

## The role of collocation in the RA

Eva Lucía Jiménez Navarro  
*Universidad de Córdoba*  
l42jinae@uco.es

Fecha de recepción: 30.06.2012  
Fecha de aceptación: 01.09.2012

**Abstract:** A collocation is a combination of two or more words which frequently occur together (McCarthy and O'Dell 2008). In this paper, we will show how the mastery of these word combinations is a must if one is to publish in a second language at an international level. As English has become the language of science (Crystal 2005; Montgomery 2009), and over 80% of scientific publication is done through this language, researchers who have a good command of it should aim at publishing in English. The work presented here aims at providing some tips on how non-native researchers who want to publish RAs at an international level can improve their collocational competence in English.

**Keywords:** collocation, collocational competence, international level, scientific publication, research article.

## El papel de la colocación en el artículo científico

**Resumen:** Una colocación es una combinación de dos o más palabras que frecuentemente aparecen juntas (McCarthy y O'Dell 2008). En este artículo, mostraremos cómo el dominio de estas combinaciones de palabras es imprescindible para publicar a un nivel internacional en una segunda lengua. Puesto que el inglés se ha convertido en el lenguaje de la ciencia (Crystal 2005; Montgomery 2009), y más del 80% de las publicaciones científicas se hace a través de este idioma, los investigadores que tienen un buen dominio de éste deberían proponerse publicar en inglés. El trabajo que presentamos pretende dar algunos consejos sobre cómo los investigadores no nativos que quieren publicar artículos científicos a un nivel internacional pueden mejorar su competencia colocacional en inglés.

**Palabras clave:** colocación, competencia colocacional, nivel internacional, publicación científica, artículo científico.

**Sumario:** Introduction. 1. Towards a definition of collocation. 1.1. Approaches to the notion of collocation. 1.2. General characteristics of the collocations. 1.3. Syntactic types of

collocations. 2. The importance of mastering collocations for English academic writing. 3. Some notion on collocation learning. Conclusions.

### **Introduction**

A collocation is a combination of two or more words which frequently occur together (McCarthy and O'Dell 2008). Their learning has proved to be an area that has not received the attention it deserves as these units have traditionally been neglected in the teaching of second languages in favour of the teaching of more idiomatic units, such as phrasal verbs and idioms, and grammar. For this reason, students are sometimes unaware of this type of combinations, regardless of their mastery of the language since "Even very advanced learners often make inappropriate or unacceptable collocations" (McCarthy 1990: 13). However, collocational knowledge becomes essential for producing natural sounding English both in the spoken and written types of discourse. According to Nation (2001: 56), the correct use of word sequences makes second language learners (henceforth L2L) sound like native speakers. It also contributes to the fluency with which language can be used. In addition to this, the largest part of the English speaker's lexicon consists of collocations (Pawley and Syder 1983), and many words are recurrently used in a limited set of collocations. Therefore learning these combinations involves learning the words that constitute them (Nation 2001: 56). In this regard, Kjellmer suggests: "In all kinds of texts collocations are essential, indispensable elements (...) with which our utterances are very largely made" (1987: 140). This way, knowing these word combinations means that learners know the language better and can use vocabulary in more appropriate set of contexts.

Although collocations do not depend on grammatical or semantic rules, what makes them very difficult to be learned by non-native speakers, it is generally accepted that they should be learned as "chunks"<sup>1</sup> of language so that they can be stored in our brain. This viewpoint is supported by Palmer who points out: "Each [collocation] (...) must or should be learnt, or is best or most conveniently learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts" (1933: 4). By the same token, Pawley and Syder (1983) claim that L2L, apart from knowing the rules of a language, should store thousands of preconstructed clauses as chunks in the memory and draw on them in their language use. Similarly, Ellis argues that both language knowledge and language use are based on associations of words stored in our brains. Ellis's

---

<sup>1</sup> This term was coined by MILLER (1956) to refer to permanent sets of associative connections in long-term storage.

point of view is summarized in the following statement (2001: 45): “Thus native-like selection is not a matter of syntactic rule alone. Speaking natively is speaking idiomatically, using frequent and familiar collocations, and learners thus have to acquire these familiar word sequences.”

## **1. Towards a definition of collocation.**

### *1.1. Approaches to the notion of collocation.*

In the study of collocations we can distinguish three main approaches: (1) the Firthian approach, (2) the Russian phraseology, and (3) the mixed approach. The first one, also known as “statistical approach”, goes back to J. R. Firth (1957) and has been further developed by the neo-Firthians M. A. K. Halliday and J. Sinclair. The two main notions they support are: (a) collocations must be identifiable on the basis of their frequency, and (b) collocations are independent of any grammatical structure. For instance, Firth defines collocations as follows: “Collocations of a given word are statements of the habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order” (1957: 12). By *habitual or customary* the author means that both of the constituent elements of a collocation tend to co-occur with each other.

The second approach, also known as “linguistic approach”, is born to a group of Russian scholars, led by V. V. Vinogradov, in the 1940s. Unless the statistical approach, they emphasize the linguistic status of collocations. So they consider a main inclusive category, which is the “phraseological unit”, and classify the different phraseological units according to the degree of semantic cohesion between their component elements. These authors regard collocations as combinations of lexical units where one of their constituents is used in its literal sense while the other one has a non-literal meaning (Cowie 1981, 1994; Mel’čuk 1981, 1988).

Thirdly, the last approach is known as mixed approach (Moreno 2009) since it combines the two approaches previously mentioned. The authors sharing this view (Greenbaum 1970, 1975; Kjellmer 1994; Carter 1998; Gyllstad 2007) take into account the frequency of occurrence of a collocation as well as the linguistic “environment” in which they occur (Carter 1998: 51). For example, Howarth (1996: 29) states: “(...) collocation is not purely a matter of probability of linear co-occurrence, since there are grammatical and semantic or purely lexical factors that constrain lexical co-occurrence in a large number of cases.” Therefore the mixed approach might be considered as a way to remedy the shortcomings of the statistical and linguistic approaches.

### 1.2. General characteristics of the collocations.

As we can see, each of the approaches described in the previous section puts more emphasis on some of the properties of collocations than on others, but even so several common characteristics can be put forward. First of all, the frequency of occurrence is the most important feature since collocations are considered this way because the linguistic elements that constitute them repeatedly co-occur. If it was not so, they would simply be “free combinations.” Besides, the more we use collocations, the quicker they become “institutionalized” (Corpas 1996: 21; Seretan 2011: 15). According to Howarth (1996: 36), “(...) institutionalization is the feature that first makes a collocation noticeable to the analyst before other criteria are applied.” Nevertheless, Copas (1996: 22) justifies the institutionalization of repeated combinations of words by arguing that “(...) los hablantes (...) no van creando sus propias combinaciones originales de palabras al hablar, sino que utilizan combinaciones ya creadas y reproducidas repetidamente en el discurso” [speakers do not create their own original combinations of words while speaking, but they use combinations which have already been created and are repeatedly reproduced in speech, translation mine].

Secondly, most collocations consist of two words, but in theory there is no length limitation. As Sinclair (1991: 170) points out: “In most of the examples, collocation patterns are restricted to pairs of words, but there is no theoretical restriction to the number of words involved.” On the other hand, the component units of a collocation do not need to appear one immediately after the other, i.e. the constituent elements of a collocation might appear within a span of words. Sinclair argues that this span should be at most of four words: “The usual measure of proximity [between the words within a collocation] is a maximum of four words intervening” (1991: 170).

Finally, arbitrariness seems to be another significant feature as these units are apparently created according to “arbitrary word usages” (Smadja 1993: 1), so their lexical items cannot be predicted because “the affinity of a word for a particular collocate (...) is unpredictable” (Seretan 2011: 16). Nevertheless, once they are used and conventionalized, alternative options are blocked, even if they were substituted by words considered synonyms out of context. That is, there are some combination restrictions as the presence of one element “often implies or suggests the rest of [the elements of] the collocation” (Smadja 1993: 147) given that “[their] constituent elements are, to varying degree, mutually selective” (Cruse 1986: 40).

Taking these characteristics into account, a collocation can be defined as follows: “a recurrent, arbitrary and unpredictable combination of two or

more linguistic elements which becomes institutionalized due to frequent usage.”

### 1.3. Syntactic types of collocations.

Following some of the most notable authors (Benson et al. 1986, Hausmann 1989, Corpas 1996, Hill 2000, Koike 2001), collocations can be classified into seven different syntactic types:

- (1) adjective + noun: bright idea, close friend, indoor game.
- (2) verb + noun / noun + verb: accumulate knowledge, break a rule, a bomb explodes.
- (3) adverb + adjective: deeply moved, fully aware, perfectly healthy.
- (4) verb + adverb / adverb + verb: act accordingly, happen spontaneously, hardly expect.
- (5) verb + adjective: sound strange, feel tired, stay awake.
- (6) noun + of + noun: attack of hiccups, bar of chocolate, twist of fate.
- (7) verb + prepositional phrase: *put into practice, bear in mind, be under investigation.*

## 2. The importance of mastering collocations for English academic writing.

As we have already said, L2L should learn collocations in order to sound like native speakers since these word combinations cannot be learned on any basis, but naturally co-occur. In Pawley and Syder's opinion, if a language learner is to achieve nativelike control, they need to distinguish “those usages that are normal or unmarked from those that are unnatural or highly marked” (1983: 194). Apart from this, as English is considered a *lingua franca* nowadays (also known as “international language”, “global language” or “world language”) (Seidlhofer 2005: 339), a good command of these linguistic units is essential if one aims at speaking and writing using this language in a fluent and natural-sounding way.

English is the most widely spoken language around the world with about 400 million native speakers and 1.4 billion who use it as a second or third language (Crystal 2005). Consequently, English has become the language of science (Gómez 2010: 289; Montgomery 2009: 13; Pérez 2001: 1), i.e. most prestigious international scientific publications (*Nature, Science, New England Journal of Medicine*) use this language and over 80% of scientific publication is done through English, as is the vast majority of

international scientific meetings, symposia, research programmes and other academic exchanges (Crystal 2005). In this sense, Swales (1990: 99) declares: "(...) there is no doubt that English has become the world's predominant language of research and scholarship."

For these reasons, researchers should aim at publishing in indexed journals in English, as well as for the reasons mentioned below:

(1) *Spread of knowledge and knowledge display*. As Ishiyama (2010: 12) suggests, publishing is essential in order to communicate research results. Also, it helps authors to demonstrate their knowledge on a topic and the quality of their research, so they would gain intellectual satisfaction from external recognition which would contribute to authors' self-esteem and encourage them to continue doing research and publishing.

(2) *Visibility and leadership*. Publications can prevent authors from being anonymous. In this regard, Ishiyama (2010: 2) claims: "Lo que se hace o se piensa, si no se publica ni se difunde, no existe" [What is done or thought, if it is not published or spread, it will not exist, translation mine]. Likewise, authors conducting groundbreaking research into a particular topic may become a leader in this field for other researchers.

(3) *Contribution to the scientific community, humanity and their institution*. One of the main characteristics of published research articles is the fact that they supply their field of study with something new. In this sense, not only do publications contribute favourably to the subjects they deal with, but also to the knowledge of their readers. Similarly, they help the universities where authors come from be known worldwide.

### 3. Some notions on collocation learning.

As we have already mentioned, the teaching of collocations has been peripheral in the teaching of second languages, hence it still remains unclear how they should be taught. In addition to learning collocational units as chunks of language, some linguists claim that teaching should focus on common collocations rather than unusual or advanced collocations, "even though the latter might seem more attractive at first sight" (Nesselhauf 2005: 259). In other words, teachers should concentrate on expanding knowledge of what is only half-known by teaching students new collocates of a known word, so teaching new and rare words should not be prioritized. Another author who supports this viewpoint is Lewis who remarks: "(...) time spent on half-known language is more likely to encourage input to become intake than time spent on completely new input" (2000: 24). Similarly, Hill promotes the idea that L2L should be taught collocations of familiar words by claiming: "(...) the main thrust of vocabulary work in most classes should be to make

students more collocationally competent with the words with which they are already partly familiar” (2000: 67).

On the other hand, Hatch and Brown (1995) and Nation (2001) propose some vocabulary learning strategies which can be applied to the learning of collocations as they are also considered to be vocabulary. For instance, Hatch and Brown (1995, based on Payne 1988) suggest the following five steps in the process of second language vocabulary acquisition (1995: 374): (1) encountering new words; (2) getting the word form; (3) getting the word meaning; (4) consolidating word form and meaning in memory, and (5) using the word. That is, the first step to acquire vocabulary is encountering new words which can be provided from different sources, e.g., books, television, the Internet, magazines, music. Then, L2L must connect the new word form to its meaning so that they can store this word in their brain. Ultimately, they must start to use this word in their everyday language in order not to forget it. These steps are also present in the process of collocation learning, i.e. L2L (1) will encounter the collocation maybe for the first time, they (2) will get the idea that the word they must use is an only unit whose meaning results from the combination of the meanings of their components (3), and they (4) will have to memorize the collocation as a chunk of language which should be repeatedly used (5).

On the other hand, Nation (2001: 218-222) proposes a more complex taxonomy of kinds of vocabulary learning strategies so that L2L could learn new vocabulary autonomously. He suggests three general types consisting of several sub-types. Firstly, L2L have to decide on what vocabulary they are going to focus on depending on the goals they want to achieve (for instance, if they are dealing with tourism, they will have to choose vocabulary related to this field). After this, they must be aware of the meanings and uses of the chosen words, and use these words so that they do not forget them. Secondly, in order to get information about the target words, L2L should take into account the context where the word appears since it is crucial when unveiling the meaning of a word or a collocation as the same word can acquire different meanings depending on the context of use (Firth 1957; Halliday 1961; Sinclair 1996, 2004). They might also try to find parallel patterns or items in the first or second language. Lastly, some strategies to remember vocabulary are: (i) to notice that a word is an item to be learned, (ii) to recall previously met items and establish a connection between them and the target words, and (iii) to generate sentences containing the word and use the word in different contexts across the four skills, i.e. listening, speaking, reading, and writing.

## Conclusions

Regarding the study of collocations, we can consider three main approaches: the first one which emphasizes their frequency of occurrence, the second one which highlights their linguistic status, and the last one which is a mixture of both. This way, although there is not an agreement on what a collocation is, most authors generally agree that it is an unpredictable combination of two or more lexical units which frequently appear together and is restricted by grammatical or semantic constraints.

Their teaching has traditionally been neglected in the second language teaching, in spite of the fact that a good command of them is extremely important for natural sounding English. L2L have still problems to learn them since they tend to memorize their constituent elements as single units instead of as chunks of language, such as most linguists suggest. Collocational knowledge is also of paramount importance for academics when writing research articles as it helps them sound like native speakers. On the other hand, English is considered the language of science nowadays and most prestigious international publications are done through it. For this reason, academic writers should aim at publishing in this language as it is the best way to reach as much audience as possible.

Finally, some tips on how to learn collocations have been provided. Students should learn them as chunks of language instead of memorizing their constituent elements as isolated words. Besides, the context in which they appear should be taken into account when unveiling their meaning as they can acquire different meanings depending on the linguistic environment around them. Once these combinations are learned, they should be used repeatedly across the four skills, i.e. written and oral production and comprehension, to prevent the non-native speaker from forgetting them.

## References

- BENSON, M., BENSON, E., ILSON, R. (1986): *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- CARTER, R. (1998): *Vocabulary*. London: Routledge.
- CORPAS, G. (1996): *Manual de Fraseología Española*. Madrid: Gredos, 1996.
- COWIE, A. P. (1981): "The Treatment of Collocations and Idioms in Learners' Dictionaries." In: *Applied Linguistics*, 2: 3, 223-235.
- \_\_\_\_\_ (1994): "Phraseology." In: *The Encyclopedia of Language and Linguistics*. R. E. Asher (ed.). Oxford: Pergamon, 3168-3171.
- CRUSE, D. A. (1986): *Lexical Semantics*. Cambridge: Cambridge University Press.



- CRYSTAL, D. (2005): *English as a Global Language*. Cambridge, New York: Cambridge University Press.
- ELLIS, N. (2001): "Memory for Language." In: *Cognition and Second Language Instruction*. P. Robinson (ed.). Cambridge: Cambridge University Press, 33-68.
- FIRTH, J. R. (1957): *Papers in Linguistics, 1934-1951*. London: Oxford University Press.
- GÓMEZ-GARCÍA, F. (2010): "The Importance of Publishing in English for Our Journal." In: *Acta Ortopédica Mexicana*, 24: 5, 289-290.
- GREENBAUM, S. (1970): *Verb-Intensifier Collocations in English: an Experimental Approach*. The Hague, Paris: Mouton.
- \_\_\_\_\_ (1975): "Some Verb-Intensifier Collocations in American and British English." In: *American Speech*, 49: 1-2, 79-89.
- GYLLSTAD, H. (2007): *Testing English collocations. Developing receptive tests for use with advanced Swedish learners*. PhD Thesis. Lund University, Sweden.
- HALLIDAY, M. A. K. (1961): "Categories of the Theory of Grammar." In: *Word*, 17, 241-292.
- HATCH, E.; BROWN, Ch. (1995): *Vocabulary, Semantics, and Language Education*. Cambridge: Cambridge University Press.
- HAUSMANN, F. J. (1989): *Wörterbücher – Dictionaries – Dictionnaires: Ein internationales Handbuch zur Lexikographie Volume I*. Berlin, New York: Walter de Gruyter.
- HILL, J. (2000): "Revising Priorities: from Grammatical Failure to Collocational Success." In: *Teaching Collocation*. M. Lewis (ed.). London: Language Teaching Publications, 47-70.
- HOWARTH, P. A. (1996): *Phraseology in English Academic Writing*. Tübingen: Niemeyer.
- ISHIYAMA, R. (2011): "El Artículo Científico y la Importancia de su Publicación." *Upch.edu.pe*. Universidad Peruana Cayetano Heredia [En línea].
- KJELLMER, G. (1987): "Aspects of English Collocations." In: W. Meijs (ed.). *Corpus Linguistics and Beyond: Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora, Volumen 59*. Amsterdam: Rodopi, 133-140.
- \_\_\_\_\_ (1994): *A Dictionary of English Collocations: Based on the Brown Corpus*. Oxford: Clarendon Press.
- KOIKE, K. (2001): *Colocaciones Léxicas en el Español Actual: Estudio Formal y Léxico-Semántico*. Alcalá de Henares: Universidad de Alcalá de Henares.

- LEWIS, M. (2000): "There Is Nothing as Practical as a Good Theory." In: *Teaching Collocation*. M. Lewis (ed.). London: Heinle Cengage Learning, 10-27.
- MCCARTHY, M. (1990): *Vocabulary*. Oxford, New York: Oxford University Press.
- MCCARTHY, M.; O'DELL, F. (2008): *English Collocations in Use: How Words Work Together for Fluent and Natural English: Self-Study and Classroom Use*. Cambridge: University of Cambridge.
- MEL'ČUK, I. A. (1981): "Meaning-Text Models: a Recent Trend in Soviet Linguistics." In: *Annual Review of Anthropology*, 10, 27-62.
- MEL'ČUK, Igor Aleksandrovič, *Dependency Syntax: Theory and Practice*. Albany, New York: SUNY Press, 1988.
- MILLER, G. A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." In: *Psychological Review*, 63, 81-97.
- MONTGOMERY, S. L. (2009): "English and Science: Realities and Issues for Translation in the Age of an Expanding *Lingua Franca*." In: *The Journal of Specialized Translation*, 11, 6-16.
- MORENO, M. (2009): *Recopilación, Desarrollo Pedagógico y Evaluación de un Banco de Colocaciones Frecuentes de la Lengua Inglesa a través de la Lingüística de Corpus y Computacional*. PhD Thesis. University of Granada, Spain.
- NATION, I. S. P. (2001): *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- NESSLHAUF, N. (2005): *Collocations in a Learner Corpus*. Philadelphia: John Benjamins.
- PALMER, H. E. (1933): *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- PAWLEY, A.; SYDER, F. H. (1983): "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." In: *Language and Communication*. J. C. Richards, R. W. Schmidt (eds.). London: Longman, 191-226.
- PAYNE, M. E. (1988): *Vocabulary Learning Strategies Used by ESL Students and their Relationship to Perceptual Learning Style Preferences*. PhD thesis. Brigham Young University, Provo, Utah.
- PÉREZ, R. (2001): "El Español de la Ciencia. Importancia de la Lengua Española en el Desarrollo de la Ciencia en México." In: *Centro Virtual Cervantes: Congresos Internacionales de la Lengua*. Iberia, Telefónica, Caja Duero.
- SEIDLHOFER, B. (2005): "English as a *Lingua Franca*." In: *ELT Journal*, 59: 4, 339-341.
- SERETAN, V. (2011): *Syntax-Based Collocation Extraction*. Geneva: Springer.

- SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- \_\_\_\_\_ (1996): "The Search for the Units of Meaning." In: *Textus*, 9: 1, 75-106.
- \_\_\_\_\_ (2004): "The Search for the Units of Meaning." In: *Trust the Text: Language, Corpus and Discourse*. J. Sinclair (ed.). London: Routledge, 24-48.
- SMADJA, F. A. (1993): "Retrieving Collocations from Text: Xtract." In: *Computational Linguistics*, 19: 1, 143-177.
- SWALES, J. M. (1990): *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

