The creation of an online English collocations platform to help develop collocational competence

Adriane Orenha-Ottaiano

SÃO PAULO STATE UNIVERSITY (UNESP)

Abstract *:* The aim of this paper is to discuss the compilation of an *Online English Collocations Platform*, designed to help Brazilian Portuguese as well as other foreign language speakers to learn English. Moreover, we address the role and importance of collocations in foreign language learning, teaching and training. The Platform activities have been firstly developed based on some difficulties Brazilian Portuguese speakers have in learning collocations in English, according to research data results (Orenha-Ottaiano, 2012a, 2012b, 2013; Brito; Orenha-Ottaiano, 2013). The data was obtained from: 1) a Translation Learner Corpus (parallel) made up of texts translated from Portuguese into English by B.A. in Translation undergraduate students, and 2) a Learner Corpus composed of essays written in English by B.A. in English language undergraduate students. In addition to that, other collocations have been added to the Platform, which may also help the target audience develop their collocational competence in English. To build up the Platform activities, the methodology involves the extraction and analysis of collocations from the referred learner corpora (to focus on Brazilian students' difficulties and needs) as well as the extraction of frequent collocational patterns from *EnTenTen15*, from the *Sketch Engine* (Kilgarriff et al., 2004), The Platform is expected to help pre- and in-service teachers work with the referred phraseologisms in the classroom more effectively and learners use them more accurately and productively with a view to have one more resource to develop collocational competence and fluency in English.

Keywords: collocations; online platform; collocational patterns; collocational competence; corpus.

Introduction

➔ Nesselhauf (2005) states that many non-native speakers and learners of a foreign language naturally have the goal of reducing the processing effort of more complex phraseological units, if they aim to convey their message more clearly and develop fluency. And for that, it is necessary to have control of a large repertoire of prefabricated units, as Wray (2002) claims.

➔ Kjellmer (1991: 125) explains that "the mental lexicon of any native speaker contains single-word units as well as phrasal units or collocations". For the author, "mastery of both types is an essential part of the linguist equipment of the speaker or writer and enables him to move swiftly and with little effort through his exposition from one prefabricated structure to the next".

➔ According to our teaching experience in English as a foreign language, we have repeatedly encountered great frustration among learners regarding oral production. They frequently mention that they fail to produce the same language as native speakers, that they usually feel linguis-

tically limited, due to the difficulty of combining words suitably. For instance, they may know the word *order*, in the sense of a requesting for a product, however, when they wish to make up a phraseological unit like *place an order*, they may, as first attitude and by means of a linguistic transfer from their mother tongue, use the combination *make an order* instead.

→ Another case is the use of adverbial collocations by Brazilian Portuguese speakers who tend to use the intensifier *very* whenever they want to express intensity. They seem to forget that they cannot freely combine "very" with all adjectives, that some adjectives have a meaning which is extreme or absolute and cannot be easily made stronger or weaker - the so-called *ungradable adjectives*. As an example, we can cite the adjective *cold*. Brazilian Portuguese speakers would tend to say *very very cold* instead of other frequent and correctly used adverbial collocations like *freezing cold* or *bitterly cold*. The same occurs to the adjective *naked*, whose collocate may be *completely* or *entirely*, but also, and more frequently used, *stark* or *buck*.

→ This situation happens because student's first attitude is to transfer the combination from his or her L1 into the L2, which may lead, in some cases, to misunderstanding or often to some strangeness on the part of the native speaker, as noted by Ter-Minasova (1992: 534): "learners of a foreign language combine words – orally and in writing – following their mother tongue's collocational patterns, which results in all sorts of errors, ludicrous mistakes […]".

→ According to Sinclair (1991, 2004), if we choose combinations which are not expected for a specific environment, the receiver's attention will be directed to form, not to content. In con-

trast, that may not happen if we choose prefabricated chunks, as they may not disturb or break the receptor's concentration. Regarding this matter, it is worth mentioning the research carried out by Altenberg and Eeg-Olofsson (1990: 2) on the understanding and production of discourse, having as research questions: "How do competent speakers of a language (or genre) manage to talk more fluently than (it appears) their limited processing capacity should permit? How do simultaneous interpreters achieve their seemingly miraculous feats?" Altenberg and Eeg-Olofsson (1990) conclude that speakers involved in a spontaneous interaction constantly need easily retrieved expressions from their mental lexicon and large repertoires of "preferential ways" of saying things.

→ To this respect, Pawley and Syder (2000), Wray (2002) and Nesselhauf (2005) claim that psycholinguistic evidence shows that the human brain is much better at memorizing than processing, and hence, having at their disposal a large number of prefabricated units may reduce their processing effort and, therefore, enhance language fluency.

→ In addition, we also have to take into account that the feeling of being linguistically limited may signal the need that some types of learners have to feel like a member of certain groups, such as the executives of a company, translators or teachers in training, the latter being part of the target audience of our investigation. Wray (2002: 75) corroborates this statement by mentioning that knowledge of a wide range of prefabricated units, and in this context we may also include collocations, serve to indicate that such individual belongs to a particular language group, that is to say, to show the learners' desire "to sound [and write] like others".

→ Even though there may be a consensus among many researchers that collocations should be taught (explicitly) in foreign language classes, and that they are key word combinations to achieve fluency (Fontenelle, 1994; Granger, 1998; Martelli, 2007; Nesselhauf, 2005; Orenha-Ottaiano, 2004, 2012, 2015; Sinclair, 1991, 2004; Thomas, 2014; Torner; Bernal, 2017; Wray, 2002), there is still no better or more accurate reply, approach or course of action on how to teach them.

→ Therefore, this paper aims to discuss the relevance of collocations to EFL teaching and learning, taking into account the fact that they may not still be extensively and systematically explored in the classroom, even in the past few years.

→ Besides that, as there has still been a lack of teaching support materials which could enable students to achieve collocational competence, we also have the purpose of describing the development of an *Online English Collocations Platform*[1]. By that, we hope to contribute to fill this gap, at least to a certain extent, in an attempt to find a course of action on how to teach them more effectively.

1. Collocations: definition and role in foreign language teaching and learning

→ The term 'collocation' has been commonly ascribed to Firth (1957), defined in his article entitled *Modes of Meaning*, when explaining cases of lexical syntactic co-occurrences, that is to say, "words that usually go together". The author is also well-known for the characterization of a collocation when he asserts "you shall know a word by the company it keeps" (Firth, 1957: 11),

Nevertheless, Handl (2008) points out that both Bartsch (2004: 30) and Mitchell (1971: 35) claim that the term was actually used in the 1950s by H. E. Palmer and still earlier by Otto Jespersen in 1917.

→ In what regards to the definition of collocations, Nesselhauf (2005: 1) characterizes them as "[…] arbitrarily restricted lexeme combinations". The author adds that they are made up of more than one word and are lexically and/or syntactically fixed to a certain degree. Heylen and Maxwell (1994: 299) give a broader definition when they mention that collocations are "cohesive, recurrent, arbitrary combinations of words which are not idioms but in which the (figurative) meaning of one part is contextually restricted to the specific combination". Hausmann (1984), in his turn, draws our attention to the fact that the most important aspect of collocations is their "status of mental disponibility as a whole, not as a creation produced ad hoc by a speaker".

→ Having these ideas or concepts in mind and also based on well-known researchers who have investigated collocations, as previously mentioned, we propose a definition for collocations which may suit the investigations we have carried out, as part of our umbrella research project called *The compilation of corpora-based teaching materials and specialized glossaries* and their contribution to a Lexicon and Translation Pedagogy and our research group FRASCORP - *Fraseologia e Colocações Baseado em* Corpora (PHRASCORP - *Corpus-based Phraseology and Collocations), duly registered at CNPq Group Directory* (Conselho Nacional de Desenvolvimento Científico e Tecnológico/*The Brazilian National Council for Scientific and Technological Development*).

---

1 Accessed from: http://www.institucional.grupogbd.com/workbook/index.

→ To our view, collocations are 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. They are a language and a culture's specific combinations and, as such, the collocability of their elements may vary significantly from a language to another, being thus made up of their own collocations network. We display below a figure which may illustrate what we understand by collocations:

Figura 1: Definition of collocations

→ The classification of collocations used in this study is based on Hausmann's (1984, 1985) and a taxonomy expanded by Orenha-Ottaiano (2004, 2009) from Hausmann's: verbal, nominal, adjectival and adverbial collocations. The fig-ure below shows the taxonomy with examples of collocations taken from *EnTenTen15*, from the bases (nodes) *platform*, *leadership*, *strong*, *developer*, *develop* and *competence*.
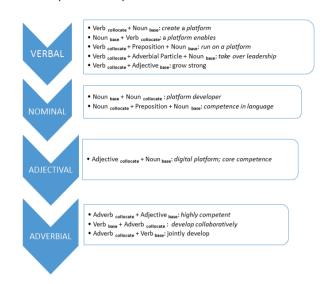

Figure 2: Taxonomy of collocations

→ With respect to the elements of a collocation, Hausmann (1985) points out two of them: a base and a collocate, each with different semantic status. There is actually a hierarchy between these two elements, as one of them (the base) determines and the other (the collocate) is determined. Simply speaking, the base is what we have already known and the collocate is the element we are searching for. The base is an independent semantically autonomous element that determines which lexical patterns can combine with it. On the other hand, the collocate works as a modifier; it is semantically interpretable within a collocation, and it is chosen by a certain base to form a collocation (Heid et al., 1991).

→ Another issue worth discussing for this investigation is concerning the degree of fixedness or restrictedness of collocations. This aspect may help explain how, besides frequency, collocations were chosen and selected for the Platform

exercises, as well as contribute to justify the selection of collocations for the three different levels of difficulty the user can choose from in order to play the Memory Game or do the Gap Fill available on the Platform. Obviously, that is such a challenging, arduous and time-consuming task, if we take into account the huge number of collocations that comprises a language and their pervasiveness or ubiquitous feature. Nevertheless, it is important to have these features in mind when categorizing the collocations on the Platform according to their degree of difficulty as it may be a way out to help learners retrieve them.

→ Regarding then the degree of fixedness or restrictedness of collocations, Lewis (1997) uses the terms *unique*, *strong*, *medium-length* and *weak* collocations when it comes to collocational strength. We drew the figure below to show these terms in a spectrum with a brief explanation and examples given by the author:

→ Lewi's classification for collocations learning seems interesting, however, how can we set the limits for distinguishing a strong from a medium-length collocation? The reflections and contributions coming out of his views are highly important, but using this classification for choosing the collocations which would comprise the platform exercises would be fairly difficult.

→ O'Dell and McCarthy, in both 2008 and 2017 editions of *English Collocations in Use* for advanced students, classify collocations as *strong*, *fixed* and weak. A *strong collocation* is "one in which the words are closely associated with each other. For example, the adjective *mitigating* almost always collocates with the *circumstances* or *factor*. It rarely collocates with any other word" (O'Dell; McCarthy, 2017, 2008: 8).

→ As for the *fixed collocations*, the authors define them as collocations which are "so strong
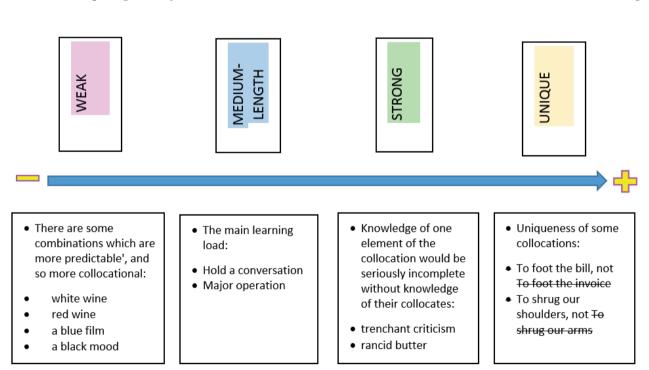

Figure 3: Adaptation of Lewis' classification for degree of fixedness or restrictedness of collocations

that they cannot be changed in any way". The example given is *walk to and fro*, and they explain that no other words can replace *to* or *and* or *fro* in this collocation as it is completely fixed. They add that these collocations are called idioms as their meanings cannot be guessed from the individual words (O'Dell; McCarthy, 2017, 2008: 8).

→   *Weak collocations* are made up of words that collocate with a wide range of other words. They exemplify with the collocation *in broad agreement*, explaining that the collocate *broad* can be used with several other words, such as a *broad avenue, a broad smile, a broad accent* etc., and that is why it is considered as a weak collocation.

→   According to the authors, most collocations lie somewhere between the two types of collocations (strong and weak) as in *a picturesque village* (p. 8).

→ Nevertheless, we once more end up by having some difficulty in identifying or classifying a collocation as strong, something in between or weak. Besides that, if a collocation is so fixed and is taken to be compositional, it may not be a collocation, but an idiom - as the authors themselves added -, and as such, would not be called or considered to be a *collocation*, according to the taxonomy and theoretical perspective here adopted..

→   Howarth (1996), in his turn, aimed at an investigation focusing on where free collocability ends and restriction begins, taking into account that "learners of a language need guidance on when generative mechanisms of language production are appropriate and when they need to draw on phraseological competence" (Howarth, 1996: 102), For this reason, he establishes a scale of restrictedness for verbal collocations.
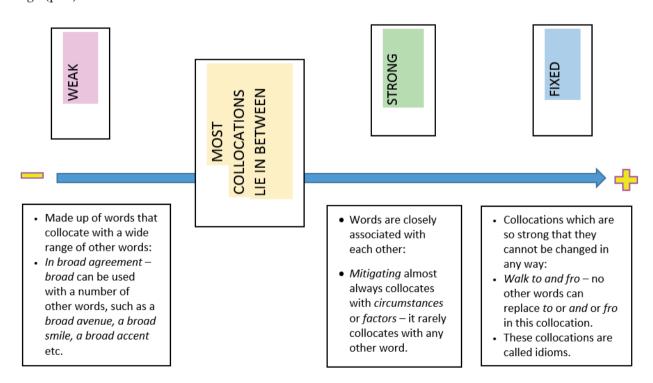

Figure 4: Adaptation of O'Dell & McCarthy's classification for degree of restrictedness of collocations

→ In the first phase of the scale, the author identified that there is some freedom of substitution in the noun while there is some restriction on the choice of verb. This means, according to them, that there is an open set of nouns and a small number of synonymous verbs, as in *adopt/accept/agree* to a *proposal /suggestion /recommendation /convention /plan*.

→ The second phase is characterized by having some substitution in both elements of the collocation, that is to say that "a small range of nouns can be used with a verb in that sense there are a small number of synonymous verbs" (Howarth, 1996: 102), as in *introduce/table/bring forward a bill/an amendment*.

→ The third category is said to allow some substitution in the verb, however, there is a complete restriction on the choice of the noun, which implies that no other noun can be used with the verb in that sense and that there are a small number of synonymous verbs: *pay/take heed*.

→ As for the fourth phase, there can be identified a complete restriction on the choice of the verb and some substitution of the noun or, as the writer adds, a small range of nouns can be used with the verb in the given sense and there are no synonymous verbs. For example: *give the appearance/impression*.

→ The fifth and last category describes a combination with a complete restriction on the choice of both elements, that is, no other noun can be used with the verb in the given sense there are no synonymous verbs, such as the verbal collocation *to curry favour.* In the figure below we try to summarize the scale of restrictedness proposed by Howarth as described.
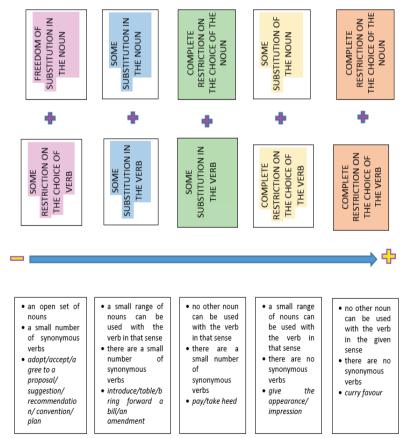

Figure 5: Adaptation of Howarth's classification for degree of fixedness or restrictedness of collocations

→ Even though Howarth's classification and exemplification are restricted to verbal collocations, it seems to be a more detailed and comprehensive scale and, therefore, it may help identify the different degrees of collocation restrictedness and understand how collocational competence may be enhanced. Due to that, Howarth's scale helped us define one more criteria for selecting and categorizing collocations according to the three levels of difficulties provided by the Platform (*Easy*, *Medium* and *Hard*), besides the criteria of frequency.

2. Methodology for collocations extractions, selection and insertion into the Platform

→ The Platform activities had been initially planned to target only Brazilian Portuguese speakers who were learning English as a foreign language.

→ Due to that, some collocations that comprise the activities on the Platform were extracted based on the difficulties these students had as well as on their collocational errors, according to research data results (Orenha-Ottaiano, 2012a, 2012b, 2013; 2019; Brito & Orenha-Ottaiano, 2013) from the analysis of:

1) a Translation Learner Corpus (parallel) made up of texts translated from Portuguese into English by undergraduate students (B.A. in Translation), of approximately 100,000 words and compiled from 2010 to 2013;

2) a Learner Corpus composed of essays written in English by undergraduate students (B.A. in English language), of approximately 100,000 words, compiled from 2010 to 2013; and

3) a second Translation Learner Corpus, of approximately 140,000 words, which has been feeding the *Multilingual Student Translation Corpus* (*MUST*), This corpus compilation started in 2017 and will end in December 2019.

→ The MUST corpus is an international project under the direction of Drs. Sylviane Granger and Marie-Aude Lefer, from the Center for English Corpus Linguistics, Université Catholique de Louvain, and under our coordination, at São Paulo State University (UNESP), Brazil. It aims to collect a large multilingual translation corpus for students with standardized metadata that combines precepts of Learner Corpus Research and Corpus-Based Translation Studies. The corpus MUST corpus can be exploited through the Hypal4MUST interface, a custom version of Hybrid Parallel Text Aligner (Hypal) designed by Obrusnik (2014) for parallel word processing. The project currently includes 31 researchers from 18 countries and covers more than 15 languages (German, Chinese, French, Gallego, Dutch, English, Greek, Italian, Lithuanian, Macedonian, Norwegian, Polish, Portuguese, Brazilian Portuguese, Slovenian, Spanish).

→ Firstly, collocations were extracted from the words with the highest keyness from the aforementioned corpora. The decision to use these corpora lies in the fact that we wanted to make sure that the collocations which are difficult to be produced by Brazilian learners of English as a foreign language would be included in the Platform activities.

The most frequent collocations produced by the translation learners were extracted from the keywords. The ones which were not correctly or frequently used were then analyzed, and more suitable and frequently used collocational

patterns were raised instead, by searching for the same lexical item on *Sketch Engine*'s *English Web 2015* (*EnTenTen15*) corpus, with the use of *Word Sketch*. Later on, these collocational patterns were added to the Collocations Platform. It is worth mentioning that not only the collocations extracted from the referred learner corpora which posed problems to the learners were chosen for the exercises, but also other patterns taken from the most frequent words, generated from *EnTenTen15*.

→ It is also worth mentioning that over the last two years, more collocations have been added to the category *General Collocations* (collocations from generic language), taken from the collocational pattern data bank created for an *Online English-Portuguese Collocations Dictionary Platform* we have been compiling since 2012. The *Collocations Dictionary Platform* relies on the collocations extracted from *EnTenTen15*'s most frequent words from the *WordList*, with the help of *Word Sketches*.

→ As for the other categories, we have used different corpora to extract more specialized collocational patterns:

• *Medicine Category* - from the *House MD* and *Gray's Anatomy* corpora, formed by the cap-

tions of the two series Orenha-Ottaiano; Fiel, 2013; Rocha; Orenha-Ottaiano, 2012;

• *Business Category* - from a Corpus of Business made up of texts from the Business section taken from *The New York Times*, *The Financial Time*; *Businessweek* and *Time* magazine (Orenha-Ottaiano, 2004);

• *Politics Category* - from a Corpus of Politics constituted by texts from the Politics section from *The New York Times*, *Financial Time*; *Businessweek*, *Time*;

• *Academic Language Category* - from *Cambridge Academic English Corpus*, available at the *Sketch Engine*; and

• *Investigation Category* - from NCIS Corpus, *CSI - Crime Scene Investigation* Corpus, *Criminal Minds* Corpus, compiled for a Master's thesis by Caldas (2017); and *Law & Order: SVU* Corpus, compiled for an undergraduate scientific research, by Real (2016), both work under our supervision.

→ The figure below summarizes the Platform categories and the corpora that were responsible for feeding them with the collocations from the different fields as well as from generic language:
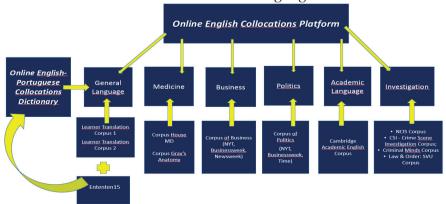

Figure 6: Corpora used for collocation extraction

→ Regarding the insertion of collocations, the Platform provides an online form, so that the researcher and team members can insert the collocation itself (node + collocate), choose the category, the taxonomy as well as the level of difficulty, as shown by the figure below:
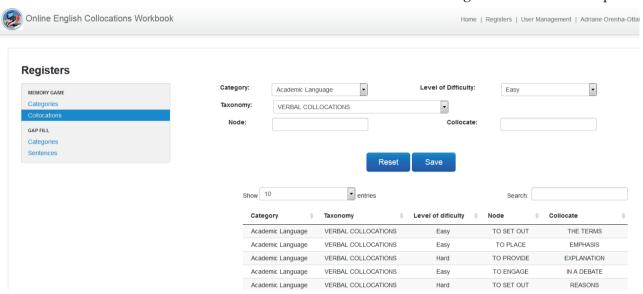

Figure 7: Platform form for collocations insertion

→ The next section will describe the *Online English Collocations Platform* in more details.

3. The Online Corpus-based English Collocations Platform

→ The purpose of compiling a corpus-based supporting material is to contribute to the teaching and awareness of the conventional and collocational aspects of a language and to the development of collocational competence, to teach English collocations to Brazilian Portuguese learners of English (mainly), as well as any other learner of English whose mother language is not Portuguese. Another reason that may justify this work is the scarcity of supporting material to teach phraseological units and, more specifically, to teach English collocations (refer to O'Dell,

McCarthy, 2011; McCarthy, O'Dell, 2017), as Lewis (1996, 1997, 2000), Howarth (1998), Aramo (2010), Higueras (2017), Jiang (2009) and Nesi (2009) have pointed out.

→ In addition, the number of supporting materials to teach English collocations to speakers of specific mother languages is rare. The significance of the latter lies on the idea that supporting material which focuses on a specific target audience may contribute to overcome specific difficulties learners of that particular L1 have and, hence, help them achieve collocational competence more easily.

→ Mackin (1978) had already stressed the importance of the development of supporting material to teach collocations and also the design of materials aiming at specific audiences of particular L1s: "what EFL teachers need for an effective training of collocational competence in their classes (as long as these are homogenous with regard to the L1) are workbooks presenting a selection of collocations geared to the specific difficulties of learners […]".

→ The researcher believed that the development of specific material would allow teachers to actually teach collocations and "thus shorten, at least to a certain extent, the long and laborious process of acquiring collocational competence through years of study, reading, and observation of the language" (Mackin, 1978: 151).

→ Other writers and researchers have also mentioned the importance of developing a more organized and systematic teaching material for collocations (Aramo, 2010; Higueras, 2017; Jiang, 2009; Nesi, 2009).

→ Jiang (2009) targets on the effect of the materials on collocation learning and Nesi (2009), for instance, reports on the concern with the design of materials to help learners recognize and reproduce appropriate collocations.

→ Having these points in mind, this part of the paper will describe how the Platform has been designed and developed so far.

→ The platform has been developed by the Data Bank Group (GBD), under the coordination of Professor Carlos Roberto Valêncio, from the Computing Department, also from São Paulo State University.

→ To have access to the *Online English Collocations Platform*, the user will have to register on the Platform web. The first page of the Platform will explain how this can be made, besides giving a brief context and definition of collocations:

Figure 8: Access to the Online English Collocations Platform

→ Both *Memory game* and *Gap Fill* have interactive activities, with different levels of difficulty: Easy, Medium and Hard, as shown in the figure below.

→ Furthermore, if one considers that one of the main target audience is Brazilian Portuguese-speaking learners of English, one of the first criteria would be to select collocations which would pose some difficulties regarding the influ-
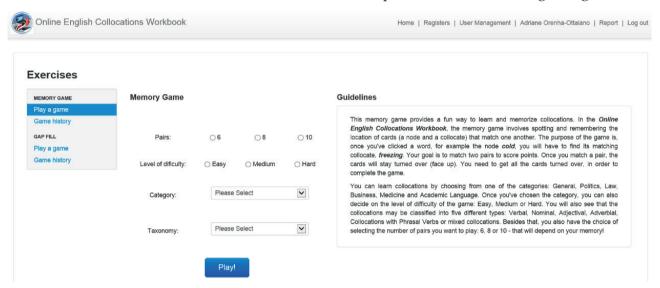

Figure 9: Screenshot with levels of difficulty for each collocational activity

→ The collocations that comprise each level of difficulty were chosen, besides frequency, according to their degree of lexicalization or degree of restrictedness. In fact, we would rather use the term "degree of restrictedness", borrowing Howarth's view of describing five different levels of collocational restrictions (Howarth, 1996), as previously mentioned. This decision implies that the more restricted collocations are, the harder they are to produce. Hence, they should be included in medium or hard level of difficulty.

→ Nevertheless, that may not be the sole criterion, as it is known that, even though some collocations may apparently be easier to produce, due to their low degree of restrictedness or because they are more literal to a specific language, learners may face difficulties in using them. Hence, some of these collocations could also be included in a medium or hard degree of difficulty.

ence of their mother tongue. This may imply that the aim would be to choose from those collocations which they would, by influence of Portuguese, transfer into English by using a not suitable collocation. If these are B2 or C1 students and would select *hard* degree of difficulty on the Platform, these collocations may not necessarily be the more restricted ones. Hence, all these aspects have to be taken into consideration before preparing the exercises.

→ Another issue to be taken into account is that the lexicalization or restrictedness boundaries of collocations are very difficult to draw (Agirre; Aldezabal; Pociello, 2006; Contreras; Sueñer, 2004; Cowie, 1990), turning our task even more complex.

→ For these reasons, some collocations were or have been evaluated by three different target audiences:

- BA in Translation undergraduates (every year since 2012),

- BA in English language undergraduates (every year since 2012), and

- Public School teachers who took part in an Extension course called "Phraseology and Corpus Linguistics applied to Public School English teachers", of 180 hours, under our supervision and duly approved by Sao Paulo State University Pro-rectorship of Extension (in 2015).

→ Through a questionnaire, the students and teachers evaluated the Platform exercises as well as the level of difficulty of collocations from the exercises they had worked on, giving feedback on whether they agreed with the classification (easy, medium or hard) of the collocations studied. Besides that, they also provided the researcher with information on how to improve

→ According to the questionnaire result, approximately 95% of the collocations evaluated were considered correctly classified. The other 5% were re-evaluated by the coordinator of the research project and classification was changed. It is worth mentioning that before this phase, the collocations were carefully classified and discussed by the research team that was responsible for selecting and inserting the collocations on the Platform. Later, they were laboriously checked out by the project coordinator.

→ In addition to choosing the levels of difficulty, the user can also opt for the number of pairs he or she would like to play: 6, 8 or 10 pairs, for the Memory Game.

→ After having chosen that, the user will have to decide on the category he or she would like to learn: *Academic Language*, *Business*, *General*, *Medicine*, *Politics* or *Investigation*.
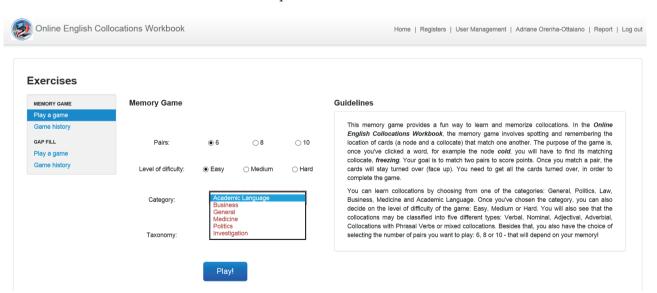

Figure 10: Screenshot of the different categories

the platform and reported on difficulties or any negative impression they had while doing the exercises.

→ Then, the last choice will be regarding the taxonomy of the collocations: whether verbal, adjectival, adverbial or nominal:
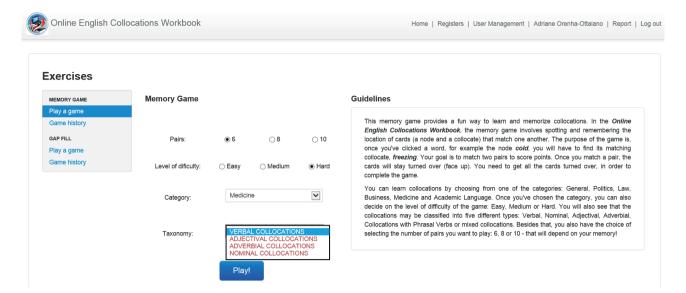
Figure 11: Screenshot of collocations taxonomy

→ In what regards the *Memory Game*, the user will click on *Play a game* and will receive a message saying he or she will have 25 seconds to memorize the pairs of words that will make up the collocations. After that, he will have to click the OK button so that he will be able to visualize them. At the end of the 25 seconds, all words will be hidden, and he/she will start playing the *Memory Game*.
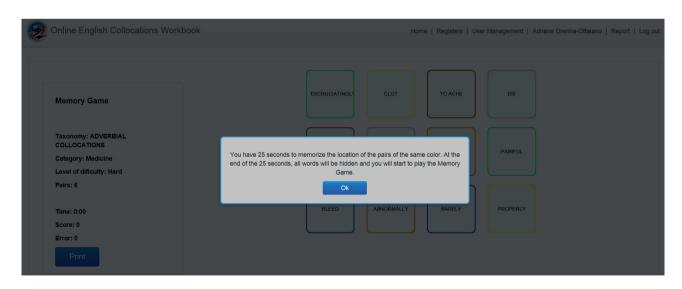

Figure 12: Screenshot of time for memorization of words

→ Below, there is a screenshot with all the words (pairs) that will make up the collocations to be learnt. As previously mentioned, each pair of words has one color, in order to show the collocation to the user and help memorizing where each element of the collocation is:

→ The next figure shows the user playing the *Memory Game*, in search for the pairs that will make up the collocations. It is important to note that, on the left side of the screen, the user will have a summary of his or her choices (*Category*, *Level of difficulty* and number of *Pairs*), as well as the time it took him or her to finish the game
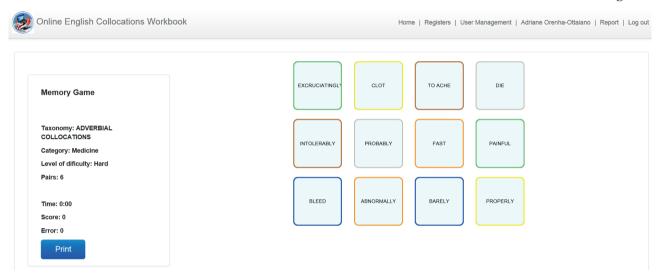

Figure 13: Screenshot of the pairs that will make up a collocation

→ When time is up, there will come up a screenshot informing the user he or she is ready to start playing:

(*Time*), his or her score (*Score*) and the number of combination mistakes he or she had (*Errors*). In the near future, learners will be able to choose
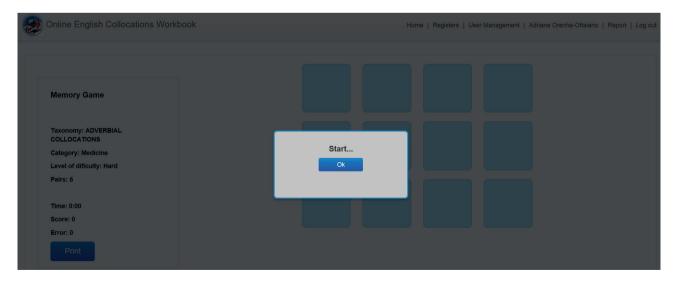

Figure 14: Screen informing the game can be started

whether they want to have these options at their disposal.

→ We decided to change it to optional, as some learners reported to feel either pressured or bored by such a resource, according to the results of questionnaires administered in the classroom since 2012, as stated in the methodology of this paper:

→ In the figure below, you can see a screen from which the user will start the *Gap Fill* exercise. The *Gap Fill* consists of filling in the blanks with predefined and frequently used collocations. The user had been previously instructed that, in order to fill in the gaps, he or she should drag the collocate that will make up the suitable collocations in context provided. It is worth reminding that some collocations that comprise the *Gap Fill* exercise are those that might have
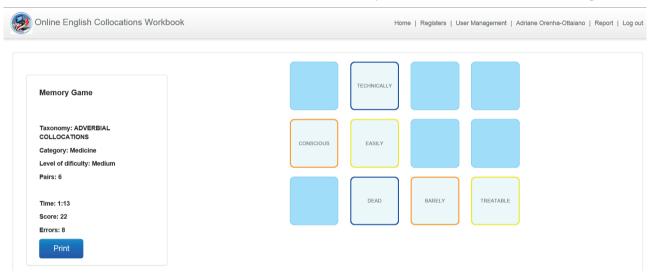

Figure 15: Playing the *Memory Game*

→ In the following figure, we present the *Gap fill* exercise screen, following the same pattern as the *Memory Game*.

already been worked on in the *Memory Game.* Thus, the user may have an opportunity to see them in context. It is a random selection for making up both games, however, if the learner
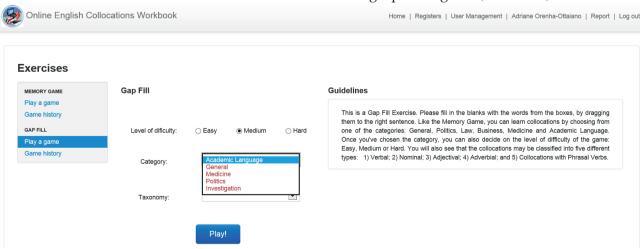

Figure 16: Selection of items to play the *Gap Fill*

uses the Platform, he may have higher chances of working again the same collocation practiced before in the *Memory Game*.
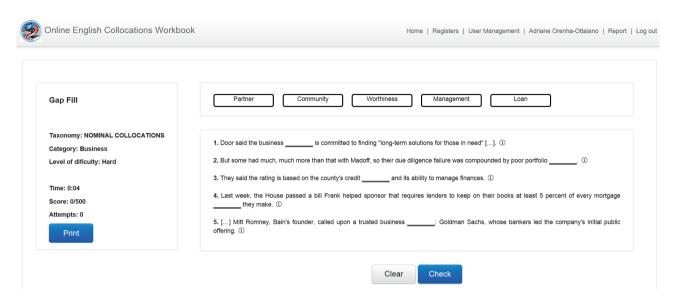

Figure 17: Screenshot of a *Gap Fill* activity

→ The following figure shows how verbs are dragged to the sentence which the user believes that will make up a collocation:
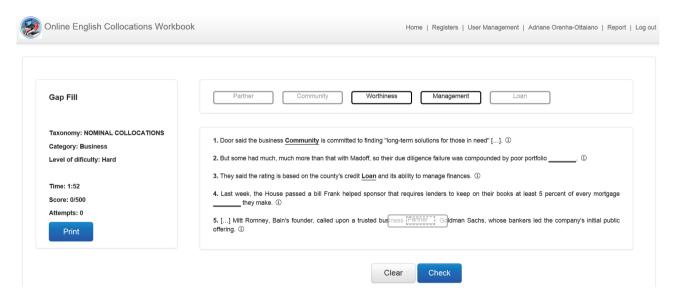

Figure 18: The user dragging a verb that will make up a collocation in context

➔ We simulated some collocational errors for this work, so that the user can notice that, once finished the exercise, he or she can click on the button *Check*, in order to verify if the options chosen are correct. By doing so, he or she will obtain the following result:

➔ Due to the complexity of corpus-based collocational pattern extractions, analysis and categorizing (taxonomy, level of difficulty, for example), not to mention funding limitations, we have so far developed two types of activities, the *Memory Game* and the *Gap Fill*. In a near future,
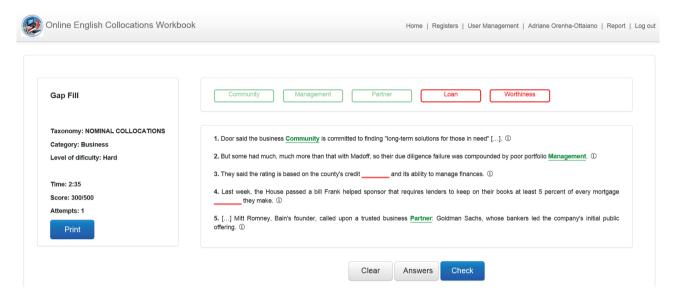

Figure 19: Correction of the *Gap Fill* exercise

➔ After that, the user may choose whether he wants to keep on trying or to be shown the Answers. By analyzing the next figure, the user can have access to the source of the contexts from which the collocations were taken, by clicking the icon : ⓘ

more exercises and different types of activities will be added, so that the Platform may be more user-friendly, more resourceful and practical to a wider range of English learners.
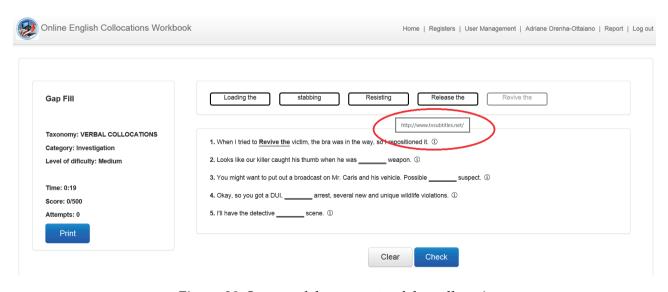

Figure 20: Source of the contexts of the collocations

Conclusion

→ In this paper, we discussed the role and importance of collocations in foreign language learning and teaching, drawing attention to the relevance of producing supporting material whose goal is to develop learners' collocational competence.

→ As a contribution, we described the methodological steps for the creation of an Online English Collocations Platform, designed to help Brazilian Portuguese as well as other foreign language speakers to learn English collocations.

→ Under a Corpus Linguistics perspective, the benefits and practical results from this investigation may reflect on the target audience's environment. The resource offered may provide a more systematized and interactive learning of collocations and enable them to achieve collocational competence and fluency in English.

→ Nevertheless, the challenges still lie ahead, in the sense that we should aim to constantly improve and review what it has already been developed on the Platform, as well as create other activities and exercises that may fulfill learners' collocational needs.

References

AGIRRE, Eneko / ALDEZABAL, Izaskun / POCIELLO, Eli. (2006), Lexicalization and multiword expressions in the Basque WordNet. *Third International WordNet Conference Proceedings*, GWC.

BAHNS, Jens / ELDOW, Moira. (1993), Should we teach EFL students' collocations? *System* 21(1): 101-114.

BARFIELD, Andy / GYLLSTAD, Henrik (eds), (2009), *Researching Collocations in Another Language: Multiple Interpretations*. London: Palgrave MacMillan.

BENSON, Morton. (1985), Lexical Combinability. *Papers in Linguistics* 18/1: 3-15.

BRITO, Yasmin / ORENHA-OTTAIANO, Adriane. (2013), Análise das colocações extraídas do Corpus de Aprendizes de Tradução (CAT), In: *XXXV Congresso de Iniciação Científica da UNESP,* São José do Rio Preto (SP).

CALDAS, Ariane / RIBEIRO Donizete Delgado. (2015), *A identificação de colocações especializadas extraídas do Corpus CSI e do Corpus Comparável Criminal para a elaboração de atividades didáticas*. Master's Thesis, Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP).

CONTRERAS, Joan-Miquel / SUÑER, Avel·lina (2004), Los procesos de lexicalización. In: E. Perez Gaztelu / Zabala, I. / Gràcia, L. (eds.), *Las fronteras de la composición en lenguas románicas y en vasco*. Bilbao: Universidad de Deusto, 47-109.

CONZETT, Jane. (2000), Integrating collocation into a reading & writing course. In: Lewis, Michael (ed.) *Teaching Collocation*. Boston: Heinle. 70-87.

CSI: CRIME SCENE INVESTIGATION. Available at: http://www.tvsubtitles.net. Accessed: July 2014.

FERRANDO ARAMO, Verónica. (2010), *Materiales didácticos para la enseñanza-aprendizaje de las colocaciones*. Master's thesis. Tarragona:

Universitat Rovira i Virgili. Available at Biblioteca Virtual RedELE 11.

FONTENELLE, Thierry. (1994), 'Towards the construction of a collocational database for translation students.' *Meta* 39/1: 47-56.

GRANGER, Sylviane. (1998), Prefabricated patterns in advanced EFL writing: collocations and formulae. In: Cowie, A. (ed.), *Phraseology: theory, analysis and applications*. Oxford: OUP, p. 145-160.

GRANGER, Sylviane. (eds.), (2008), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam & Philadelphia: John Benjamins.

HANDL, Susan. (2008), Essential collocations for learners of English. In: Meunier, F. / S. Granger (eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins. 43-66.

HEID, Ulrich / 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.

HIGUERAS GARCÍA, Marta. (2017), Pedagogical principles for the teaching of collocations in the foreign language classroom. In: Torner, S. / Bernal, E. (eds), *Collocations and Other Lexical Combinations in Spanish*. London: Routledge.

HILL, Jimmy. (2000), Revising priorities: from grammatical failure to collocational success. In: Lewis, Michael (ed.), *Teaching Collocation*. Boston: Heinle. 70-87.

HOWARTH, Peter (1996), *Phraseology in English Academic Writing. Some implications for language learning and dictionary making*. Tübingen: Niemeyer.

HOWARTH, Peter (1998), Phraseology and second language proficiency. *Applied Linguistics*,19 (1), 24-44.

JIANG, Jingyi. (2009), Designing Pedagogic Materials to Improve Awareness and Productive Use of L2 Collocations. In: Barfield, A. / Gyllstad, H. (eds), *Researching Collocations in Another Language: Multiple Interpretations*. London: Palgrave MacMillan.

KJELLMER, Göran. (1991), A mint of phrases. In: Aijmer, K. / Altenberg, B. (eds.), *English Corpus Linguistics: studies in Honour of Jan Svartvik*, London: Longman, 111-127.

LAW & ORDER: SPECIAL VICTIMS UNIT - Episode List. Available at: <http://www.imdb.com/title/tt0203259/episodes?ref_=tt_ql_6 >. Access: April 2015.

LEWIS, Michael. (1996), *The Lexical Approach*. London: Language Teaching Publications.

LEWIS, Michael. (1997), *Implementing the Lexical Approach*. London: Language Teaching Publications.

LEWIS, Michael. (2000), *Teaching collocations*. London: Language Teaching Publications.

MARTELLI, Aurelia (2007), *Lexical Collocations in Learner English: a corpus-based approach*. Alessandria: Edizioni dell'Orso.

MACKIN, Ronald (1978), On collocations: words shall be known by the company they keep. In: Strevens, P. (ed.) *In Honor of A. S. Hornby*. Oxford University Press. 149-165.

McCARTHY, Michael / O'DELL, Felicity (2017), [2005]. *English Collocations in Use Intermediate Book with answers: how words work together for fluent and natural English* . Cambridge: University Press.

MCINTOSH, Collin / FRANCIS, Ben / POOLE Richard (eds.) (2009), *Oxford Collocations Dictionary for Students of English*. 2nd ed. Oxford: Oxford University Press.

MEUNIER, Fanny / GRANGER, Sylviane (2008), Phraseology in foreign language learning and teaching. Where to and from?' In: Meunier, F. / Granger, S. (eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins. 247-252.

NESI, Hilary (2009), Exploring Materials for the Study of L2 Collocations. In: Barfield, A. / Gyllstad, H. (eds), 2009. *Researching Collocations in Another Language: Multiple Interpretations*. London: Palgrave MacMillan.

NESSELHAUF, Nadja. (2005), *Collocations in a Learner Corpus*. Amsterdam / Philadelphia: John Benjamins.

O'DELL, Felicity / McCARTHY, Michael. (2017), [2008]. *English Collocations in Use Advanced Book with answers: how words work together for fluent and natural English* . Cambridge: University Press.

ONLINE ENGLISH COLLOCATIONS PLATFORM. [Accessed from: http://www.insti-tucional.grupogbd. com/workbook/index. 34(1), 241-251.

ORENHA-OTTAIANO, Adriane (2004), *A compilação de um glossário bilíngue de colocações, na área de jornalismo de Negócios, baseado em corpus comparável.* Master's thesis, Universidade de São Paulo.

ORENHA-OTTAIANO, Adriane (2012ª), English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. *Acta Scientiarum*, 34 (2), :241-251.

ORENHA-OTTAIANO, Adriane (2012b), Compilação de um corpus de aprendizes de tradução e análise de aspectos colocacionais. In: Abralin em cena, *Proceedings of Abralin* Cuiabá.: UFMT.

ORENHA-OTTAIANO, Adriane (2013), Corpus de aprendizes de tradução: subsídios para o desenvolvimento de uma pedagogia do léxico e da tradução. In: *61ᵉ seminário do GEL / Proceedings of 61th seminar of GEL*, São Paulo: UNESP.

ORENHA-OTTAIANO, Adriane (2019), Escolhas colocacionais a partir de um Corpus de Aprendizes de Tradução e a importância do desenvolvimento da competência colocacional. Submitted to *Cadernos de Fraseoloxia Galega* (under review).

ORENHA-OTTAIANO, Adriane / FIEL, Roberta Pereira. (2013), Glossário de colocações da língua geral e especializadas baseado em corpus paralelo: uma contribuição para o ensino de LE e para a tradução. *Confluência* (Rio de Janeiro), vol. 44-45, 309-332.

OBRUSNIK, Adam. (2014), Hypal: A User-Friendly Tool for Automatic Parallel Text Alignment and Error Tagging. Paper presented at the *Eleventh International Conference Teaching and Language Corpora*, Lancaster, 20-23 July 2014, 67-69.

PAWLEY, Andrew / SYDER, F. H. (2000. The One-Clause-at-a-Time Hypothesis. In: Riggenbach, H. (ed.), *Perspectives on Fluency*. Ann Arbor: The University of Michigan Press,

REAL, Laís Magalhães. (2016), *O estudo das colocações extraídas do Corpus Law and Order*. Undergraduate Monograph, São Paulo: Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP).

ROCHA, Jean Michel Pimentel / ORENHA-OTTAIANO, Adriane. (2012), Colocações especializadas na área médica extraídas a partir do corpus house MD. *Cadernos do IL*, vol. 44, 295-318.

SINCLAIR, John. McH. (1991), *Corpus, concordance and collocation*. Oxford: University Press.

SINCLAIR, John. McH. (2004), *How to use corpora in Language Teaching*. Amsterdam: John Benjamins.

THOMAS, J. (2014), Stealing a march on collocation. *TALC 10 Proceedings*.

STUBBS, Michael (2005), *Words and phrases:* Corpus studies of lexical semantics. Oxford: Blackwell Publishing.

TER-MINASOVA, S. (1992), The freedom of word-combinations and the compilation of learners' dictionaries. In *Euralex '92 Proceedings*, Tampere: University of Tampere, 533-539.

TORNER, Sergi / BERNAL, Elisenda (eds), (2017), *Collocations and Other Lexical Combinations in Spanish*. London: Routledge.

WRAY, Alison. (2002), Formulaic language and the lexicon. Cambridge: Cambridge University Press.

Biographic profile: Adriane Orenha-Ottaiano is a Lecturer at the Department of Modern Languages at the São Paulo State University (UNESP). She also works at the UNESP's Graduate Program in Linguistic Studies, in the research lines "Translation Studies" and "Corpus-based Translation and Lexicon Pedagogy". She has a Postdoctoral Fellowship in Translation from Université de Montréal, Canada (FAPESP scholarship – Process 2018/22943-8), a PhD in Linguistics Studies, from UNESP, a Master's in Linguistics and English Studies from University of Sao Paulo (USP) and a B.A. in Translation (UNESP). She coordinates the research projects "A phraseographical methodology and model for an Online Corpus-Based Multilingual Collocations Dictionary Platform", funded by FAPESP (Process 2020/ 01783-2) and "The compilation of an Academic Collocations Dictionary". She is the leader of the CNPq Research Group "PHRASCORP - Corpus-based Phraseology and Collocations". Her main research interests are Corpus-Based Translation Studies, Teaching English and Portuguese as a foreign language, Corpus Linguistics, Teacher Training and Corpus-based Phraseology and Phraseography, specially on collocations and collocations dictionary making.

e-mail: adriane.ottaiano@unesp.br