

Laura Pinnavaia*

*Collocations in Twenty-first Century English Monolingual
Lexicography: a State of the Art*

«Compared to teachers of other languages, teachers of English are in a rather privileged position: English is the Lingua Franca of our times. This unique status of a vernacular for which there is no parallel in the history of languages has led to a thorough and in-depth study of the English language, with the result that it is one of the best described modern languages.» (Stein, 2002: 101)

ABSTRACT:

Many studies in the field of lexicography have been devoted to collocations in order to analyse their coverage and position in different English monolingual dictionaries. Indeed, given that it is learners who have the greatest difficulty in understanding and in using collocations, researchers have been examining language users' ability to find and select collocations in such dictionaries. The outcome of these researches has prompted a series of studies that stress the need for a more suitable dictionary for collocations. This essay will be devoted to providing a synoptic report about studies that feature the collocation, the user, and the dictionary.

KEYWORDS: Collocations, Dictionary, English, Lexicography, Studies, Users

1. *The term* collocation

That words are combined in recurring patterns in order to create discourse is a relatively recent scholarly realization, testified to by the fact that the definition of the term *collocation* as “the habitual juxtaposition or association, in the sentences of a language, of a particular word with other particular words; a group of words so associated [...]” appears in the *Oxford English Dictionary* no sooner than 1951, on the input of J.R. Firth studies in Linguistics¹. Firth insisted that all linguistic utterances were in different ways meaningful activity, and that phonology, morphology,

* Università di Milano.

¹ “1951 J.R. Firth in *Ess. & Stud.* IV. 123 I propose to bring forward as a technical term, meaning by ‘collocation’, and to apply the test of ‘collocability’” (OED s.v. *collocation*).

and syntax, along with lexicon, should be brought within an extended theory of semantics. His basic concept in this approach was the context of situation and the insistence on studying language in its social and cultural context, contributing to fostering the ‘Linguistic War’ of the 1970s, during which linguists became interested in the social uses of language (see *Oxford National Dictionary of Biography* s.v. J.R Firth). Firth (1957: 14) argued that collocations are “actual words in habitual company” and that “meaning by collocation” is one of the modes in which meaning can be expressed because the meaning of any word is generated by the company it keeps.

But it was not Firth who first stated that words tend to occur in fixed and predictable combinations. It was Harold Palmer who, involved in English language teaching and concerned with second language acquisition, first described the term *collocation* as “a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts” (Palmer, 1933: i). Unlike Firth, however, Palmer failed to expand on this idea, to provide a detailed description of it, and to integrate it into a sound theoretical framework, which is why traditionally it is Firth who is considered the father of collocations. And yet Palmer’s impact upon the development of linguistic studies was no smaller than Firth’s: indeed, he influenced “phraseological research throughout Europe, thereby creating a phraseological tradition of collocation studies concerned by fixed or semi-fixed units.” (Williams & Millon, 2011: 3). Though proceeding along parallel lines, both scholars’ works were equally important for the evolution of the concept of *collocation*, which has now come to refer to two slightly different types of word combinations, depending on whether it was influenced by Palmer or by Firth’s studies. Indeed, if Palmer’s *Second Interim Report on Collocations* (1933) gave rise to the continental tradition of restricted collocation in phraseology, from Firth’s studies (1957) came the other major approach based on a statistical method, developed within the Birmingham school of corpus linguistics.

2. *The development of the two concepts of collocation*

2.1 *The phraseological collocation*

Palmer's studies were of great inspiration to the traditions of phraseological studies first in the former Soviet Union and Eastern Europe and later to those in Western Europe and North America. Thanks to the work carried out by many scholars, among whom stand out the names of Vinogradov, Amosova, Hausmann, Cowie, Mel'čuk, and Burger, phraseology is now the linguistic discipline we all recognize as being devoted to categorizing and analyzing phraseological units. However, because scholars over the decades have prioritized different features in the classification of phraseological units, different models and typologies still strongly characterize the discipline and it is often difficult to find in every scholar's work the same stable definition for each subcategory. That said, all phraseologists seem to concur that language is a continuum that extends from free to non-free combinations of words.

In the attempt to describe the features that characterize a collocation, we will begin by reporting its position along the free/non-free continuum and its difference from other lexical units, commencing with free combination of words, which see "two elements that do not repeatedly co-occur, that are not bound specifically to each other, that occur with other lexical items freely". Benson *et al.* (1986: xxx) illustrate this definition with the word *murder* that may be combined in accordance with the general rules of English syntax with hundreds of words, such as *abhor*, *accept*, *acclaim*, *advocate*, and freely substituted with each (1986: xv). This is not at all possible for a collocation. *Commit murder* is an example of this because "the verb *commit* is limited in use to a small number of nouns, meaning 'crime', 'wrongdoing' and collocates specifically with *murder* and cannot be replaced by any other word (1986: xxx). Thus, collocations consist of one word that is freely chosen on the basis of the speaker's communicative intent (e.g. to report a *murder*), while the other is an obliged choice because dependent upon the first (e.g. *commit*). These two components are referred to as *base* and *collocator* (see Hausmann, 1984) or *keyword* and *value* (see Mel'čuk, 1998). At the other end of the continuum, collocations can be differentiated from word combinations that are considered as being not free at all. Indeed, despite the varying terminology, Cowie (1988, 1994), Burger (1998) and Mel'čuk (1998), among others, all agree that *collocations* differ from

partial/figurative/quasi-idioms because the latter have both a figurative meaning and a literal interpretation (e.g. *do a U-turn*), and differ even further from *idioms/pure idioms/full phrasemes*, which have only and exclusively a figurative meaning (e.g. *spill the beans*).

Whilst the divisions along the language continuum may seem neat and clear, much less exact are the descriptions of the subcategories. Especially collocations, positioned in between free combinations and idioms, can be seen to host a range of semantic combinations. At one extreme, one may find a collocation that includes words all having a literal meaning, like *to make a comment*; at the other extreme, one may find a collocation like *heavy rain*, made up of a base *rain*, which has a literal meaning and a collocator *heavy* with a figurative meaning. In varying their semantic makeup, collocations can thus range from being completely literal to partially figurative, known as opaque. Regardless of their level of semantic compositionality, such combinations – in which the collocator and the base are syntactically and semantically bound and cannot be substituted – are referred to as restricted collocations. This term relegates them to the sphere of phraseology, and distinguishes them from the other concept of collocation that stems from the sphere of corpus linguistics.

2.2. *The corpus linguistic collocation*

From the Firthian idea that words should be known by the company they keep comes the other concept of collocation, borne from linguistic studies on corpora by John Sinclair and his followers. In Sinclair's pioneering work *Corpus Concordance Collocation* (1991:170), this term takes on a completely different meaning from the original phraseological one:

«A collocation is the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening. Collocations can be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated.»

Unlike phraseologists, corpus linguists look for frequently recurring groups of words, they also call collocations. Unlike the phraseological ones, however, these multiword units can extend beyond two or three

words, becoming longer lexical chunks. What they aim to show is that instead of the open choice principle, whereby language users have at their disposal the choice of an unlimited number of single terms, communication comes about via “a large number of semi-preconstructed phrases that constitute single choices”, referred to as the idiom principle (Sinclair, 1991: 110). If phraseologists are concerned with analyzing the syntactic makeup and the degree of semantic compositionality between the components of collocations, thus distinguishing bases/keywords from collocators/values, corpus linguists are concerned with language users’ intentions within the context of the situation. Consequently, any component of the collocation may be either *node* or *collocate* depending on the speaker’s communicative purpose within the discourse. For instance, in corpus linguistic terms, the words *heavy* and *rain* can both be nodes, depending on whether the focus of the communication is on the adjective *heavy* or on the noun *rain*, with *rain* and *heavy* being possible but not exclusive collocates. In this approach, words are not connected to one or two collocates only, but rather to sets of collocates. The connection between the components in the collocation surpasses the limits of the semantic field, typical of the phraseologists’ approach, embracing all the contexts and co-texts that the components encounter. It is the principle upon which Hoey’s (2005) theory of lexical priming rests, in which every time a word is encountered it is associated with other words, which in turn are associated with other words. There is no doubt that in the corpus linguistic approach collocations are a dynamic process in which the lexical environment plays a key role, as opposed to collocations that in the phraseological approach are products to be studied and listed in reference works (Williams & Millon, 2011).

2.3 The ‘hybrid’ collocation

And it is indeed in Hornby’s *Advanced Learner’s Dictionary of Current English* (1942) that we find the first fruits that Palmer’s studies bore on the history of English lexicography: it was the first dictionary to include a great deal of collocational and phraseological information², opening up a brand new tradition in English monolingual learner lexicography that has developed at an exponential rate since then, not only in

² “Early dictionaries were full of chunks of discourse that were glossed, explained, or defined, [...], but collocations then disappeared from most dictionaries after the eighteenth century in England, [...] only to reappear recently.” (Béjoint, 2010: 316).

the production of general monolingual dictionaries for learners, but also in the output of specialized dictionaries devoted solely to collocations. The first English collocations dictionary, of course, is Benson, Benson and Ilson's *BBI Dictionary of English Word Combinations* (1986). In line with the phraseological approach, the authors devote a good number of pages to explaining the differences between lexical and grammatical collocations as well as to providing a detailed syntactic description of the seven lexical structures and the eight grammatical ones. Clearly, the compilation of this dictionary was at the time manual and the selection of the collocations to include prevalently based upon the lexicographers' language instinct. With the advent of corpus linguistics not only did the compilation of dictionaries change from manual to electronic, but so did the inclusion of collocations that were no longer selected according to their syntactic and semantic makeup, but also according to their recurrence in texts.

Emblematic of this change is *The Collins COBUILD Dictionary*, published in 1987, which makes a hard break with former lexicographical tradition, as Sinclair states in its Preface (xv):

« [...] for the first time a dictionary has been compiled by the thorough examination of a representative group of English texts, spoken and written, running to many millions of words. This means that in addition to all the tools of the conventional dictionary makers – wide reading and experience of English, other dictionaries and of course eyes and ears – this dictionary is based on hard, measurable evidence. No major uses are missed, and the number of times a use occurs has a strong influence on the way the entries are organized. »

In becoming one of the most important triggers of inclusion of words in learner lexicography, usage also comes to influence the inclusion and treatment of collocations too, starting from the end of the twentieth century. The inclusion of examples taken from real language use, which is possibly one of the most characterizing features of the new corpus-based dictionaries, endorses the importance that the linguistic environment has for corpus linguists. For Sinclair, in his Firthian-inspired view of language, the context is indispensable for explaining the meaning of words:

«The most important result that has come from the work of preparing this dictionary concerns the way in which patterns of words with each other are related to the meanings and uses of the

words. [...] It is not really possible to talk about the meaning of the word in isolation – it only has a particular meaning when it is in a particular environment. » (Sinclair, 1987: xvii)

Since words have no meaning on their own, collocations rightly take the lion's share in Sinclair's dictionary. He includes them to explain and illustrate the meaning of words, both as examples of use under lemmas and as lemmas themselves. Though only a general learner's dictionary, the Cobuild's coverage of collocations is certainly no less reliable than the specialised BBI's. On the contrary, the precision and the detail that corpus linguistics confers Sinclair's dictionary cannot be wholly matched by Benson et al's work, which is manually compiled. The resultant word patterns are thus ideologically different: if Benson et al's collocations represent a unit that stems from the phraseological tradition, typical of pre-corpus lexicographical experience, Sinclair's collocations represent a cross between the phraseological unit, typical of lexicography, and the statistically-significant one, typical of corpus linguistics, leading to a new concept of collocation for lexicographic purposes in the twenty-first century:

«Lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in direct relation with each other. » (Bartsch, 2004:76)

As we can see from this definition, the semantic relation between components is not put to one side, as often happens when collocations are defined in the domain of corpus linguistics. For the purpose of lexicography, a definition that represents a middle road between the theoretical (phraseological) and the empirical (distributional and frequency-based) (see Evert, 2008: 1213) seems to be a functional compromise that helps to facilitate the extraction of collocations—both lexical and grammatical. That said, the definition of collocation even in the domain of English lexicography is still far from homogeneous and remains a crucial topic of debate, with linguists and lexicographers putting forward what each considers the most appropriate functional or working definition. To enter into this debate would be too complicated and too long for the scope of this essay, but it certainly remains one of the thorniest issues regarding studies of collocations in twenty-first century lexicography of English and other languages too³.

³ For more detailed information regarding this aspect see Orlandi & Giacomini (2016).

3. *Research issues in twenty-first century English lexicography*

Next to questioning what lexicographers should look for in order to extract from corpora the most numerous and most appropriate instances of collocations to be included in dictionaries, scholars' studies regarding collocations in English lexicography have been focussing on another three important issues lately. An important number of studies have been devoted to analysing the coverage and position of collocations in different English monolingual dictionaries, both of the general learner and specialized kind. Aware that it is the language learner who has the greatest difficulty in understanding and in using collocations, researchers have thus also been examining language users' ability to find and select collocations in such dictionaries. The outcome of these researches has prompted an interesting new series of studies that stress the need for a more suitable dictionary if collocations are to be learnt effectively. A synoptic report about studies that feature the collocation, the user, and the dictionary is thus what this essay will be about.

3.1 *The coverage and position of collocations*

That the new 'hybrid' definition of collocation has been fruitful in the extraction of collocations from corpora can be seen by the increased number of collocations included in more recently published specialized dictionaries. Indeed, compared to the second revised edition of the *BBI Combinatory Dictionary of English* (1997)—“considered its main competitor”—Crowther *et al.* (2002: 58) point out that the *Oxford Collocations Dictionary for Students of English* (2002), has a much more extended coverage, owing to the editors' choice to implement a more versatile and operative definition of collocation that combines statistical saliency and learners' needs with phraseological norms. It is possibly because this working definition is so efficient that the *Oxford Collocations Dictionary for Students of English* not only fares better than *The LTP Dictionary of Selected Collocations* (1999), underlines Walker (2009: 288) in a later research between the three dictionaries, but also better than the *Collins COBUILD English Collocations on CD-Rom* (1995) add Crowther *et al.* (2002: 61).

This discrepancy of coverage emerges also in research carried out on the learners' dictionaries. Walker (2009), who carries out contrastive

analyses between the *Collins COBUILD Advanced Learners' Dictionary* (2005), the *Longman Dictionary of Contemporary English* (2003), and the *Oxford Advanced Learners' Dictionary* (2005) finds that, even among the general dictionaries, the *Collins COBUILD Advanced Learners' Dictionary* includes the fewest collocations and the *Oxford Advanced Learners' Dictionary* the most, not only confirming the findings Mittmann (1999) had obtained when examining earlier editions of the same three dictionaries, but also endorsing the findings reported above regarding the specialized dictionaries of collocations. While these differences between the COBUILD and the other dictionaries, both special and general, may be partly due to the more stringent editorial policies adhered to by COBUILD, whose "collocate listings are restricted to maximally twenty items" (Crowther *et al.*, 2002: 61) because what is statistically not salient is left out, there is no doubt that each learners' dictionary proceeds very differently as to what they decide to include. This is also largely testified to by the fact that only 29% of all the collocates listed appear in two or more of the three learners' dictionaries (Walker, 2009: 288), and that more than 80% of collocations in the three specialized dictionaries appear in only one of the three (Walker, 2009: 297).

Research has shown that the lack of agreement in the contents of the general learner and specialized dictionaries does not only regard the number of collocations, but also the way they are entered in dictionaries. This fact has resulted from analyses, pointed at examining where collocations are placed. In claiming that "lexicographers may have to do more than inserting a collocation in an example which illustrates the meaning of the headword", Laufer (2011: 45) in her analysis of the *Cambridge International Dictionary of English* (1995), *Cambridge Advanced Learners' Dictionary* (2005), *Longman Dictionary of Contemporary English* (2003), *Oxford Advanced Learner's Dictionary of Current English* (2005), *Collins COBUILD Advanced Dictionary* (2008) supports the results Walker also obtains regarding the position of collocations in learner dictionaries. Indeed, Walker (2009) shows that most collocations in the entries of the learners' dictionaries he examines seem to be chosen in order to exemplify a definition, or part of a definition. Like many earlier findings, these twenty-first century studies reiterate the need for general learner dictionaries to give collocations more prominence and not include them simply "to exemplify the different polysemous meanings explained in the entry" (Walker, 2009: 287).

Whether collocations are entered as lemmas or as examples of use, it is not of any relevance in the case of specialized dictionaries: in these

works, collocations are all given headword status. What has concerned researchers, instead, is the choice of which lexeme of the multiword unit should figure as the headword. In the examination of the *Macmillan's Collocation Dictionary* (2010), Coffey (2011) states that the headwords are either nouns, adjectives or verbs, which is also true for the *BBJ Combinatory Dictionary of English* (1997), *The LTP Dictionary of Selected Collocations* (1999), and the *Oxford Collocations Dictionary for Students of English* (2002), with “the approximate proportions for each grammatical category [in the former being]: nouns 55%, adjectives 24%, verbs 21%” (2011: 329). That does not mean that the *Macmillan's Collocation Dictionary* lacks innovativeness. On the contrary, compared to the previously published works, it includes three lexico-grammatical patterns in which a verb or adjective base leads to a noun collocate (for example V. DESERVE + N. applause; N. injuries + V. HEAL; ADJ. DESIRABLE + N. attribute) (2011: 334). However, if research has shown that even in more recent publications, nouns still predominate as headwords of collocations, it is because contemporary lexicography has continued to abide by the principles of phraseology when positioning them, regardless of the enormous impact corpus linguistics has had on their extraction and on language study in general. It may be possible to say then that the concept of semantic tension extant between lexemes composing a collocation has resisted the test of time and continues to influence twenty-first century dictionary-making. Following Hausmann's (1984) tenet, nouns – more commonly the autonomous lexemes of the unit – are entered as *bases* to be looked up first in order to find their semantically related *collocates*.

3.2 *Learners' ability to locate and select collocations*

That the noun seems still to be the most frequently chosen base in the listing of collocations may be explained by the fact that the noun has traditionally been considered the point of departure in looking up collocations, thus becoming the focus of much twenty-first century research. Indeed, next to the continued interest in the lexical and grammatical composition of collocations and their treatment in dictionaries, recent research has been directed towards the dictionary-users themselves. Learners of English, in need of decoding what they cannot understand and/or encoding what they would like to express, have in the last decade or so become the target of more pedagogically-

oriented researches bent on testing the ease with which they can locate and select collocations in and from both general learner and specialized dictionaries. Despite the general appeal by linguists to step up studies of this kind (see Chen, 2016), some important findings have already begun to come to light. The very first is that learners “have inadequate dictionary use skills”. (Chen, 2016: 246). Indeed, from his study, in which fifty-two English majors at a Chinese university were asked to fill in the missing verbs in twelve v + n collocations gapped sentences, Chen noticed that students were reluctant to use the hyperlink function of the electronic dictionary to look up further information, were unable to distinguish between entry sub-senses, and lost their patience when faced with overcrowded entry information. In another study carried out by Lew and Radlowska (2010), intermediate pre-university Polish learners of English were requested to supply missing words in 13 gapped sentences by looking up collocations in the *Longman Dictionary of Contemporary English* (2003), and in the *Oxford Collocations Dictionary*. It is almost needless to say that the findings showed that “even fairly advanced learners experienced serious problems with locating and selecting appropriate collocations” (2010: 43). These findings, corroborated by similar results regarding learners of other languages (e.g. Alonso-Ramos, 2008; Alonso-Ramos & Garcia-Salida, 2019), have not just highlighted the need to teach language learners dictionary skills but rather and more precisely to teach them “collocation dictionary skills” (Kim, 2018: 322). This is even more impelling, when studies on language acquisition have long shown the difficulty learners have in gaining and developing strong collocational competence in general,⁴ let alone when combined to dictionary use.

Next to the need to teach learners how to use dictionaries, research in the last few years has also shown that dictionaries ought to be more user-friendly if they are to be of any real help for learners having to locate and select collocations. Since Herbst’s (1996: 336) claim that “the value for the learner is much greater if the special character of these combinations is pointed out by giving them typographical prominence of some sort,” experiments on general learner and specialized dictionaries have shown that learners find collocations more easily if they are highlighted (in colour or in bold print) as well as organized in boxes. (Götz-Votteler & Herbst, 2009, Heid, 2004, Laufer, 2011, Mittmann, 1999, Siepmann, 2006). In analyzing and comparing *Collins COBUILD Advanced Dictionary* (2008), *Longman Dictionary of Contemporary English*

⁴ For the state of the art of these studies see Henrikson (2013)

(2009) and *Oxford Advanced Learners' Dictionary of Current English* (2010), Dzemianko (2014: 272) more specifically finds that learners do better at finding and using collocations when they are highlighted in bold before or within examples as well as being placed at the bottom of an entry, and especially if the dictionary is online. Indeed, in an earlier experiment carried out by 64 upper-intermediate and advanced students of English at Poznan University, Dzemianko (2010) showed how the online version of *Collins COBUILD Advanced Dictionary* is more useful for students having to deal with receptive and productive tasks than its paper equivalent and that its use “results in better retention of meaning and more effective retrieval of collocations” (Dzemianko, 2010: 264). Unlike paper dictionaries in which it is difficult to look up multiword units, owing to the orthographical organization of entries (Lew, 2012), online dictionaries, where cross-referencing prevails, reduce lookup time (Dai *et al.*, 2019), thus helping learners to remain focused on the task and obtain better results. Contemporary empirical research on learners' skills in locating, selecting, and using collocations has in fact begun to show that many of the problems learners face could be of lesser importance in electronic dictionaries. The endless space in such dictionaries not only would allow for greater coverage but also for a more ubiquitous positioning of collocations (Gabuyian, 2019). Moreover, the shortcomings that the strict separation of the denotational information from the collocational one—typical of learners' dictionaries—along with the reduced amount of contextual information—typical of both general learner and specialized dictionaries—could be eliminated in dictionaries that have no spatial limits (Handl, 2009).

3.3 *Towards a new kind of dictionary*

Recent research has thus highlighted the need for a new type of dictionary that might overcome the shortcomings related to the selection and treatment of collocations as well as the weaknesses learners show when having to look for them. In her analytic study of seven dictionaries of English collocations: *BBI Combinatory Dictionary of English* (1997), *Selected English Collocations* (1988, 1998) and its companion *English Adverbial Collocations* (1991, 1998) the *Oxford Collocations Dictionary for Students of English* (2002), the *Collins COBUILD English Collocations on CD-Rom* (1995), *A Deskbook of most frequent English Collocations* (1986), and *A Dictionary of English*

Collocations (1994), Nuccorini (2003) underlines the need to strive for an ‘ideal’ dictionary, rid of the inconsistencies that feature in and across general learner and specialized dictionaries, in which the learners’ needs come first and foremost. It has already been suggested that, owing to their flexible structure, electronic dictionaries might be one step in the right direction towards meeting this need. Indeed, in having to select the correct collocations for a writing task, learners in Nurmukhamedov’s (2016) experiment performed better when using the online *Longman Dictionary of Contemporary English* than when they used the paper *Macmillan Collocation Dictionary* (2010), but did best of all when they used the collocation tool *wordandphrase.info*, which is not a dictionary in the strict sense of the word and does not even represent its traditional structure. Nurmukhamedov (2016: 472) motivates this by arguing that, besides including a clearer presentation of collocations, the *wordandphrase.info* tool has a “what-you-see-is-what-you-get interface” which learners in this digital age are now more used to, fostering more positive attitudes and better results.

To be able to locate, select, and use collocations successfully, learners therefore seem to necessitate tools that they are familiar with, and which are easy and fast to use. Digital tools may surely satisfy this requirement, provided they do not simply replicate the existing printed ones (Gabuyian, 2019: 213). As mentioned earlier, the advantage of an electronic dictionary is that it can include much more information than a printed one. This would allow, as Laufer (2011: 46) suggests, to include contrastive word-focused information that would draw attention to “the differences between the L2 and L1 ways of expressing similar meanings”. After all, since research has shown that learners opt for bilingual dictionaries more than monolingual ones for decoding but especially for encoding purposes (Atkins, 1985; MacFarquhar & Richards, 1983; Piotrowski, 1989; Rundell, 1999; Scholfield, 1999), one way forward could therefore be to create electronic bilingualized dictionaries of collocations.

“To provide a sound basis for the production of unabridged onomasiological bilingual learners’ dictionaries which focus on collocation” is what the Bilexicon project headed by Siepmann (2005b: 3) has been pursuing in the last decade or so. Besides fostering the production of electronic works that integrate monolingual with bilingual information, this project interestingly also encourages an onomasiological approach in entering collocations and looking them up. As opposed to the semasiological method, whereby collocations are positioned under

a word following an alphabetical order, the onomasiological method positions the collocation under a concept. A thematic organization of dictionaries, argues Siepmann, avoids the difficulty of deciding where to place the collocation, under the base or the collocate, which necessitates time to study the phraseological structure of each multiword unit. This of course is driven by Siepmann's idea of collocation: as "any holistic lexical, lexicogrammatical or semantic unit normally composed of two or more words which exhibits minimal recurrence within a particular discourse community (2005a: 438), it is the collocation itself, in Siepmann's mind, that should determine "the setting up and internal structuring of subareas and situation types" in a dictionary. And it is indeed the onomasiological method that fits in better with this idea of a dictionary rather than the semasiological one, which is better suited for the insertion of collocations in a "fully pre-determined ontological structure" (Siepmann, 2005b: 8). Within this line of thought, recent research has also been advocating for the inclusion in electronic dictionaries of visual networks of collocations, grouped together according to semantic content. Torner and Arias-Badia (2019: 271) claim that this would make "an easily-readable representation of complex lexical relations possible, avoiding the use of metalinguistic apparatus which can be difficult to manage for non-expert users". For example, the learner would undoubtedly benefit from finding the visual grouping of the four verbs *mostrar* ('show'), *expresar* ('express'), *demostrar* ('demonstrate'), and *manifestar* ('exhibit') in combination with *amor* ('love') under the concept of 'showing love' (see Torner & Arias-Badia, 2019) in a future bilingualized English-Spanish onomasiological electronic dictionary.

4. *Conclusions*

If "the idiomaticity of a language is perhaps best revealed in the errors committed by learners" (Fellbaum, 2007: 2), then we can safely say that English is still a highly idiomatic language, owing to the difficulty learners have in selecting the correct word from a vast pool of near-synonyms. As combinations of words that are seemingly unmotivated, collocations are particularly difficult for learners to understand and/or to formulate, often leading to a breakdown in communication. Consequently, since the mid-twentieth century lexicographers have been devoting time and

attention to selecting, entering, and explaining as many collocational patterns as possible in dedicated reference works in order to aid learners in their acquisition and use of them. This summative report on the state of the art of research in the field of monolingual learner lexicography and collocations has, however, shown that, despite the evident progress made, scholars in the last two decades are aware of the need to improve the way collocations are handled in such dictionaries. The inconsistent inclusion and treatment of collocations in and across learner and specialized dictionaries, combined with learners' unwillingness and incapacity to use them, still seems to hamper correct usage. In addressing this problem, researchers have come to realize that the solution might be a totally new kind of tool. Electronic, thematically structured, and bilingualized, this new kind of dictionary could foster prompt and easy searches that, departing from one's own mother-tongue, could result in finding equivalents more quickly and in selecting them more efficiently, thanks to the additional contextual information provided. Indeed, although research has shown that learners prefer bilingual dictionaries to monolingual ones because they address their own language issues, the stringent translational equivalents typical of bilingual tools often fail to provide the necessary contextual details required for optimal decoding and/or encoding purposes. By supporting these equivalents with an added wealth of illustrative examples, typical of monolingual works, the electronic bilingualized English dictionary of collocations could indeed become an exciting challenge for lexicographers and a better solution for learners.

References

- ALONSO-RAMOS, M. (2008). Papel de los Diccionarios de Colocaciones en la Enseñanza de Español Como L2. In Bernal, E. & DeCesaris, J. (eds.). *Proceedings of the XIII EURALEX International Congress*. Barcelona: Universitat Pompeu Fabra, pp.1215-1230.
- ALONSO-RAMOS, M. & GARCIA-SALIDA, M. (2019). Testing the use of a collocation retrieval tool without prior training by learners of Spanish. *International Journal of Lexicography*, 32(4), 480-497.
- ATKINS, B. T. S. (1985). Monolingual and Bilingual Learners' Dictionaries: A Comparison. In Ilson, R. (ed.), *Dictionaries, Lexicography and*

- Language Learning*. Oxford: Pergamon Press and British Council, pp. 15-24.
- BARTSCH, S. (2004). *Structural and Functional Properties of Collocations in English*. Tübingen: Gunter Narr Verlag Tübingen.
- BÉJOINT, H. (2010). *The Lexicography of English*. Oxford: Oxford University Press.
- BENSON, M., BENSON, E & ILSON, R. F. (1986). *The BBI Dictionary of English Word Combinations*. Amsterdam: John Benjamins Publishing Company.
- BURGER, H. (1998), *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt.
- CHEN, Y. (2016). Dictionary use for Collocation Production and Retention: a CALL-based study. *International Journal of Lexicography*, 30(2), 225-251.
- COFFEY, S. (2011). A New Pedagogical Dictionary of English Collocations. *International Journal of Lexicography*, 24(3), 328-341.
- COWIE, A. P. (1988). Stable and Creative Aspects of Vocabulary Use. In Carter, R. & McCarthy, M. J. (eds.), *Vocabulary and Language Teaching*. London: Longman, pp.126-137.
- COWIE, A. P. (1994). Phraseology. In Asher, R. E. (ed.), *The Encyclopedia of Language and Linguistics*. Oxford: Oxford University Press, pp. 3168-3171.
- CROWTHER, J., DIGNEN, S., & LEA, D. (eds.) (2002). *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.
- DAI, Y., WU, Z., XU, H. (2019). The Effect of Types of Dictionary Presentation on the Retention of Metaphorical Collocations: Involvement Load Hypothesis vs. Cognitive Load Theory. *International Journal of Lexicography*, 32(4), 411-43.
- DZIEMIANKO, A. (2010). Paper or electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography*, 23(3), 257-273.
- DZIEMIANKO, A. (2014). On the Presentation and Placement of Collocations in Monolingual English Learners' Dictionaries: Insights into Encoding and Retention. *International Journal of Lexicography*, 27(3), 259-279.
- EVERT, S. (2008). Corpora and Collocations. In Lüdeling, A. & Kytö, M. (eds), *Corpus Linguistics. An International Handbook*. Vol. 2. De Gruyter: Amsterdam, 1212–1248.
- FELLBAUM, C. (2017). Introduction. In Fellbaum C (ed.), *Idioms and Collocations*. London: Continuum, pp.1-22.

- FIRTH, J. R. (1957). Modes of Meaning. In Palmer, F.R. (ed.), *Papers in Linguistics 1934–1951*. Oxford: Oxford University Press, pp. 190–215.
- GARIBYAN, A. (2019). Interpreting Collocations: An analysis of Collocation Dictionaries of: English, German and Russia. *Lexicographica*, 35(1), 107-132.
- GÖTZ-VOTTELER, K. & HERBST, T. (2009). Innovation in Advanced Learner's Dictionaries of English. *Lexicographica*, 25, 47-66.
- HANDL, S. (2009). Towards Collocational Webs for Presenting Collocations in Learners' Dictionaries. In Barfield, A. & Henrik, G. (eds.), *Researching Collocations in Another Language. Multiple Interpretations*, Hampshire: Palgrave Macmillan, pp. 69-85.
- HAUSMANN, F. J. (1984). Wortschatzlernen ist Kollokationslernen. *Praxis des Neusprachlichen Unterrichts* 31: 395–406.
- HEID, U. (2004). On the Presentation of Collocations in Monolingual Dictionaries. In Williams, G. & Vessier, S. (eds.), *Proceedings of the 11th EURALEX International Congress*. Lorient: Université de Bretagne-Sud, pp. 729-738.
- HENRIKSEN, B. (2013). Research on L2 Learners' Collocational Competence and Development—A Progress Report. In Bardel, C., Lindqvist, C. & Laufer, B. (eds), *L2 Vocabulary Acquisition, Knowledge and Use*, Eurosla Monographs Series 2, pp. 29-56.
- HERBST, T. (1996). On the Way to the Perfect Learners' Dictionary: A First Comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography*, 8(4), 321–357.
- HOEY, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London:Routledge.
- KIM, S. (2018). EFL Learners' Dictionary Consultation Behaviour during the Revision Process to Correct Collocational Errors. *International Journal of Lexicography*, 31(3), 312-326.
- LAUFER, B. (2011). The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. *International Journal of Lexicography*, 24(1), 29–49.
- LEW, R. & RADLOWSKA, M. (2010). Navigating Dictionary Space: the Findability of English Collocations in a General Learner's Dictionary (LDOCE4) and Special-purpose Dictionary of Collocations (OCD). In Ciuk, A. & Molek-Kozakowska, K. (eds), *Exploring Space: Spatial Notions in Cultural, Literary and Language Studies*, Volume 2: Space in Language Studies. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 34-47.
- LEW, R. (2012). How Can We Make Electronic Dictionaries More

- Effective? In Granger, S. & Paquot, M. (eds.), *Electronic Lexicography*. Oxford: Oxford University Press, pp. 343-361.
- MACFARQUHAR, P. D. & RICHARDS, J. C. (1983). On Dictionaries and Definitions. *RELC Journal* 14, 111-124.
- MEL'ČUK, I. (1998). Collocations and Lexical Functions. In Cowie, A. P. (ed.), *Phraseology. Theory, Analysis and Applications*. Oxford: Clarendon Press, pp. 23-53.
- MITTMANN, B. (1999). The Treatment of Collocations in OALD5, LDOCE3, COBUILD2 and CIDE. In Herbst, T. & Popp, K. (eds), *The Perfect Learners' Dictionary*. Lexicographica: Series Maior, 95. Tübingen: Max Niemeyer Verlag, pp. 101-111.
- NUCCORINI, S. (2003). Towards an 'ideal' Dictionary of English Collocations. In van Sterkenburg, P., *A Practical Guide to Lexicography*. Amsterdam: John Benjamins, pp. 366-387.
- NURMUKHAMEDOV, U. (2016). The Contribution of Collocation Tools to Collocation Correction in Second Language Writing. *International Journal of Lexicography*, 30(4), 454-482.
- ORLANDI, A. & GIACOMINI, L. (2016). *Defining Collocation for Lexicographic Purposes: From Linguistic Theory to Lexicographic Practice*. Linguistic Insights. Bern: Peter Lang.
- OXFORD ENGLISH DICTIONARY* (online). Oxford: Oxford University Press.
- OXFORD NATIONAL DICTIONARY OF BIOGRAPHY* (online). Oxford: Oxford University Press.
- PALMER, H. E. (1933). *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- PIOTROWSKI, T. (1989). Monolingual and Bilingual dictionaries: Fundamental differences. In Tickoo, M. L. (ed.), *Learner's Dictionaries: State of the Art*. Singapore: Seamo Relc, pp. 72-83.
- RUNDELL, M. (1999). Dictionary Use in Production. *International Journal of Lexicography*, 12, 35-54.
- SCHOLFIELD, P. J. (1999). Dictionary Use in Reception. *International Journal of Lexicography* 12, 13-35.
- SIEPMANN, D. (2005a). Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography*, 18(4), 409-443.
- SIEPMANN, D. (2005b). Collocation, Colligation and Encoding dictionaries. Part I: Lexicographical aspects. *International Journal of Lexicography*, 19(1), 1-39.
- SIEPMANN, D. (2006). Collocation, Colligation and Encoding Dictionaries. Part II: Lexicographical Aspects. *International Journal of Lexicography*, 19(1), 1-39.

- SINCLAIR, J. M. (1987). *Collins Cobuild English Language Dictionary*. Birminham: Harper Collins.
- SINCLAIR, J. M. (1991). *Corpus, Concordance, Collocation (Describing English Language)*. Oxford: Oxford University Press.
- STEIN, G. (2002). *Better Words. Evaluating EFL Dictionaries*. Exeter: Exeter University Press.
- TORNER, S. & ARIAS-BADIA, B. (2019). Visual Networks as a Means of Representing Collocational Information in Electronic Dictionaries. *International Journal of Lexicography*, 32(3), 270-295.
- WALKER, C. (2009). The Treatment of Collocation by Learners' Dictionaries, Collocational Dictionaries and Dictionaries of Business English. *International Journal of Lexicography*, 22(3), 281-299.
- WILLIAMS, G. & MILLON, C. (2011). Palmer, Firth and Internet: Drawing together Collocational Threads. *Corpus Linguistics*, 1-34.

Valeria Fiasco*
Virginia Pulcini**

*'Overt' Calques from English and their Currency in Italian*¹

ABSTRACT:

This paper analyses a selection of loanwords and calques triggered by the influence of English and used in Italian. We have introduced the term 'overt' calques to denote borrowings that are used in Italian both as loanwords and calques, such as *full time* and *tempo pieno*. This particular status of 'overt' calques raises interesting questions concerning the existence of near synonyms, the typological profile of the replica with respect to the source word, the semantic features of equivalents and the underlying sociolinguistic and pragmatic components which influence the speakers' preferences either for the foreign or for the domestic form. The analysis is based on a sample of 22 'overt' calques and their equivalent loanwords. On the basis of lexicographic information, the chronology of borrowing and the development of calques is presented; we also illustrate common typological patterns of Italian calques and compare the usage frequency of the synonymic pairs (loanwords and calques) in three corpora of present-day Italian.

KEYWORDS: Anglicisms, Calques, Loanwords, Italian corpora

1. Introduction

The primary outcome of language contact is the transfer (borrowing or lending) of lexical units and phrases across speech communities and national languages. Scholars agree that borrowings can be grouped into two major categories, i.e. loanwords and calques. Most of the research on English borrowings, or Anglicisms, is focussed on loanwords, the type of borrowings that are imported into another language in the original 'foreign' appearance, with minor adaptation in form and pronunciation (e.g. *week-end*). While loanwords remain recognizably English, calques are formally made up of units belonging to the receiving language (RL), so that the meaning of the English source word is reproduced with a

* Università Roma Tre.

** Università di Torino.

¹ Both authors are responsible for the overall structure of this paper. V. Pulcini drafted sections 1 and 2.2; V. Fiasco was responsible for sections 2.1 and 2.3.

translation equivalent (*fine settimana*) or a new meaning is taken on by an already existing Italian word (e.g. *realizzare*, with the meaning of ‘to become aware of’, from English *realize*).

Because of the high degree of ‘camouflage’ in the RL, calques are difficult to identify in Italian discourse. Their lack of salience is further enhanced by the fact that English and Italian, though genetically unrelated, share a large stock of Latin-based vocabulary, so that the formal similarity of the source word and its Italian replica makes it difficult, if not impossible for the lay speaker, to be aware of the provenance of a term or phrase. This may be established only with the support of historical and sociolinguistic evidence, as in the case of the Italian terms *convenzione* (from En. *convention* < Fr. *convention*; etymon: Latin *conventiōn-em*), *impatto* (from En. *impact*; Fr. *impact*; etymon < Latin type **impactus* noun, < participial stem of *impingĕre*) or *ostruzionismo* (from En. *obstructionism*; etymon: Latin *obstructiōn*). In this respect, as argued by Bombi (2005), English has played an important role in the creation of Latin-based specialist terms and then in transferring them into Italian, often through the mediation of French. Hundreds of lexical items, be them in the form of adaptations or translations of exogenous terms, belong to the Italian word stock, but their historical identity remains ‘under cover’, as it were, by virtue of their Italian form.

As explained by Rodriguez Gonzalez & Knospe (2019), although English has been the most active donor language over the past century, quite a few calques mistakenly associated to English originated in other European languages, like the word *superman*, which actually comes from Ge. *Übermensch* and the Sp. neologism *centro comercial*, borrowed from Fr. *centre commercial*, though in turn adapted from En. *shopping centre*. In the complex scenario of European cultural history from the Renaissance to the present, much vocabulary travelled across speech communities so that multiple origins are the rule rather than the exception. It follows that independent national genesis seems to be a more plausible reason rather than borrowing for things or concepts that emerged in the same historical period: an emblematic example is the It. adjective *romantico* (from Lat *romanticus*), reportedly borrowed in 1824 from 17th c. En. *romantic* (with the meaning of ‘characteristic of a movement or style during the late 18th and 19th centuries in Europe marked by an emphasis on feeling, individuality, and passion [...]’ [OED], but also attested from Fr. *romantique*). Another independent outcome of language contact across European languages, to set apart from borrowing, is the category of ‘internationalisms’, i.e. lexical items

of Latin/Greek etymology that are formally and semantically similar across unrelated languages, a prototypical case being that of *telephone* and many scientific and technical terms (discussed by Pulcini, 2019).

Because of the difficulties outlined above to recognize routes of transmission and the origin of borrowings, when they are ‘camouflaged’ in domestic disguise, the vast literature on English-Italian contact and on the cultural and linguistic exchanges (Iamartino, 2001; Pulcini, 2002, 2017, 2020; Pulcini *et al.*, 2012) has mostly focussed on ‘direct’ Anglicisms, i.e. words or multi-word units borrowed from English without any formal integration or with some orthographic, phonological and morphological adjustments, which however leave the word ‘recognizably’ English. In Italian most Anglicisms are actively used in Italian in their original form with no competition with domestic words: among the hundreds of examples, suffice it to quote the names of some music genres (*rock, blues, hip-hop, rap*), names of sports (*tennis, rugby, curling*) and internet terms (*hashtag, blog, doodle*).² On the other hand, for several different reasons related to language contact and interference modes, some terms are readily adopted and rendered only with a domestic equivalent, and the English term from which they originated is never integrated or quickly falls into disuse: for example, *forno a microonde* (*microwave oven*), *aria condizionata* (*air conditioned*), *arrampicatore sociale* (*social climber*) and *disco volante* (*flying saucer*). Very often, however, the Anglicisms start being used alongside a domestic equivalent, which may be a newly created term or an already existing term/phrase, which then enters in competition with the English neologism.

This last outcome of interference leads to the category of borrowings that we will focus on in this paper, which we have termed as ‘overt’ calques. By ‘overt’ calques we refer to domestic lexical units that coexist with a loanword expressing the same semantic content, such as, for example, *tempo pieno* and *full time*, or *week-end* and *fine settimana*.³ We have chosen the term ‘overt’ because it clearly expresses the fact that the very existence of the loanword confirms the motivation and

² The number of entries having English origin in the GDU amounts to 8,196, of which 5,850 are labelled as ES (‘exoticism’) and have an English form; it follows that the number of adapted Anglicisms and calques is 2,340 (28.5% of the total). All the foreign words contained in the GDU are also recorded in a separate dictionary (De Mauro & Mancini, 2003)

³ This phenomenon has already been addressed by Winter-Froemel & Onysko (2012), who introduced the terms ‘catachrestic’ and ‘non-catachrestic’ innovations to refer to loanwords adopted to name something new vs. loanwords that convey a meaning already expressed by a domestic lexical unit of the RL, and their pragmatic values.

the origin of the calque. This particular status of ‘overt’ calques raises interesting questions related to the co-existence of synonymic doublets. This analysis is based on a sample of 22 ‘overt’ calques used in Italian: on the basis of previous research, lexicographic and corpus-based data, we will focus on the following features:

- the chronology of the selected loanwords and the development of calques;
- common typological patterns of Italian calques with respect to their English models;
- the usage frequency of the synonymic pairs (loanwords and calques).

Finally, on the basis of our data, we will try and suggest the reasons which may lead users to opt for one or the other form.

2. ‘Overt’ calques and synonymic loanwords

The present analysis is based on a sample of ‘overt’ calques and synonymic loanwords collected during the compilation of the Global Anglicisms Database (GLAD)⁴. Although GLAD’s word list contains mostly direct Anglicisms, we also considered candidate calques and checked their currency in dictionaries (Zingarelli, 2020; GDU, 2007; Treccani 2020; Devoto Oli, 2020), and in other lexicographic sources⁵. We also used newspaper archives (*La Repubblica* and *La Stampa*)⁶ for checking dates of adoption and finding authentic examples, which allowed us to antedate the borrowing of some of the focus items. Finally, the frequency of the competing forms were searched for in three Italian corpora, namely Coris⁷, Italian Web 2016 and Timestamped JSI Italian Corpus, the latter two accessed through the Sketch Engine platform⁸.

⁴ <<https://www.nhh.no/en/research-centres/global-anglicism-database-network/>>

⁵ The portal <aaa.italofonia.info> and *ArchiDATA*, *Archivio di (retro)datazioni lessicali* <<https://www.archidata.info/>>

⁶ www.repubblica.it; www.archiviolaStampa.it

⁷ <http://corpora.dslo.unibo.it/TCORIS/>

⁸ <https://www.sketchengine.eu/>

2.1

Table 1 shows the list of the selected English loanwords and their synonymic Italian calques, accompanied by the earliest date of adoption retrieved from the above-mentioned sources. The borrowing process normally begins from the adoption of the loanword, often within a specialist domain or sector of the general language, followed by the creation of the corresponding calque in the RL: this process is exemplified by the term *countdown*, introduced in the context of the launch of a spacecraft or of a missile, the meaning of which was later figuratively extended to a period of time preceding an important event. This loanword and its calque *conto alla rovescia* are attributed the same time of importation (1958). The same or a close date of adoption are attested for many of the listed items, such as, for example, *pay-tv/televisione a pagamento* (1936), *password/parola d'ordine* (1966), *self-control/autocontrollo* (1911), *supermarket/supermercato* (1956), *politically correct/politicamente corretto* (1991/1993).

This is not the only order of transmission. In other cases the creation of the calque precedes the borrowing of the underlying loanword, like the syntagmatic calque *conferenza al vertice*, introduced in 1960 on the compound *summit conference*, preceded by several unsuccessful replacements (cf. Bombi, 2005: 121) and followed a few years later by the elliptic English calque *summit*, which gradually won out in use over Italian *vertice*. Another case is *posta elettronica*, which started being used in 1982, much earlier than the shorter and more successful loanword *e-mail* (1992). Also *dopobarba* appeared before *after-shave*, initially as a calque of *after-shaving lotion* (*lozione dopobarba*), probably mediated by French *après rasage* (Bombi, 2005: 55). ArchiDATA provides an earlier attestation (1946) with respect to Italian dictionaries in the following citation [1]:

[1] “Marchio d’impresa depositato il 18 aprile 1946 da S.A.P.P.A. [...] a Milano, per lozione per *dopobarba* e profumeria. SMOOTH prodotto italiano” (source: Ministero dell’Industria e del Commercio, *Bollettino dei brevetti per invenzioni, modelli e marchi*, pt. III, Roma, Istituto Poligrafico dello Stato, 1947 [ottobre 1946], p. 1183).

English loanword	Attestation in Italian	Italian calque	Attestation in Italian
after-shave	1959	Dopobarba	1946
all inclusive	1995	tutto compreso tutto incluso	1972
basket	1906	Pallacanestro	1923
case study	1989	caso di studio studio di caso	1993 1992
corner (kick)	1934	(calcio d') angolo	1925
countdown	1958	conto alla rovescia	1958
e-mail	1992	posta elettronica	1982
full time	1963	(a) tempo pieno	1967
hard disk	1985	disco rigido	1988
live	1975	dal vivo	1981
pay-tv	1936	televisione a pagamento	1936
part time	1963	tempo parziale	1978
password	1966	parola d'ordine	1966
politically correct	1991	politicamente corretto	1993
self-control	1911	Autocontrollo	1911
step by step	1989	passo dopo passo	1999
talent scout	1936	scopritore di talenti	1935
shuttle	1981	Navetta	1983
Star	1929	Stella	1856
summit (conference)	1967	(conferenza al) vertice	1960
supermarket	1956	Supermercato	1956
week-end	1905	fine settimana	1911

Table 1. Selected Anglicisms and calques with first attestation in Italian

Another case of prior adoption of the calque with respect to the loanword is that of the sports term *calcio d'angolo*, normally reduced to *angolo*. Rather than a semantic calque of the English term *corner*, it seems that *calcio d'angolo* appeared as a loan translation of *corner kick*, which was then borrowed in its elliptic form *corner*, featuring in Italian side by side the domestic form *angolo*. It is also worth noting that most of the sports terminology was subject to intense 'Italianization' in the first half of the 20th century because of the political pressure of national purism (Cappuzzo, 2008; Pulcini, 2017).

The development of calques from the model English term can trigger

more than one replacive forms. For example, *all inclusive* coexists with the Italian expressions *tutto compreso* and *tutto incluso*, which can be used in the field of tourism. Previous research (Pulcini, 2012) has shown that *tutto compreso* is generally preferred (*formula tutto compreso, viaggi tutto compreso*), whereas *tutto incluso* is commonly found in connection with prices (*tariffa di lancio a partire da 736 euro tutto incluso a/r*).

The synonymic pair *week-end/fine settimana* deserves particular attention. Scholars agree that *week-end* appeared as early as 1905, followed by the syntagmatic calque *fine settimana*. It seems logical that the specific acceptation of ‘end of the week’ as a moment of relaxation, an outing or entertainment after a Monday-to-Friday working routine is a new modern meaning of the generic expression devoided of its social value. It was possible to antedate this acceptation to 1911 from the archive of the daily newspaper *La Stampa* (earlier uses specifically refer to the fields of economy and finance) (see example [2])

[2] Sezione “Annunzi vari”

Margherita troverai due annunci miei [...] partirò forse *fine settimana*, manderò l’indirizzo.

Finally, *case study* is an interesting case of deviant rendition caused by conflicting word order in English and in Italian. Following the compositional patterning of English noun phrases, the head element of *case study* is the one on the right, i.e. *study*, and the left-hand element is the modifier. Therefore, the correct equivalent calque should be *studio di caso*. A word for word rendition of the model term has led to the creation of the Italian equivalent *caso di studio* (187 hits in la Repubblica archive), which is in fact more frequently used in Italian than the correct calque *studio di caso* (only 4 hits in la Repubblica archive).

Turning the attention to semantic calques (or loans) such as *stella* and *navetta*, we can see that the process involves the acquisition of a new meaning of an already existing word in the RL. The term *stella* with the meaning of ‘famous person’ dates back to 1856 as a semantic calque of English *star* (introduced decades later in 1929). In the case of *navetta* (used in Italian since the 14th century to denote a small boat), the modern meaning of ‘means of transport operating a transfer service to and from a certain destination, like airports and stations, at regular times’ was taken on at the time of increased mass tourism, possibly favoured by the term *navetta spaziale* (in turn a calque of *space shuttle*, cf. Bombi, 2005: 139).

2.2

The typology of structural and semantic calques can be quite complex, if we consider the Italian replicas triggered by English source models. The most straightforward categorization of indirect loans, adopted by Görlach (2001) and derived from earlier taxonomies (e.g. Weinreich, 1953), distinguishes between calques and semantic loans. In turn, calques can take the form of loan translation (faithful reproduction of the model), loan rendition (divergent reproduction of the model) and loan creation (free reproduction of the model). A similar categorization of types of lexical borrowings (starting from the distinction between direct and indirect loans) is presented by Pulcini *et al.* (2012: 6), extending the possible patterns from lexical to phrasal (e.g. *step by step* in our sample). For the specific categorization of Italian calques, more refined models are those proposed by Klajn (1972) and by Bombi (2005), in turn drawing on Gusmani (1986). As anticipated in the introduction, calques can reproduce both the structure and the meaning of the foreign model or attach a new meaning to an already existing word in the RL. According to Bombi (2005) in the former case we obtain a ‘structural calque’, in the latter case a ‘semantic calque’. The term calque is largely shared in the literature on language contact, whereas for the type of interference that involves only the development of a new meaning for an already existing word, also the terms ‘semantic loan’ (Pulcini *et al.*, 2012) and ‘*prestito semantico*’ (Klajn, 1972) are used.

The selected calques presented in Table 1 are current in Italian and easily recognizable by speakers, although most of them were first introduced in a specialized domain of vocabulary and then spread to the general language. A common structure is the compositional one (*calco strutturale di composizione* cf. Bombi, 2005), whereby the model word is reproduced in the RL: *dopobarba*, *autocontrollo* and *supermercato* are the precise replica of *aftershave*, *self-control* and *supermarket* and therefore considered as ‘perfect’ calques (*calchi perfetti*). The order of the elements follows the pattern modifier+modified (determinans+determinatum), which was common in old Italian, but today has given way to the more frequent reversed order, as in *pallacanestro* for *basket-(ball)*. Frequently one of the elements is a neoclassical combining forms such as *auto-* and *super-* (Pulcini & Milani, 2017). Some solid compounds are instead translated with analytic phrasal patterns, such as *countdown* and *password*, which are rendered in Italian as *conto alla rovescia* and *parola d’ordine*.

Another common type of calque in Italian involves a phrasal pattern, such as *dal vivo* for *live*, used as adjective or adverb. In general phrasal calques are modelled on a similar phrasal pattern in English (*calco sintagmatico*, cf. Bombi, 2005), but not always. An example of a ‘perfect’ phrasal calque is *politicamente corretto* (adverb+adjective) for *politically correct*. A common type of phrase in English involves the pattern adjective+noun (*full time, part time, hard disk*) where we can see that in Italian the replicas display a reversed order (*tempo pieno, tempo parziale, disco rigido*). Other parts of speech may be involved as in *pay-tv* (verb+noun), in which case we may observe the frequent Italian pattern constituted by a substantive+prepositional phrase (*televisione/tv a pagamento*). We can say that Italian calques are generally ‘imperfect’ with respect to the English model. Moreover, the divergent structure is accompanied by divergent meaning (loan rendition, cf. Pulcini *et al.*, 2012), as in *talent scout*, rendered as *scopritore di talenti* (literally ‘discoverer of talents’).

2.3

The competition between the loanword and the equivalent calque can be observed and measured by searching for the focus items in Italian corpora. For the present study, three corpora of present-day Italian have been queried, namely, the *CORIS*, the *Italian Web 2016*, and the *Timestamped JSI Italian Corpus*, in order to extract and compare the frequency of the English loanwords and calques in Italian. This is a research question that dictionaries cannot answer, as only a few dictionaries provide information about usage frequency. *CORIS* (*Corpus di Italiano Scritto*) is a general reference corpus of contemporary written Italian, including 150 million running words from 1980 to 2016 and texts from the press, fiction and academic prose. The *Italian Web 2016* corpus, also known as *itTenTen16*, is an automatically collected and processed corpus consisting of web-based texts, collected (crawled) in 2016, consisting of 4.9 billion words and available on the *Sketch Engine* platform. The *Italian Timestamped Corpus* is made up of news articles obtained from RSS feeds, covering the period 2014-2020, with a size of 5.8+ billion words. This new suite of corpora is most promising for the analysis of frequency trends of neologisms, as data can be searched according to times and subjects.

The figures listed in Table 2 allow a comparison between the

usage frequency of Anglicisms and their Italian equivalents⁹. For better comprehension of the data, we have discussed the focus terms according to three main trends: the first group includes the cases when the Anglicisms are prevalent in all three corpora, the second contains the cases where the Italian calques are preferred, and the third features cases where preferences diverge between the million-size traditionally sampled corpus (CORIS) and the two web-based billion-size corpora.

Starting from the terms for which there is a consensus among the Italian corpora on the prevalent use of the Anglicisms, these include *basket*, *e-mail*, *hard disk*, *pay-tv*, *part time*, *password*, *talent scout*, *star*, *week-end* and *summit*. It is not surprising that some of the words belong to information technology, a field that has rapidly grown since the 1990s, spreading from specialist the general use, which is today the most productive field of English neologisms (Gianni, 1994; Pulcini, 2017). Other qualities favouring Anglicisms against Italian equivalents may be English brevity (cf. *pay-tv* vs. *tv a pagamento*) but also the aura of modernity and prestige of the donor culture. The fact that Anglicisms are monoreferential (*star* vs *stella*; *summit* vs *vertice*) may also play a role in favour of Anglicisms.

The second group includes Italian words that are more frequently selected in actual use than their synonymous Anglicisms, featuring *dopobarba*, *calcio d'angolo*, *autocontrollo*, *passo dopo passo*, *navetta*, and *supermercato*. A feature that is readily evident is the presence of the neoclassical combining forms *auto-* and *super-*, which are quite productive in Italian, and therefore may be more readily combined with another Italian element (Pulcini & Milani, 2017). The phrase *calcio d'angolo* is prevalent in all corpora, but the choice between *corner* and *angolo* is pretty balanced. The preference for *passo dopo passo* and *navetta* could be explained resorting to semantic opacity of *step by step* or the difficult pronunciation of *shuttle*, the latter giving way to the much nicer-sounding, feminine noun *navetta* (the Italian suffix *-etto* conveys an affective connotation of something small and pretty). These conclusions are based on intuition, to be tested empirically.

⁹ In order to compare corpora of different sizes, the usage frequencies have been normalized to 1 million.

	CORIS freq/pmw	Italian Web 2016 freq/pmw	Timestamped JSI Italian Corpus freq/pmw
after-shave <i>dopobarba</i>	0.03 0.62	0.05 0.45	0.01 0.06
all inclusive <i>tutto compreso</i> <i>tutto incluso</i>	0.03 1.61 0.14	1.95 1.53 0.76	6.95 0.41 0.29
basket <i>pallacanestro</i>	5.8 1.48	13.45 4.5	25.65 6.64
case study <i>studio di caso</i> <i>caso di studio</i>	0.09 0.07 0.2	0.50 0.28 1.24	1.16 0.02 0.38
corner (kick) 10 <i>(calcio d') angolo</i> <i>calcio d'angolo</i>	0 0 0.75	0.35 0.67 1.35	0.37 1.21 1.99
countdown <i>conto alla rovescia</i>	0.2 0	1.17 1.9	1.9 3.69
e-mail <i>posta elettronica</i>	20.24 15.6	101.5 17.5	63.7 9.64
full time <i>(a) tempo pieno</i>	0.26 13.76	3.81 7.42	4.93 5.3
hard disk <i>disco rigido</i>	2.52 1	6.97 1.4	1.67 0.34
live <i>dal vivo</i>	4.22 7.71	32.56 19	89.95 15.85
pay-tv <i>televisione a pagamento</i> <i>tv a pagamento</i>	1.63 0.20 0.42	1.15 0.05 0.11	3.97 0.03 0.19
part time <i>tempo parziale</i>	7.04 3.6	7.79 1.65	7.73 0.73
password <i>parola d'ordine</i>	5.83 4.84	25.72 3.09	10.84 3.07
politically correct <i>politicamente corretto</i>	1 0.73	0.74 0.82	0.62 1.23
self-control <i>autocontrollo</i>	0.2 2.34	0.20 2.73	0.11 1.42

Table 2. Frequency of calques and loanwords in Italian corpora.

¹⁰ The terms *angolo* and *corner* are polysemous in Italian. *Angolo* denotes multiple referents such as 'geometric shape', 'part of a building', 'hidden place', and several others, including the football term, while *corner* is used in football but may also refer to

The third group includes words whose frequency diverges between Coris and the other two corpora. While Coris seems to prefer the Italian units *tutto compreso*, *tempo pieno* and *dal vivo*, the two corpora containing articles from the web and newsfeeds are more in favour of the Anglicisms *all inclusive*, *full time* and *live*. By contrast, the Anglicisms *countdown* and *politically correct* are more frequently used in Coris whereas *conto alla rovescia* and *politicamente corretto* are preferred by the other two corpora. In these cases, it would be necessary to carry out a more fine-grained qualitative analysis of the usage contexts to come up with more solid conclusions, which lies outside the scope of the present study.

Final remarks

The development of calques from English loanwords is considered by many linguists an enrichment for the Italian language both in terms of lexical growth and for the study of language contact (Bombi, 2005). The continuous inflow of Anglicisms and the creation of calques is favoured by the classical roots of many English loanwords, which also blurs and hides the origin of transmission and makes the loanword look domestic in form and meaning, when it is adapted or translated into Italian. It is therefore important to distinguish lexical items that may have developed out of independent genesis across a globalized world from words that have been imported from Anglo-American societies, integrated and translated into different languages and cultures.

In this paper we looked at a sample of ‘overt’ calques in Italian, namely lexical items that coexist with the equivalent Anglicisms from which they developed. Considering the dates of adoption, we could confirm that 16 out of 22 items (more than 70%) developed soon after or simultaneously to their synonymic Anglicism, whereas in other cases, typically for semantic calques, already existing words underwent

an area of a shop selling a single brand or product. Since it was not possible to isolate meanings, we calculated frequency roughly on the basis of their collocations. The three most frequent collocates of *angolo* and *corner* were taken into account, their absolute frequencies were summed, compared and then normalized to 1 million. The data suggest that both words are used in the field of football in Italian with a slight variation and that they often co-occur with the noun ‘*palla*’ (*ball*), the verbs ‘*battere*’ (*kick*) and ‘*deviare*’ (*deflect*).

a semantic extension or switch to another or a more general meaning.

As far as usage frequency is concerned, corpus data confirmed that Anglicisms and related calques are low-frequency items; in fact only a few show a frequency above 10/ pmw, namely *e-mail/posta elettronica*, *week-end/fine settimana*, *live/dal vivo*, *password* (but not *parola d’ordine*), and *basket* (but not *pallacanestro*), *supermercato* (but not *supermarket*). Most of the randomly chosen examples are more frequently used in the form of loanword rather than calque, with some exceptions regarding Latin-derived lexical items. Furthermore, a comparison between a smaller size, sampled corpus of Italian like Coris reflected a preference for calques with respect to very large web-based corpora, which display higher figures in favour of Anglicisms.

The ongoing reaction of Italian linguists and language observers, through official institutions like the *Accademia della Crusca* (cf. Marazzini & Petralli, 2015), awareness raising campaigns against the excessive use of English and other popular forms of linguistic nationalism (e.g. the online petition “dillo in italiano”¹¹) may indeed reverse the tide and align Italy to countries like France and Spain, whose institutions systematically propose/impose domestic translation equivalents for loanwords to obscure the interference of exogenous influences on the national language. So far, the influence of the mass media in Italy seems to have overruled speakers’ attitudes and preferences in favour of Anglicisms rather than of calques.

References

- BOMBI, R. (2005). *La Linguistica del contatto - Tipologie di anglicismi nell’italiano contemporaneo e riflessi metalinguistici*. Roma: Il Calamo.
- CAPPUZZO, B. (2008). Calques and Loanwords from English in Italian Specialized Press. An Investigation on the Names of Sports. *Rivista della Facoltà di Scienze Motorie dell’Università degli Studi di Palermo*, 1, 129-139.
- DE MAURO, T. (1999). *Grande dizionario italiano dell’uso*. 6 vols + cd-rom (GDU 2007). Torino: UTET.

¹¹ <https://www.change.org/p/un-intervento-per-la-lingua-italiana-dilloinitaliano>. See also Giovanardi *et al.* (2008) and Zoppetti (2017; 2018).

- DE MAURO T. & MANCINI, M. (2003) [2001]. *Parole straniere nella lingua italiana*, 2nd ed., Milano: Garzanti.
- DEVOTO, G. *et al.* (2020). *Nuovo Devoto Oli. Il vocabolario dell'italiano contemporaneo*, Milano: Mondadori Education, Le Monnier.
- GIANNI, M. (1994). Influenze dell'inglese sulla terminologia informatica italiana, *Studi di lessicografia italiana*, XII, pp. 273-299.
- GIOVANARDI, C. *et al.* (2008). *Inglese-italiano I a I. Tradurre o non tradurre le parole inglesi?* Lecce: Manni. *Global Anglicisms Database* (GLAD). <<https://www.nhh.no/en/research-centres/global-anglicism-database-network/>> (accessed August 2022).
- GÖRLACH, M. (ed.) (2001). *A Dictionary of European Anglicisms*. Oxford: Oxford University Press.
- GUSMANI, R. (1986). *Saggi sull'interferenza linguistica*, 2nd ed., Firenze: Le Lettere.
- IAMARTINO, G. (2001). "La contrastività italiano-inglese in prospettiva storica". *Rassegna Italiana di Linguistica Applicata (RILA)*, XXXIII, 7-130.
- KLAJN, I. (1972). *Influssi inglesi nella lingua italiana*. Firenze: Olschki.
- MARAZZINI, C., PETRALLI, A. (2015). *La lingua italiana e le lingue romanze di fronte agli anglicismi*, Firenze, goWare & Accademia della Crusca. *Oxford English Dictionary* <www.oed.com> (last accessed, October, 2022)
- PULCINI, V. (2002). "Italian". In GÖRLACH M. (ed.), *English in Europe*, Oxford: Oxford University Press, pp. 151-167.
- PULCINI, V. (2012). "Register Variation in Tourism Terminology" . In FACCHINETTI, R. (ed.). *A Cultural Journey through the English Lexicon*. Newcastle Upon Tyne: Cambridge Scholars Publishing, pp. 109-131.
- PULCINI, V. (2017). "Anglicisms in Italian: Moving on into the third millennium". In BOGGIO, C., MOLINO, A. (eds.). *English in Italy: Linguistic, educational and professional challenges*. Milano: Franco Angeli, pp. 13-35.
- PULCINI, V. (2019). Internationalisms, Anglo-Latinisms and other kinship ties between Italian and English. In MAŃCZAK-WOHLFELD E. (ed.). *Special issue on Anglicisms, Studia Linguistica Universitatis Jagellonicae Cracoviensis*, 136, 121-141.
- PULCINI, V. (2020). The English language and Anglo-American culture in Twentieth-Century Italy. In BONSAVER, G. *et al.* (eds.) *Understanding Cultural Change: Language and Narrative between Italy and the USA*, Oxford: Legenda, pp. 31-46.

- PULCINI V. & MILANI, M. (2017). Neo-classical Combining Forms in English loanwords: Evidence from Italian. *ESP Across Cultures*, 14, 175-196.
- PULCINI, V. *et al.* (2012). The lexical influence of English on European languages: From words to phraseology. In FURIASSI, C. *et al.* (eds.), *The Anglicization of European Lexis*, Amsterdam & Philadelphia: John Benjamins, pp. 1-24.
- RODRÍGUEZ GONZÁLEZ, F., KNOSPE, S. (2019). The variation of calques in European languages, with particular reference to Spanish and German: Main patterns and trends, *Folia Linguistica*, 53(1), 233–276.
- TRECCANI, *Il Vocabolario Treccani* online, <http://www.treccani.it/vocabolario/> (Accessed August 2022).
- WEINREICH, U. (1953). *Languages in Contact: Findings and Problems*, New York: Linguistic Circle of New York.
- WINTER-FROEMEL, E. & ONYSKO (2012), “Proposing a pragmatic distinction for lexical Anglicisms”, in FURIASSI C. *et al.* (eds.), *The Anglicization of European Lexis*, Amsterdam & Philadelphia: John Benjamins, pp. 43–64.
- ZINGARELLI N. (2019). *Lo Zingarelli 2020, Vocabolario della lingua italiana* (12 ed.), Bologna: Zanichelli.
- ZOPPETTI, A. (2017). *Diciamolo in italiano. Gli abusi dell'inglese nel lessico dell'Italia e incolla*. Milano: Hoepli.
- ZOPPETTI, A. (2018). *L'etichettario. Dizionario di alternative italiane a 1800 parole inglesi*, Firenze: Cesati.