***AOO. Orenha-Ottaiano, Adriane***

***MEOOS. Oliveira Silva, Maria Eugênia Olímpio de***

**A CORPUS-BASED PLATFORM OF MULTILINGUAL COLLOCATIONS DICTIONARIES (PLATCOL): SOME LEXICOGRAPHICAL ASPECTS AIMING AT PRE- AND IN-SERVICE TEACHERS**

**Abstract**: This paper aims at describing an Online Platform for Multilingual Collocations Dictionaries (PLATCOL), highlighting its relevance to FL teaching and learning. We discuss some lexicographical aspects to develop a customized platform to meet pre- and in-service teachers’ needs. Its design, layout and part of the methodological procedures are based on the Bilingual Online Collocations Dictionary Platform (Orenha-Ottaiano 2017). The methodology relies on the combination of automatic methods to extract candidate collocations (Garcia et al., 2019a). Statistical measures (Garcia et al. 2019b), and distributional semantics strategies are applied to select the candidates (Garcia et al., 2019c), and extract examples (Kilgarriff et al., 2008).

 **Keywords**: collocations, collocations dictionary, corpus, corpus linguistics, lexicography, pre-service teachers, in-service teachers, customized platform.

1. **Introduction**

Corpus Linguistics has been successfully applied in different fields of Linguistics, especially in Foreign Language (FL) Teaching and Learning. More specifically, learner corpora have also contributed to a better understanding of the second language acquisition process as well as L2 vocabulary learning and teaching (Granger 1998, 2015). Besides that, Corpus Linguistics has also been regarded as a powerful tool as well as a suitable match to Lexicography, mainly concerning the retrieval of phraseological units and collocations.

Numerous studies have highlighted the challenges in the FL teaching and learning of collocations [Laufer 2011, Nesselhauf 2005, Martelli 2007, Orenha-Ottaiano 2013, 2020, Torner Castells and Bernal 2017] and a lot of proposals have been put forward to surmount these obstacles.

With a view to contribute to the teaching and learning of collocations and fill one of the gaps regarding the availability of specific lexicographical work, such as a collocations dictionary, we proposed the creation of an *Online Platform for Multilingual Collocations Dictionaries* (*PLATCOL*), in English, Portuguese, French, Spanish and Chinese, whose lexicographical aspects and features are here briefly described.

In Section 1, we address the issue of raising collocational awareness in pre- and in-service teachers to develop collocational competence and the importance of collocations dictionaries use to achieve that. Section 2 describes PLATCOL and focuses on some lexicographical aspects and other features necessary to develop a customized platform suitable to meet pre- and in-service teachers’ needs. Section 3 describes our research corpora and briefly outlines the methodological aspects for automatic extraction of corpus data. Finally, Section 4 presents concluding remarks and highlights some ideas for further work.

1. **Developing collocational competence in pre- and in-service teachers: the importance of collocations dictionaries use**

Developing collocational competence[[1]](#footnote-2) is one of the most challenging tasks for FL learners as well as pre-service teachers. This challenge can also be applied to in-service teachers who still face the same problem even after having graduated, according to our 20-year experience as university professors both in graduate (B.A. in Languages and B.A. in Translation) and post-graduate courses.

In this study, also as a major motivational factor for the creation of PLATCOL, we claim that, among the many ways of helping FL learners, pre- and in-service teachers to achieve collocational competence, the use of online collocations dictionaries can be regarded an important pedagogical tool. In order to get the most out of them, professors should raise pre-service teachers' awareness of collocations and in-service teachers should recognize their relevance to FL proficiency and fluency development. This way, they will be able to better identify the collocational patterns they are looking up and enhance retention.

In general, collocations are mentioned, directly or indirectly, as an object of study in different researches that address the topic of pedagogical lexicography [Higueras 2005, 2006; Pérez Serrano 2014; Torner and Bernal 2017;) which confirms the relevance of this type of unit for the teaching and learning of foreign languages.

According to Neshkovska (2018), training students to regularly consult collocation dictionaries is also highly recommended by FL researchers in general. She adds that collocation dictionaries can be «a significant tool when it comes to mastering collocations». The author also addresses the issue of FL learners being encouraged to look up collocations in dictionaries of collocations on a regular basis, not only in class but also outside the classroom environment. Teachers are also advised to develop activities based on collocation dictionaries.

Rezaeil and Davoudi (2016) carried out a research on the influence of electronic dictionaries on vocabulary knowledge. Even though the investigation does not focus on collocations, its results can also be extended or applied to collocation knowledge, regarding that a high proportion of language is formulaic in nature and is formed by the co-occurrence of lexical items that, in turn, make up the vocabulary of a language. They concluded that the use of electronic dictionaries can indeed improve vocabulary learning.

The authors also pointed out that students reported more interest and motivation to learn new words when they use electronic dictionaries – and we claim that it can also be extended to online dictionaries or collocation dictionaries. They added that «owing to the multiple modes of presentation and interesting and eye-catching nature of electronic dictionaries, they are embraced more easily by language learners in their attempt to learn new vocabularies in a second of foreign language». Besides these findings, it is worth mentioning the special issue of *RILE* (*Revista Internacional de Lenguas Extrangeras*) on digital pedagogical lexicography, whose presentation stresses the importance of digital dictionaries for language teaching (Nomdedeu 2019). It can be inferred that this type of work must be accessible, reliable and up-to-date and that the concept of dictionary is also a serious issue, as it must offer careful lexicographic treatment (good definitions, adequate lexical selection, inclusion of pragmatic and cultural data, etc.).

Another important aspect to be highlighted here is that electronic, and especially online dictionaries, which is the proposal of this paper, have changed the way users interact with the data they are looking up and that contributes to one of PLATCOL’s goals: help users master collocations. A clearer and eye-catching structure for the entries, with a neat and uncluttered page layout, with visually obvious features so that users do not need to understand how it is organized, etc., are all aspects this project lexicographers must have in mind if they aim at compiling an online customized dictionary.

We thus acknowledge the useful role of collocations dictionaries in raising pre- and in-service teachers’ awareness of collocations and we recognize them as a valuable pedagogical tool.

1. **A Customized Online Platform for Multilingual Collocations Dictionaries for pre- and in-service teachers**

 Researches that highlight the importance of collocations in the FL teaching-learning process ​​undoubtedly have a long and solid tradition. Many of them are based on neurolinguistic studies that confirm the relevance of the prefabricated elements in the construction of the oral and written discourse of speakers of any language [Lewis 1993]. However, most researches focus on aspects related to students, their needs and their learning experience; on rare occasions, attention is paid to teachers. Thus, in most situations, the previous knowledge that teachers have or should have on the subject, for example, are not considered, even though this may be considered a crucial issue. López Jiménez (2017), in a comparative study on lexical collocations in English and Spanish as L2 textbooks, found that many of the 208 teachers who participated in the research were unaware of the concept of lexical collocation. On the other hand, the information which teachers must have access to in order to be able to properly conduct the teaching process of collocations or the resources they usually use for their didactics is also not considered.

Taking into account this complex context, it is necessary, when developing a dictionary aimed at pre- and in-service teachers, to inquire about the specific needs of this type of user, in order to adapt its design and the lexicographic treatment given to collocations according to these needs.

We follow the recommendations of the functional theory [Fuertes Olivera and Tarp 2008, 2014; Tarp 2004] to define, on the one hand, the profile of users and their potential needs, and, on the other hand, to select and present the corresponding lexicographic data.

Hence, users’ profile to whom this proposal is directed is:

* in the case of pre-service teachers; language learners (student teachers) from higher education institutions trained to become professional language teachers.
* with respect to in-service teachers: additional language teachers, native or non-native ones, with specific training or degree in Languages.

Based on the definition of the profile, we establish an initial delimitation, of general character, of the typology of needs these users may have. For this, we draw on our experience as professors, researchers and lexicographers. We understand that pre- and in-service teachers will need to use the platform in tasks such as: proofreading and correction of written or oral texts or to expand their knowledge on the topic, when preparing teaching materials. These needs are linked, thus, to two of the specific extra-lexicographic or social situations identified, by the functional theory: 1. Communicative, in which a user can try to solve a problem related to production, reception, translation, proofreading and correction of written or oral texts; and 2. Cognitive, when the user needs or wants to expand their knowledge of something.

These situations have a direct influence on the configuration of the dictionary. Furthermore, following Tarp’s recommendations (2004: 233-234), in order to define this configuration, it is convenient to consider the lexicographic data that dictionaries generally offer about collocations. This reflection makes it possible to judge which data is useful, which ones must be improved and which new data must be considered.

On the one hand, we understand that the platform should provide a wide range of examples of usage of collocations, as well as data on their frequency of use. Thus, as shown in figures 1 and 2, each collocation is accompanied by an illustrative example. At the end of the example, users can click on “See more”, which will allow them to consult other usage examples. This way, we do not clutter the article with information that may demotivate the dictionary consultation. Data on collocation frequency and collocation statistics are also offered in "Advanced options".

The data available on the platform should allow users to develop pedagogical proposals, according to the methodological perspective they want to adopt. Hence, it is hoped that, after consulting the platform, they can:

a) propose pedagogical materials that serve both for the explicit and implicit teaching of collocations;

b) create activities that provide and encourage:

i) the development of autonomous learning;

ii) memorization of collocations.

iii) working with the different communicative skills.

iv) linking the collocations with lexical fields or communicative functions.

 We must not lose sight of the fact that it is essential to present this information in a clear, attractive and accessible way. Let us remember the results presented by Chen (2011), in a research about the effectiveness of several computer assisted English collocation tools. The author showed that teachers prefer to use resources that provide quick and easy access to colloquial information and offer many examples of usage.

On the other hand, it is also necessary to pinpoint that teachers do not always have the necessary training to enable them to use dictionaries properly. As a consequence, these works end up being underused [Hernández 2005; Ureña and Penadés 2020]. For this reason, it is essential to include a consultation guide specially designed for teachers, in order to instruct them in its correct use and show them their pedagogical applications.

The entries of the multilingual collocation dictionaries consist of the following elements:

● a headword, which corresponds to the basis of the collocations. This way, the headwords are made up of lexical items such as nouns, verbs and adjectives;

● a word class: a word class is placed right after the headword (the base of the collocation). In the case of these collocation dictionaries, they will be either a noun (n.), a verb (v.) or an adjective (adj.). If a word belongs to more than one word class, such as abstract (n.), abstract (v.) and abstract (adj.), they are shown in separate entries, so that the collocations, collocations structures and other pieces of information are easily organized;

● frequency of each headword;

● a definition – a brief definition of the different senses of the base will be provided. The decision of including a base definition is that the collocations can be duly organized according to each sense of the headword;

● usage examples: to illustrate how collocations are used, based on a specific meaning. Users will have the chance to choose from displaying from 1 to 5 examples.

Hence, users will be able to have a quicker access to the collocations they are searching for:



*Figure 1. Screenshot of PLATCOL’S basic structure of an Entry*

Besides the basic microstructure, *Advanced options* will be available if a user opts to sign in. Hence, according to a users’ profile, new dictionary structure will be available as follows:



*Figure 2. Screenshot of Advanced Option Microstructure (Entry is a verb)*

As this is an ongoing research, the lexicographical aspects and other features here reported are still under review and there are other issues that should be considered and tested.

1. **Automatic retrieval of bases, collocations and corpus-based examples**

A large corpus for each of the five languages of the platform was compiled using different source data, as the Table 1 below illustrates:



 *Table 1. Corpora Sizes and Sources*

This research methodology relies on the combination of automatic methods to extract candidate collocations [Garcia et al., 2019a] and takes advantage of NLP tools to annotate large corpora with lemmas, PoS-tags and dependency relations in the five languages.

Statistical measures [Evert et al. 2017; Garcia et al. 2019b], and distributional semantics strategies are applied to select the candidates [Garcia et al., 2019c], and retrieve corpus-based examples [Kilgarriff et al., 2008].

1. **Concluding remarks and future research**

PLATCOL aims to be user-friendly and user-based, hoping to make a contribution to the teaching and learning of collocations and to boost pre- and in-service teachers’ text production and collocational competence. When encouraging them to develop collocational competence, we automatically stimulate them to improve their fluency in a given foreign language.

Thanks to the interest collocations have aroused in the past decades, we have advanced in the knowledge of the characteristics of this type of lexical unit as well as of the aspects that must be considered in its lexicographic treatment. The great challenge facing researchers is, above all, to put this knowledge into practice and generate lexicographic works more suitable to the specificities of the object of study, the idiosyncrasies of its users and their lexicographic needs. This is undoubtedly an ambitious goal; however, we believe that the guidelines we have investigated in the development of PLATCOL will make it possible to achieve it.

**Acknowledgements**

We gratefully acknowledge the financial support provided by *The São Paulo Research Foundation* (*FAPESP*), Process ner 2020/01783-2.

**References**

*1. Chen, H.-J. H. (2011), Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. In Computer Assisted Language Learning 24, 1, pp. 59–76. DOI 10.1080/09588221.2010.526945.*

*2. Evert, S., Uhrig, P., Bartsch, S., Thomas Proisl*. (2017), E-VIEW-affilation – A large-scale evaluation study of association measures for collocation identification. In *Proceedings of eLex 2017–Electronic lexicography in the 21st century*: Lexicography from Scratch, pp. 531-549.

3. *Fuertes-Olivera, P. A., Tarp, S*. (2014), *Theory and Practice of Specialised Dictionaries*. Lexicography versus Terminography, Berlín/Boston: Walter de Gruyter.

4. *Fuertes-Olivera, P. A. & Sven Tarp*. (2008), La Teoría Funcional de la Lexicografía y sus consecuencias para los diccionarios de economía del español [Functional theory of Lexicography and its consequences for Spanish dictionaries of Economy]. *Revista de Lexicografía*, XIV, pp. 75-95.

5. *Garcia, M., García-Salido, M., Alonso-Ramos, M*. (2019a), Towards the automatic construction of a multilingual dictionary of collocations using distributional semantics. In *Proceedings of eLex 2019*: Smart Lexicography: pp. 747-762, Sintra.

6. *Garcia, M., García-Salido, M., Alonso-Ramos, M*. (2019b), A comparison of statistical association measures for identifying dependency-based collocations in various languages. In *Proceedings of the Joint Workshop on Multiword Expressions and WordNet* (MWE-WN 2019) at the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), pp. 49-59, Florence.

7. *Garcia, M., García-Salido, M., Alonso-Ramos, M*. (2019c), Weighted compositional vectors for translating collocations using monolingual corpora. In *Computational and Corpus-Based Phraseology* (EUROPHRAS 2019). Lecture Notes in Artificial Intelligence, 11755. Springer: pp. 113-128.

8. Granger, S. (ed.) (1998), *Learner English on Computer*. London and New York: Longman.

9. Granger, S.; Gilquin, G. & Meunier, F. (eds.). (2015), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 761 pp.

10. *Hernández Hernández, H*. (2005), Quince años después: estado actual y perspectivas de la lexicografía del español para extranjeros. In: Castillo, M.ª A. et al. (eds.), Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del XV Congreso Internacional de ASELE. Sevilla: Universidad de Sevilla, pp.465-472.

11. *Higueras, M.* (2005), Necesidad de un diccionario de colocaciones para aprendientes de ELE. In: Castillo, M.ª A. et al. (eds.), Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del XV Congreso Internacional de ASELE. Sevilla: Universidad de Sevilla, pp. 480-490.

12. *Higueras, M.* (2006), *Las colocaciones y su enseñanza en la clase de ELE.* Madrid: Arco Libros.

13. *Laufer, B.* (2011), The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. In *International Journal of Lexicography*, 24, 1, pp. 29-49.

15. *Lewis, M*. (1993), *The Lexical Approach. The State of ELT and a Way Forward.* Londres: Language Teaching Publication.

15. *López Jiménez, M. D*. (2017), Colocaciones léxicas: Estudio comparativo de libros de texto de inglés y de español como L2. In *Odisea***,** 14, pp. 101-120.

16. *Ureña Tormo, C., I. Penadés Martínez* (2020), Análisis del uso del diccionario en L2. In: *Logos: Revista de Lingüística, Filosofía y Literatura*, 30, 1, pp. 154-170.

17. Orenha-Ottaiano, A. (2020), The creation of an Online English Collocations Platform to help develop collocational competence. In *PHRASIS*. *Rivista di Studi Fraseologici e Paremiologici*. pp. 59-81.

18. Orenha-Ottaiano, A. (2017), The compilation of an Online Corpus-Based Bilingual Collocations Dictionary: motivations, obstacles and achievements. In *Proceedings of E-Lex Conference 2017*. Leiden, The Netherlands, pp. 458-473.

19. Orenha-Ottaiano, A. (2013), Collocations and the design of teaching materials for second language learners. In: *Proceedings of the 10th International Conference on Teaching and Language Corpora*. Varsovia, Polony: Warsaw: Institute of Applied Linguistics, pp. 93-103.

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

21. *Nesselhauf, N.* (2005), *Collocations in a Learner Corpus*. Amsterdam & Philadelphia: John Benjamins.

22. *Neshkovska, S.* (2018), *What do advanced ESL\EFL students’ need to know to overcome ‘collocational’ hurdles*? Thesis. Vol. 7. Iss. 2. Pristina: AAB College, pp. 53-74.

23. Nomdedeu Rull, A. (coord.), (2019), Lexicografía pedagógica digital [Digital pedagogical lexicography]. In *Revista Internacional de Lenguas Extranjeras*, 10, <https://doi.org/10.17345/rile10>.

24. Pérez Serrano, M. (2014), ¿Son indispensables los diccionarios combinatorios? [Are combinatory dictionaries indispensable?] In: *Revista de Lexicografía*, 20, pp. 121-145.

25. *Tarp, S.* (2004), Basic Problems of Learner’s Lexicography. In *Lexikos,* 14, pp. 222-252.

26. Torner Castells, S., E. Bernal (2017), *Collocations and Other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. New York: Routledge.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Adriane Orenha-Ottaiano**

São Paulo State University – UNESP (Brazil)

***E-mail:*** ***adriane.ottaiano@unesp.br***

**Maria Eugênia Olímpio de Oliveira Silva**

University of Alcalá (Spain)

***E-mail:*** ***eugenia.olimpio@uah.es***

1. By collocational competence we understand the ability to identify, understand and, mainly, to produce collocations in a given language context, for translational purposes, for teaching or communicating in a given language. [↑](#footnote-ref-2)