *COCA*, out of the keyword *primaries*.

The *Corpus-based Bilingual Collocations Dictionary* will deal with all types of collocations: verbal, adjectival, nominal and adverbial, according to the grammatical relations of the entry. For example, if the entry is a noun as in *Interview*, there may be three possible types of collocations: verbal (verb + noun or noun + verb), adjectival (adjective + noun) and nominal (noun + noun). If the entry *Interview* is a verb, there may be a possible combination: adverbial collocation (adverb + verb or verb + adverb), as shown in the table below:

INTERVIEW (n.)	ENTREVISTA (s.)
VERBAL – Verb + INTERVIEW	Verbo + ENTREVISTA
Conduct an ~ Last March, during the South by Southwest Music Festival in their hometown of Austin, I conducted a long interview with the " three Charlies " who lead C3: Jones, the mover and shaker behind the revitalized Lollapalooza; [...]	Conduzir uma ~ Fred Durst, frontman do Limp Bizkit , ficou nervoso com a maneira com a qual o jornalista espanhol Jordi Meya <i>conduziu uma entrevista</i> com o cantor, por telefone.
ADJECTIVAL – Adjective + INTERVIEW	ADJETIVA – Adjetivo + ENTREVISTA
Live ~ The show’s host called in sick, and a desperate station manager recruited Diane to conduct a live interview with a spokeswoman from the Dairy Council about the wonders of milk and cheese.	~ Ao vivo Uma repórter desmaiou durante uma entrevista ao vivo em um programa americano.
NOMINAL – Noun + INTERVIEW	NOMINAL – ENTREVISTA + Adjunto Adnominal
Job ~ A friend of mine was actually laid off and she went for a job interview just the other day in a place she didn’t know that she’d enjoy, and suddenly all	~ de Emprego O espírito que anima uma <i>entrevista de emprego</i> deve ser sempre positivo.

of her creative juices started flowing. ⓘ	
INTERVIEW (v.)	ENTREVISTAR (v.)
ADVERBIAL – INTERVIEW + adverb	ADVERBIAL – ENTREVISTAR + advérbio
Formally ~	~ formalmente
Golden State has talked to former Cleveland coach Mike Brown and Mavericks assistant Dwane Casey and plans to formally interview them as well. ⓘ	O tenista escocês afirmou também que ele não precisa entrevistar formalmente todos aqueles de sua lista. ⓘ

Table 3: Example of a collocation entry (*Interview*).

Thus, based on Hausmann (1985) and with examples taken from Orenha-Ottaiano (2015), the Collocations Dictionary will have entries structured according to the following types and grammatical relations:

Verbal

– with four basic structures:

- Verb _{collocate} + Noun _{basis}: *acquire shares*
- Noun _{basis} + Verb _{collocate}: *investments dropped*
- Verb _{collocate} + Preposition + Noun _{basis}: *dispose of shares*
- Verb _{collocate} + Adverbial Particle + Noun _{basis}: *set up a business*
- Verb _{collocate} + Adjective _{basis}: *grow strong*

Nominal

– with two basic structures:

- Noun _{basis} + Noun _{collocate}: *share subscription*
- Noun _{collocate} + Preposition + Noun _{basis}: *holder of shares*

Adjectival

– with one basic structure:

- Adjective _{collocate} + Noun _{basis}: *bearer shares*

Adverbial

– with three basic structures:

- Adverb _{collocate} + Adjective _{basis}: *fully eligible*
- Verb _{basis} + Adverb _{collocate}: *drop dramatically*
- Adverb _{collocate} + Verb _{basis}: *fully paid; duly appointed*

Figure 1: The entry structure of the Collocations Dictionary.

The *Corpus-based Bilingual Collocations Dictionary* has up to now more than 440 entries, and more than 3,500 collocations, both in English and in Portuguese, all of them with their corresponding contexts (examples) as shown in table 2. For the online version, the examples will be followed by the icon ⓘ, indicating the sources of the examples for the collocations, so that the dictionary gets visually cleaner. If the user clicks on it, he or she will have access to the source of the example:



<http://corpus.byu.edu/coca/>

Figure 2: Icon indicating the source of the example.

The idea of having the proposed dictionary in online format will allow us to incorporate collocational information more qualitatively and quantitatively. Besides that, more examples can be included. Having access to all types of collocations, the user will also have the chance to choose from different types of collocations in case he or she wants to restrict the query, by clicking *taxonomy*, as shown in the figure below, taken from the prototype of the online version of the dictionary:



Figure 3: Screenshot of the prototype of the online version of the dictionary.

On the right of the screen, the user can see all the entries of the dictionary in English. If the user clicks on the flag (Brazilian and English flag) on the top right, he or she can switch languages and see the collocations in the Portuguese direction. The screen will show one language on the left and the equivalent collocations in the other language, as illustrated below:

The screenshot displays the 'Corpus-based Bilingual Collocations Dictionary' interface. At the top, the title is followed by the name 'Adriane' and flags for Brazil and the USA. On the left, there is a sidebar with an 'Add Entry' button, an 'Advanced Search' field, and a list of words starting with 'A' and 'B'. The main content area is a grid of collocation entries:

Primária (n.)	Primary (n.)
Verbal – Verbo + Primária	Verbal – Verb + Primary
<p>Participar de uma ~</p> <p>Romney, que participou das primárias de 2008, liderou as enquetes apoiado em seu programa econômico e em sua longa experiência como empresário e político. ⓘ</p>	<p>Participate in a ~</p> <p>Nowhere is the support for big changes stronger than among voters who participate in Democratic presidential primaries. ⓘ</p>
Realizar uma ~	Hold a ~
Ganhar uma ~	Win a ~
Concorrer a uma ~	Enter a ~
Conduzir uma ~	Conduct a ~

Figure 4: Screenshot showing the collocation from the Portuguese-English direction.

Grounded on the importance of this work to the above mentioned target audience, and taking into account the potential it has to generate a two-version product, printed and online, we strongly believe that, considering the way it is methodologically organized, the *Corpus-based Bilingual Collocations Dictionary* will boost the access of more people, from different parts of Brazil and abroad.

5 Conclusion

Compiling a dictionary of collocations is an extremely challenging task due to the many details a lexicographer has to take into account. We hope the proposed *Corpus-based Bilingual Collocations Dictionary* may become a user-friendly and valuable new resource for its target audience.

Having shed some light on the relevance of collocation learning, development of collocational competence as well as the potential benefits of the *Corpus-based Bilingual Collocations Dictionary*, considering the referred work to be the first collocations dictionary in the English-Portuguese and Portuguese-English directions, we expect its publication, printed and online version, to be widely accessed and extremely useful.

Believing that the traditional format of a print dictionary would limit its appeal for nowadays learners, the online version of the dictionary is hoped to meet students' specific needs.

6 References

- Adamska-Sałaciak, A. (2013). Issues in compiling bilingual dictionaries. In: H. Jackson. *The Bloomsbury Companion to Lexicography*. London: Bloomsbury Publishing Plc.
- Allan, D. (2004). *Oxford Placement Tests 2*. Oxford: OUP.
- Altenberg, B., Granger, S. (2002). *Lexis in Contrast*. Amsterdam/Philadelphia: Benjamins.
- Atkins, S. B. T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Benson, M., Benson, E. (1993). *Russian-English Dictionary of verbal collocations*. Amsterdam: John Benjamins.

- Benson, M., Benson, E. & R. Ilson (1997). *Longman Dictionary of English Collocations* (Bilingual English-Chinese). London: Longman.
- Benson, M., Benson, E. & R. Ilson (1997 [1986]). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins.
- Bosque, I. (ed.) (2004). *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- Bosque, I. (ed.) (2006). *Diccionario combinatorio práctico del español contemporáneo*. Madrid: SM.
- Burger, H., Dobrovolskij, D., Kuhn, P. & Norrick, N. R. (eds.) (2007). *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*. 2 Halbbände (=HSK 28.1/2). Berlin/New York: de Gruyter.
- Davies, M. (2008-2012). *The Corpus of Contemporary American English: 425 million words, 1990-present*. [online] Accessed at: <http://corpus.byu.edu/coca/> [20/11/2015].
- Granger, S. & Paquot, M. (2012). *Electronic Lexicography*. Oxford: Oxford University Press.
- Halliday, M.A.K., Teubert, W., Yallop, C. & Čermáková, A. (eds.) (2004). *Lexicology and Corpus Linguistics: an introduction*. London: Continuum.
- Hausmann, F. J. (1985). 'Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels'. In: H. Bergenholtz; J. Mugdan, eds., 1985. *Lexikographie und grammatik*. Tübingen: Niemeyer.
- Heid, U., Martin, W. & Posch, I. (1991). An Overview of approaches towards the description of collocations. *Feasibility of standards for collocational description of lexical items. Eurotra 7-Report*, Stuttgart/Amsterdam.
- Hill, J., Lewis, M. (1999). *LTP Dictionary of Selected Collocations*. Hove, London: Language Teaching Publications.
- Jackson, H. (ed.) (2013). *The Bloomsbury Companion to Lexicography*. London: Bloomsbury Publishing Plc.
- Kilgarriff, A.; Rychly, P.; Smrz, P.; Tugwell, D. The Sketch Engine. In G. Williams; S. Vessier (eds.). *Proceedings of the 11th Euralex International Congress*. Lorient: Université de Bretagne-Sud, p. 105-116, 2004.
- Kilgarriff, A.; Tugwell, D. (2002). Sketching Words. In: M. -H. Corréard (ed.). *Lexicography and Natural Language Processing*. A Festschrift in Honour of B.T.S. Atkins. Euralex.
- Konecny, C, Autelli, E. (2014). *Kollokationen Italienisch-Deutsch*. Vydavateľ: Helmut Buske Verlag.
- Mayor, M. (ed.) (2013). *Longman Collocations Dictionary and Thesaurus*. Harlow: Pearson.
- Mcintosh, C.; Francis, B. & Poole, R. (eds.) (2009 [2002]). *Oxford Collocations Dictionary for Students of English*. 2nd ed. Oxford: Oxford University Press. 963 p.
- Meunier, F., Granger, S. eds. (). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Moon, R. (2008). Dictionaries and collocations. In: S. Granger and F. Meunier (eds.) (2008). *Phraseology: an interdisciplinary perspective*. Amsterdam: John Benjamins, pp. 247-252.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Orenha-Ottaiano, A. (2004). *A compilação de um glossário bilíngue de colocações, na área de jornalismo de Negócios, baseado em corpus comparável*. Unpublished Master's thesis, Universidade de São Paulo (USP), Brazil.
- Orenha-Ottaiano, A. (2009). *Unidades fraseológicas especializadas: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado*. Unpublished Ph.D. Thesis, Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Brazil.
- Orenha-Ottaiano, A. (2013). The proposal of an electronic bilingual dictionary based on corpora. In *X International School on Lexicography*, Florence, Italy. Life Beyond Dictionaries. Ivanovo: University of Ivanovo. 1: p. 405-408.
- Orenha-Ottaiano, A. (2015). Collocations workbook: um material de apoio pedagógico on-line

- baseado em corpus para o ensino de colocações em inglês. *Revista de Estudos da Linguagem*, v. 23, p. 833-881.
- Orenha-Ottaiano, A. (2016). The compilation of an online corpus-based bilingual Collocations Dictionary. In: G. Corpas Pastor (ed.). *Computerised and corpus-based approaches to Phraseology: Monolingual and Multilingual Perspectives*. 1ed. Geneva: Editions Tradulex, v. 1, p. 486-493.
- Rundell, M. (2010) *Macmillan Collocations Dictionary for Learners of English*. Oxford: Macmillan Publishers Ltd.
- Rundell (2012). The road to automated lexicography: an editor's viewpoint. In S. Granger & M. Paquot, (2012). *Electronic Lexicography*. Oxford: Oxford University Press.
- Sinclair, J. McH. (1987). Collocation: a progress report. In R. Steele & T. Treadgold (eds.) *Essays in honour of Michael Halliday*. Amsterdam: John Benjamins, p. 319–331.
- Sinclair, J. McH. (2007). Data-derived multilingual lexicons. In: W. Teubert (ed.). (2007). *Text corpora and multilingual lexicography*. Amsterdam: John Benjamins, p. 69-82.
- Teubert, W. (2004). Language and Corpus Linguistics. In M.A.K., Halliday, W. Teubert; C. Yallop & A. Čermáková (eds.). (2004). *Lexicology and Corpus Linguistics: an introduction*. London: Continuum.
- Teubert, W. (ed.) (2007). *Text corpora and multilingual lexicography*. Amsterdam: John Benjamins.
- Scott, M. (2008). *WordSmith Tools*, version 5.0. Liverpool: Lexical Analysis Software Ltd.

Acknowledgements

I gratefully acknowledge the financial support to my research provided by *FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo/ São Paulo Research Foundation)* that made my participation in Euralex 2016 Conference possible.