

## Introduction

While initially understood as a type of cognitive restriction (Firth 1957), very much in line with Coseriu's *lexical solidarities* (1967), the term *collocation* is now most often used as a kind of distributional restriction. This notion has had two different developments. The first is the phraseological one, that can be found in frameworks such as Meaning-Text Theory, in relation to the notion of *lexical functions* (Mel'cuk/Clas/Polguère 1995; Mel'cuk 2003; Mel'cuk/Polguère 2006), and in the idea of a binary relation between two lexical components, the *base* and the *collocate*, where the collocate fully realizes its meaning only when coupled with its base (see Paillard [1997], Hausmann [1998], Grossmann/Tutin [2002, 2003]). The second approach, originated in the works of the late John Sinclair in Great Britain, is strongly grounded in statistics and corpus analysis. The emphasis here is on the frequency of co-occurrences of word pairs, and on the distribution of meanings and lexical uses of words. The former approach led to extensive researches in lexicology (Cruse, 1988) and lexicography (Hausmann 1989, Mel'čuk 1998). The latter (Sinclair 1991, Evert 2008) underlies research in corpus linguistics.

In this volume, we take a lexicographic perspective. The aim of this volume is to promote a discussion on the definition of collocations that can be useful to lexicographic purposes. Problems with the definition of collocations are related, first, to the boundaries between collocations and free combinations, and, second, to those between collocations and idioms. At this level, the question to be answered is: do lexicographers need to take these distinctions into consideration? If we analyse the boundaries existing between collocations and free combinations, for instance, what is under investigation is the very notion of linguistic restriction. Given the large amount of space made available by the digital medium, in the near future electronic dictionaries might make this distinction redundant and enable lexicographers to include all

combinatory possibilities of a word, even for lexical items that are not part of a collocation. This does not mean, however, that a classification of multi-word combinations, and, consequently, criteria to define collocations are superfluous to lexicographic needs. As Bergenholtz and Gouws (2013: 11) point out, what is under question is not the necessity of a classification system, which is actually becoming more and more crucial for lexicography, but the way in which lexicographers convey such a classification to dictionary users.

A second issue that has to be taken into consideration when focusing on the notion of collocation is the possibility, or even necessity, of adapting the definition of collocation to the ends of different types of dictionaries (Tarp 2008). The structure of a given dictionary, i.e. the sum of its micro- and macro-structural properties, results from and reflects the objectives of the lexicographer, who is aiming to serve specific user's needs in specific usage situations (e.g. text production or active translation). The lexicographic formalization of the concept of collocation should take all this into account. The central issue is thus how to fit a theoretical model into the practice of lexicography. As suggested by Rundell (2012: 71), while lexicographers require a theory, the particular purpose of a dictionary precludes its uncritical application. However, the question could also be turned around by asking: is there any theoretical model for the description of collocations that is directly applicable to dictionary making?

One further issue that should be addressed concerns the methods for collocation extraction. As a matter of fact, computational linguistics operationalizes morphosyntactic and statistical criteria (Heid/Weller 2010, Evert 2005 respectively), however these criteria are not able to draw a clear distinction between collocations and other word combinations, since they rely on an empirical definition of collocation (Evert 2008) which often fails to capture cognitive and semantic distinctions. In order to tailor the definition of collocation to the actual dictionary function it is therefore necessary to develop hybrid methods combining data processing with other criteria such as native speakers' evaluation and contrastive analysis.

The primary aim of this volume is to reflect upon the relation between lexicographical practice and theorization of the notion of

collocation, and stimulate discussion of issues relevant for future research in the field of lexicography. Each of the papers in this volume addresses in detail one or more aspects of the above mentioned issues. The book opens with a preliminary chapter by ADRIANA ORLANDI, in which the author outlines the state of art of research in monolingual collocation lexicography. The aim is to reflect upon the definition of collocation in lexicography and to point out the domains where lexicography needs further improvements in order to compile collocation dictionaries really adequate to the users' needs. The paper shows that a functional definition of collocation does not necessarily contradict the theoretical views upon collocations, and can be useful to determine the role of colligations in a collocation dictionary, as well as the place of free combinations and idioms. The paper also tries to discuss the possibility of envisaging a prototypical approach to collocations, emphasizing some paradoxical aspects that characterize the search for a prototype. Finally, the role of electronic lexicography is emphasized as a way to improve features which miss or lack accuracy in nowadays lexicography, such as information about frequency and fixedness of collocations, authentic examples, direct access to corpora, and usage notes.

After this overview of collocation lexicography, the volume investigates some general questions very often underestimated by lexicographers, but rich in implications for collocation lexicography and overall lexicography: the nature of distributional restrictions underneath word combinations, and boundaries between collocations, free combinations and idioms. The first issue is grafted on a typology of syntactic and conceptual restrictions, and it has been treated in the present volume mostly in relation to the boundary between lexical combinations and free combinations. According to VINCENZO LO CASCIO, all word combinations, included free combinations, are regulated by some congruency principles, so that the study of word combinations should be based upon the knowledge of these principles. Lo Cascio distinguishes between formal congruency principles (syntactic and functional syntactic) and encyclopaedic-semantic ones. There are congruency principles that are not language-bound, and this is the case of free combinations, that only satisfy requirements of the general congruency principles. When congruency principles are specific and cultural-bound, they generate

idiosyncratic word combinations, to which collocations belong. Thus, according to the author, it is only within a comparison between languages that we can speak of collocations. Contrastive analysis becomes the central criterion of distinction between collocations and free combinations. The main consequence for lexicography is that the lexicographic description of the lexicon should concern the entire range of combinations allowed by a word, and that all the properties which determine the combinatorial preferences of a word should be described. Finally, Lo Cascio introduces the online version of his collocation dictionaries with the help of selected excerpts.

MICHELE PRANDI's paper focusses more deeply on Lo Cascio's notion of "encyclopaedic-semantic congruency principles". The very notion of content-based restrictions is investigated, and three different types of restrictions are taken into account: selection restrictions (confining for instance death to living beings), lexical solidarities (restricting barking to dogs) and cognitive models (restricting flying to birds). Unlike lexical solidarities, that are language-specific, selection restrictions, which correspond to consistency criteria and belong to a natural ontology, are not language-specific but universal. Cognitive models are conceptual structures shared on a very large scale, which admit the possibility of being falsified (birds that don't fly). Consistency criteria are never stated in dictionaries, but if the aim of a dictionary is to account for the distribution of lexemes within sentence structures, and not simply to describe the content of isolated words, consistency criteria should be taken into account. Prandi proposes Gaston Gross' model of "generative lexicon" as a model that makes consistency criteria as well as lexical solidarities and cognitive models explicit.

An interesting point that differentiates Lo Cascio's approach from Prandi's is their position vis-à-vis the relationship between syntax and lexicon. Whereas Lo Cascio considers that the lexical component has "a primary role above syntax", in Prandi's view "syntax goes far beyond lexicon".

The paper by BÉATRICE LAMIROY focusses on the distinction between collocations and idioms. Starting from the description of a research project called "BFQS project" (an enquiry on idioms in the francophone area that takes as its starting point Maurice Gross' corpus

of French idioms), Lamiroy raises the problem of settling a protocol enabling researchers and lexicographers to easily recognize idioms and distinguish them from collocations. Two main problems are recalled: the multifaceted nature of *expressions figées*, and differences in the continuum of lexicalization due to their diachronic dimension. A detailed description of similarities and differences between idioms and collocations is then provided, giving the reader a comprehensive view of the subject.

Lo Cascio, Prandi and Lamiroy's papers raise an issue that will require further investigation in the future, that is the problem of the relationship between different types of word combinations and encoding/decoding tasks. Just to make one example, selection restrictions and cognitive models described by Michele Prandi enable speakers of different cultures to decode collocations not based on figures of speech quite easily, while these restrictions are not sufficient to correct encoding (see for instance it. *aspettare per un bel pezzo* which can be translated into English using the expression *to wait for a fair amount of time* and not *to wait for a nice piece of time*). On the other hand, collocations based on figures of speech represent to a non-native speaker, especially when they are not transparent, a difficulty not only in encoding but also in decoding tasks. Thus, the French collocation *peur bleue* (great fear) is quite difficult to decode for a non-native speaker if the context does not support interpretation. These considerations hint at the possibility of a different treatment for different types of collocations in collocation dictionaries, for instance providing more explanations and examples for opaque collocations.

The second part of the volume investigates more deeply some aspects of the definition of collocation in lexicography. The aim of DANIELA CAPRA is to show the inherent instability of the definition of collocation. Her observations focus on the domain of Spanish linguistics, but the remarks delivered are generalizable. The issues concern discrepancy whether collocations are part of phraseology or not, fixity and the compositional character of collocations, frequency, and finally the semantic determination in the combinatory of collocations. She takes as an example the treatment of Spanish light verbs, and of some *Noun + Prep. + Noun* combinations. Special attention is given

to Bosque's combinatory dictionary REDES (2004), where the choice is made to avoid the term *collocation*. The conclusion of the paper invites linguists to consider the instability of the concept of collocation as "part of its nature", and encourages the application of the Prototype Theory to the description of this complex and multi-layered category of word combinations.

GLORIA CORPAS PASTOR analyses collocation dictionaries for English and Spanish classifying them on the basis of the importance given to corpora, and from the viewpoint of their theoretical and methodological underpinnings. Standard dictionaries of collocations are based on the lexicographer's intuition and do not take into account important information such as frequency or evidence of usage. Dictionaries of collocations rely on statistical and frequency-based theories of collocation make use of corpora as sources of information, but here again the way in which corpora are used can largely differ from one dictionary to another, with some dictionaries being corpus-based, and other corpus-driven (Tognini-Bonelli 1991). The paper provides a thorough classification of English and Spanish collocation dictionaries, offering an overview of the underlying approaches to collocation and its definitorial properties.

LAURA GIACOMINI's paper focusses on the definition of collocation for lexicographic purposes. According to the author, a well-designed concept of collocation is fundamental to have criteria for data selection, but it is necessary for a lexicographer to base theoretical considerations upon the user's needs. Starting from her own experience as a dictionary compiler, Giacomini discusses the advantages of a "functional definition" of collocation as it has been employed for modelling an electronic dictionary of Italian collocations concerning the semantic field of fear. This working definition includes on the one hand a phraseological concept of collocation including idiomatic expressions, and on the other hand a wider concept of collocation as a combination with a high degree of familiarity in the speaker's mental lexicon. The use of the electronic medium makes it possible for each headword to be very accurate at the microstructural and mediostructural level, exploiting both formal and conceptual parameters to the description of collocational meanings and their syntactic patterns.

VERONICA BENIGNO and OLIVIER KRAIF discuss the concept of ‘core vocabulary’ and ‘core collocations’ and its implications for the treatment of collocations in monolingual learner phraseological dictionaries. They present the findings from a corpus-based study combining statistical analysis and native speakers’ evaluation in order to isolate the features that can be used to filter out core collocations from a set of potential candidates identified from a given pivot. The study shows that statistical measures such as frequency are appropriate but not sufficient to identify core collocations in language, because native speakers show to assign more value to highly restricted and fixed units regardless of their frequency of occurrence. These findings are directly connected with the third part of the paper, which deals with phraseology from the pedagogical and lexicographical perspective of collocations’ learner dictionaries. This section argues that both frequency and usefulness should be considered as main organizing principles for this kind of dictionaries. Practical examples extracted from the *Longman Dictionary of Contemporary English* – LDOCE – 5<sup>th</sup> edition illustrate this point.

FRANCIS GROSSMANN and AGNÈS TUTIN analyse Noun Prep Noun constructions, which represent a challenge for the study and classification of collocations both in the field of general and specialised discourse. They choose to focus on cross-disciplinary scientific lexicon using a large corpus of scientific papers, and they analyse collocations candidates according to a list of parameters that includes the semantic characterization of N1 and N2, the role played by the preposition, the presence or absence of determiners behind the N2, and so on. They address a typology of five Noun Prep Noun constructions: a) objective genitive constructions, b) subjective genitive constructions, c) predicative structures, d) specification structures, and e) classification structures. The study shows that among these types of constructions, two appear to be more directly linked to the emergence of collocations: predicative and specification structures (*hypothèse de départ, recherche de terrain*). The authors thus validate an approach that can help linguists and lexicographers to evaluate the collocational status of a construction in scientific lexicon. This approach is based on various criteria (syntactic, semantic, and sometimes pragmatic), in addition to statistical measures.

The final chapter of the volume is by LUIGI MATT. Four Italian dictionaries of collocations recently published (Urzi 2009; Russo 2010; Tiberii 2012; Lo Cascio 2012, 2013) are illustrated in detail. The analysis focusses on all aspects of dictionary compiling, going from titles and target readership, to theoretical aspects, choice of headwords and collocations, treatment of collocations, definitions, examples and diaphasic markers. Matt shows that these dictionaries represent a first step taken by Italian linguists to fill the lexicographic gap between Italian and other major European languages. The most relevant features of each dictionary, including its main advantages and drawbacks, are described in the concluding chapter.

## References

- Bergenholtz, Henning / Gows, Rufus 2013. A lexicographical perspective on the classification of multiword combinations. *International Journal of Lexicography*. 27/1, 1–24.
- Coseriu, Eugenio 1967. Lexikalische Solidaritäten. *Poetica*, 1, 293–303.
- Cruse, Alan 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Evert, Stefan 2005. *The Statistics of Word Cooccurrences – Word Pairs and Collocations*. Stuttgart: University of Stuttgart, IMS. <<http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/pdf/Evert2005phd.pdf>>.
- Evert, Stefan 2008. Corpora and collocations. In Lüdeling, Anke / Kytö, Merja (eds) *Corpus Linguistics. An International Handbook*. Vol. 2. De Gruyter, 1212–1248.
- Firth, John Rupert 1957. *Papers in Linguistics 1934/1951*. Oxford: Oxford University Press.
- Grossmann, Francis / Tutin, Agnès (eds) 2003. *Les collocations. Analyse et traitement*. Travaux et recherches en linguistique appliquée. Amsterdam: De Werelt.
- Grossmann, Francis / Tutin, Agnès 2002. Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif. *Revue*