
Creating a Bilingual Italian-English Dictionary of Collocations

Barbara Berti, Laura Pinnavaia
Università degli Studi dell'Insubria, Università Statale di Milano
barbara.berti@uninsubria.it, laura.pinnavaia@unimi.it

Abstract

Collocations are among the most challenging issues that learners of a foreign language have to face. The main obstacle for Italian learners of English is encountered in active tasks, and especially, when translating or writing essays in English. The encoding of collocations in fact often results in the selection of wrong terms. There is no doubt that the production and translation of Italian collocations would be facilitated were students to have at their disposal a bilingual Italian-English Dictionary of Collocations. The aim of this work is to endorse the importance of the existence of such a tool and to show the beginnings of its compilation.

Keywords: collocations; bilingual lexicography; corpus linguistics.

It is not so much the words of English nor the grammar of English that makes English difficult ... The vague and undefined obstacle to progress ... consists for the most part in the existence of so many odd comings-together-of-words. (Cowie 1999: 52-53)

1 Introduction

At the intersection between syntax and semantics, collocations are among the most challenging issues that learners of a foreign language have to face. Not only are they massively present in languages (Hoey 2005), but they are also (or at least appear to be) arbitrary and highly specific for a given language. This poses problems in second language acquisition. In fact, at some stage of the learning process, learners will inevitably have to deal with the way in which words combine with each other and the restrictions that they are subject to. Moreover, their first language will be likely to interfere to some extent with the production of natural-sounding word combinations in the second language and will cause them to produce incorrect *verbum pro verbo* translations (e.g. *make a photo** as a calque of *fare una foto*).

The main obstacle for learners of English is, in fact, encountered in production tasks, and, especially, when translating from Italian, or writing essays in English. In reception tasks, the meaning of English collocations can often be derived quite easily, especially if the context in which they occur is suf-

ficiently clear. For example, while in isolation the meaning of the collocation *to explode a myth* might be opaque to a learner of English, in an appropriate context, such as *scientists wanted to explode the myth of a monster living in the lake*, it becomes apparent. However, to produce the collocation from scratch would undoubtedly be a problem for the learner. It is highly likely that he/she would not select the correct terms, especially if, as in the case of Italian, the collocate of the noun *myth* is semantically unrelated to the verb *explode* (*sfatare un mito*). It thus becomes apparent that learners need to have at their disposal a resource that will help them to overcome such problems.

Nowadays, English monolingual lexicography does tackle the issue of collocations quite extensively, and especially dictionaries intended for learners of English; yet, there is evidence that Italian learners of English feel more at ease using bilingual tools rather than monolingual ones (Baxter 1980; Bensoussan, Sim and Weis 1984; Atkins 1985; Atkins and Knowles 1990; MacFarquhar and Richards 1983; Piotrowski 1989; Rundell 1999; Scholfield 1999). Unfortunately, however, Italian-English bilingual lexicographical resources poorly document this phenomenon from both a quantitative and a qualitative point of view (Berti 2010, 2012). That is why, in addition to implementing bilingual dictionaries with more and more carefully selected word combinations, a bilingual Italian-English dictionary of collocations would be very useful for Italian learners of English.

Successive to a theoretically-oriented study regarding the creation of a corpus-based Italian-English dictionary of collocations (Berti & Pinnavaia 2012), this work comes as a practical presentation of what selecting collocates entails, followed by two sample entries. The dictionary has as its principal target Italian learners of English, as well as professional translators. It will be unidirectional, presenting Italian collocations with their equivalents in the English language. The aim is not just to provide viable translations of Italian collocations, but also to offer users the most suitable English equivalents based upon frequency of use and situational appropriateness. This can only be achieved if the equivalents are drawn from linguistic corpora of English and carefully evaluated by the compilers, whose linguistic expertise must be bilingual.

2 Methodology

Nowadays, the use of corpora is of fundamental importance in lexicography (Atkins 1994), offering a more “objective” perspective on language. Over the years, in fact, it has become progressively clearer that resources based on the lexicographers’ *Sprachgefühl* are subjected to a number of limitations and that they cannot be regarded as fully representative of the current state of a language. Diastratic and diatopic variations, together with a speaker’s idiolect, exert an influence on the type of material that is included or excluded in and from a lexicographic tool, and even the work of a composite team of experts cannot guarantee an objective treatment of the language, or the selection of appropriate material.

In particular, as regards collocations, research shows that acceptability is extremely subjective and that the individuals' knowledge of word combinations in their language might vary to a great extent (Berti, *forthcoming*; Nesselhauf 2003). Acceptability depends, firstly, on semantics and, secondly, and most importantly, on usage. Indeed, it is usage that ultimately provides us with the intuition we need in order to discern between acceptable and unacceptable combinations in our mother tongue. Yet, given that the linguistic experience of each individual is absolutely unique and that the factors that have an influence on it are numerous, there will be a large grey area of disagreement in the judgement of word combinations. For this reason, it is nowadays essential to query corpora for the selection of the lexical material to be used for lexicographical purposes.

Since they are the parts of speech that occur more often in collocational patterns and are the most needed by students, it is nouns and verbs that we decided to select. This choice also seemed reasonable in virtue of the manner in which our thoughts are shaped. A speaker, unaware of the adjective + noun combination *broad daylight*, will more naturally think of the noun *daylight* before its qualifier. Similarly, the retrieval of any verb and adverb combination will normally imply the consultation of the verb before that of the adverb. This seems to be the rationale behind the organization of most dictionaries of collocations, among others, Rundell (2010) and the Oxford Dictionary of Collocations (see ODC 2002). The idea is thus to create an onomasiological dictionary that relies on the well-established distinction between base and collocate.

For the present study we took the Italian noun *odore* (a smell) and verb *pagare* (to pay) as our working examples retrieving the Italian collocations from the *Dizionario delle collocazioni* (Tiberii 2012). The decision to collect our sources from an already existing dictionary of collocations as opposed to extracting collocations from an Italian corpus was taken on recognizing the efficiency of this tool and after having ascertained the overall paucity of general Italian corpora compared to English.

We proceeded to find the collocates of the English equivalents of *odore* and *pagare*, *smell* and *pay* in the British National Corpus (BNC). Each Italian collocate was then matched with its English semantic equivalent. The explanation of the procedure for and the problems encountered with the pairing off of items ensues.

2.1 The Noun *Odore*

In Tiberii (2012), the following collocates for the noun *odore* were found:

odore *nm* accattivante, acre, acuto, aspro, buono, caratteristico, delicate, disgustoso, distintivo, fetido, forte, fresco, gradevole, inconfondibile, inebriante, intense, invitante, leggero, nauseabondo, nauseante, opprimente, particolare, penetrante pungente, ripugnante, rivoltante, sgradevole, sottile, stomachevole, strano, tenue, terribile, vago.

Considering that the Italian collocates for *odore* are all adjectives, we searched for adjective-noun collocational patterns in the BNC, obtaining the following one hundred collocates for the noun *smell*, listed in order of decreasing frequency:

- sweet
- strong
- pungent
- strange
- musty
- rich
- faint
- sour
- bad
- unpleasant
- acrid
- stale
- warm
- delicious
- fresh
- heavy
- horrible
- lovely
- different
- familiar
- good
- nice
- able
- clean
- damp
- distinctive
- sickly
- dank
- funny
- other
- overpowering
- wonderful
- awful
- hot
- new
- old
- pleasant
- aware
- earthy
- lingering
- peculiar
- thick
- close
- distinct
- metallic
- oily
- foul
- full
- masculine
- musky
- open
- particular
- sharp
- spicy
- chemical
- cold
- cloying
- comfortable
- dirty
- fishy
- fragrant
- free
- great
- green
- milky
- nasty
- putrid
- real
- slight
- smoky
- soft
- tangy
- usual
- very
- whole
- curious
- dangerous
- comforting
- clinical
- characteristic
- burning
- blue
- better
- bitter-sweet
- acute
- dry
- evil
- fine
- homely
- keen
- little
- noxious
- powerful
- rancid
- rotten
- savoury
- sick
- stronger
- supposed
- sure

The next step was to match each Italian collocate from Tiberii (2012) with an equivalent English one from the above list. For example, we paired off *odore accattivante* with *captivating smell*. On finding Italian collocations that correspond to two or more English ones, as in the case of *odore gradevole* which is equivalent to both *sweet* and *pleasant smell*, none were discarded and all were kept. This was done even when one collocation was found to be more frequent than another (i.e. *sweet smell*), out of scientific rigour, on the one hand, and in order to make up for the scant number of collocations provided in bilingual dictionaries, which were naturally also consulted, on the other. When, moreover, it was noticed that some of the collocates in Italian are synonymous, these were grouped together; for instance, *acre* and *aspro*, can both be rendered as *acid*, *sour* or *sharp* in English, just as *leggero*, *tenue*, *sottile*, *vago* can be translated as *slight* and *faint*. When the close examination of the list of English collocations revealed the absence of a number of Italian collocations in the source text, it was deemed opportune to introduce them, especially if the English equivalents are frequently occurring collocations. For example, it seemed important to include the collocations *cattivo odore* and *odore stucchevole*, missing from Tiberii (2012), aware of the fact that the collocations *unpleasant/musty smell* and *heavy/cloying smell* are commonly used in English. Lastly, when equivalents for Italian collocates were not found in the BNC, we consulted WebCorp, which owing to its larger size, often managed to provide suitable equivalent adjectives, such as *captivating*, *delicate*, *intoxicating* and *penetrating smell* for *accattivante*, *delicato*, *inebriante* and *penetrante*.

2.2 The Verb *Pagare*

Under the lemma *pagare*, Tiberii (2012) lists the following adverbial collocates:

pagare v. abbondantemente, anticipatamente, caro, comodamente, completamente, generosamente, immediatamente, integralmente, interamente, obbligatoriamente, pesantemente, profumatamente, prontamente, provvisoriamente, puntualmente, regolarmente, tempestivamente, volontariamente.

While the collocates we retrieved in the BNC for *pay* amount to forty-three, only eight have been reported here. This is because, unlike the list of collocates for *smell*, the list for *pay* included many that were not pertinent. It was necessary, in fact, to shortlist the relevant ones by examining the concordances. During this process, we also decided to exclude collocates with a single occurrence, following Sinclair's tenet that a single occurrence does not represent a "settled pattern in the language" (2003: 15).

- promptly
- dearly
- hereby
- handsomely
- proportionately
- compulsorily

- upfront
- punctually

As earlier, the Italian and English items were matched. For example, *pagare obbligatoriamente* was matched with *pay compulsorily*. Similarly, synonymous collocations were grouped together, as in *pagare abbondantemente/generosamente/profumatamente*, which were all associated to *pay handsomely*. However, the pairing off of collocates proved to be more problematic than it was for *odore/smell*. Firstly, because there are very few correspondences between the Italian and English collocates; secondly, because it is not always easy to understand the real meaning of certain Italian collocations on account of the absence of contextual data in Tiberii (2012). For example, *pagare anticipatamente/comodamente/integralmente/interamente/puntualmente/regolarmente/volontariamente* do not have a corresponding collocate in the BNC. For this reason, in order to provide an equivalent, not only did we have to consult various bilingual dictionaries along with WebCorp, but we also had to rely on our own knowledge of the language. At times this resulted in equivalent collocations (e.g. *pagare comodamente* = *pay easily*), at other times in lexical sequences that do not fit the verb + adverb pattern (e.g. *pagare anticipatamente* = *to pay in advance*), at other times in single words (e.g. *pagare anticipatamente* = *to prepay*). The results we obtained can be observed in the sample provided in the following section. Not having examples of usage to refer to made the translation of the items *pagare completamente* and *pagare provvisoriamente* particularly problematic: while the former sounds like a free combination, the latter is clearly semantically ambiguous. The way in which the collocates of *pagare/pay* have been rendered in English can be observed in the sample provided in the following section.

3 The Sample Entries

Here follow the sample entries for the noun *odore* and verb *pagare*.

odore - smell n.

ADJ. + NOUN

accattivante – captivating
acre/aspro – acrid/sour/sharp
acuto/pungente – pungent
buono – good/nice/delicious
caratteristico – characteristic
cattivo/sgradevole – unpleasant/musty
delicato – delicate
disgustoso/ripugnante/rivoltante/terribile – bad/ horrible/nasty/awful
distintivo – distinctive
fetido – foul/putrid
forte – strong/overpowering
fresco – fresh
gradevole – sweet/pleasant

inconfondibile – distinct
inebriante – intoxicating
intenso – powerful/keen
invitante – warm
leggero/tenue/sottile/vago – slight/faint
nauseabondo/nauseante/stomachevole – sickly
opprimente – overpowering
particolare – curious/peculiar/particular
penetrante – penetrating
persistente – lingering
strano – strange
stucchevole – heavy/cloying

pagare - pay v.

ADV. + VERB

Abbondantemente/generosamente/profumatamente – handsomely
anticipatamente – in advance/prepay (v.)
caro/pesantemente – dearly
comodamente – easily
immediatamente/prontamente/tempestivamente – promptly
integralmente/interamente – in full
obbligatoriamente – compulsorily
puntualmente – punctually
regolarmente – regularly
volontariamente – voluntarily

As the examples show, the entries provide a list of the Italian collocates, documented for the noun *odore* and for the verb *pagare*, along with their respective adjectival and adverbial equivalents for the English noun *smell* and verb *pay*. The Italian collocates appear listed in alphabetical order, and sometimes include more than one exemplar when they have similar meanings (e.g. *odore disgustoso/ripugnante/rivoltante/terribile*; *pagare immediatamente/prontamente/tempestivamente*). Next to them appear the equivalent English collocates, retrieved from the data in the BNC and WebCorp. As can be seen, the correspondences can be of various types. It is possible to have collocations in Italian and English that have the same number of collocates (e.g. *strano odore = a strange smell*; *cattivo/sgradevole odore = unpleasant/musty smell*; *pagare puntualmente = to pay punctually*); sometimes the Italian language has more (*nauseante, nauseabondo, stomachevole odore = sickly smell*; *pagare immediatamente/prontamente/tempestivamente = to pay promptly*); at other times, fewer (e.g. *odore buono = good/nice/delicious smell*; *pagare anticipatamente = to pay in advance/to prepay*). In all cases, it is our intention that dictionary-users have all the most frequent and most appropriate equivalent English collocations for the Italian ones.

4 Conclusions

While this work is still in progress and has still not fully solved all the problems involved in retrieving and selecting equivalent English collocates for Italian ones, the objective of compiling a bilingual dictionary of Italian and English collocations is, in our view, invaluable for the quantitative and

qualitative advantages it can offer next to general bilingual and monolingual dictionaries. Compared to a normal Italian-English bilingual dictionary, such as *Il Ragazzini2011*, a bilingual collocations dictionary can, first of all, provide much more information. While certainly being an efficient tool, *Il Ragazzini2011*, like any other bilingual dictionary, has to include all aspects of the Italian language. It cannot, evidently, devote as much space to collocations. In fact, it only provides the following English collocations for *smell* under the noun *odore*: *a good smell, a pleasant smell, a nice smell, a bad smell, a nasty smell* and an *offensive smell*. Despite offering a wide range of noun + verb collocations under the lemma *pagare*, with regards to the lexical pattern adverb + *pagare*, the dictionary only includes *pagare salato* and *pagare profumatamente*, both translated as *to pay dearly*. Clearly, bilingual dictionaries need to make a selection among the various collocations to include, whereas a dictionary of collocations can be much more exhaustive.

Of course, it could be argued that learners or translators, needing to find English collocations, can directly consult the monolingual tools already available. There is no doubt that monolingual English dictionaries of collocations are excellently compiled: once the user is aware of the English base, then he/she can consult such dictionaries to find collocates. It is also true, however, that this task in a monolingual dictionary takes longer and requires greater linguistic skills than it would do in a bilingual collocations dictionary. As seen above, the Italian-English bilingual collocations dictionary could more carefully account for meaning nuances in English, based on the Italian collocates. In finding the English collocations, *an acrid smell, a sour smell, a sharp smell, and a pungent smell*, an Italian learner or translator might be tempted to think that they are all equivalents. We have seen, however, that while the first three can be considered synonyms, the last, *a pungent smell*, means something a little different as its Italian equivalents endorse. The use of a bilingual collocations dictionary should, thus secondly, save the dictionary-user time in trying to individuate what collocation to choose, resulting in a more qualitative and effortless task. The combination of a fine lexicographic and linguistic analysis of Italian and English words will hopefully result in a bilingual reference work that will be quantitatively and qualitatively valid.

5 References

- Atkins, B. T. S. (1985). Monolingual and bilingual learners' dictionaries: A comparison. In R. Ilson (ed.), *Dictionaries, Lexicography and Language Learning*. Oxford: Pergamon Press and British Council, pp. 15-24.
- Atkins, B.T.S. (1994) A corpus-based dictionary. In *Oxford-Hachette English- French Dictionary* (Introductory section). Oxford: Oxford University Press: xix – xxxii.
- Atkins, B.T., Knowles, F.E. (1990). Interim report on the EURALEX/AILA research project into dictionary use. In I. Magay, J. Zigany (eds.), *Budalex 88 Proceedings*. Budapest, Akademiai Kiado, pp. 381-392.
- Baxter, J. (1980). The dictionary and vocabulary behavior: a single word or a handful? In *TESOL Quarterly* XIV, 3, pp. 325-336.

- Bensoussan, M., Sim, D. & Weiss, R. (1984). The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. In *Reading in a Foreign Language*, 2, pp. 262-276.
- Berti, B. (2010). The treatment of lexical collocations of six adjectives related to feelings in a sample of bilingual dictionaries English-Italian. In A. Dykstra, T. Schoonheim (eds.) *Proceedings of XIV Euralex International Congress, Leeuwarden, 6-10 July 2010*, Fryske Akademy, Leeuwarden. [contribution on CD-rom]
- Berti, B. (2012). Le collocazioni lessicali nel dizionario bilingue. Uno studio su tre dizionari italiano-inglese. In G. Gobber *Usi dei vocabolari nell'apprendimento delle lingue*, Milano: Vita e Pensiero, pp. 5-25.
- Berti, B. (forthcoming). 'This sounds odd to me'. Issues of collocational acceptability in a bilingual English-Italian dictionary". In *Proceedings of the conference Norma e uso nella lessicografia bilingue, 18-20 October 2012*, Ragusa.
- Berti, B., Pinnavaia, L. (2012). Towards a Corpus-driven Italian-English Dictionary of Collocations. In R. Facchinetti (ed.) *English Dictionaries as Cultural Mines*, Newcastle-Upon-Tyne: Cambridge Scholars Publishing, pp. 201-222.
- Cowie, A. (1999). *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press.
- Facchinetti (ed.) *English Dictionaries as Cultural Mines*, Cambridge Scholars Publishing, Newcastle-Upon-Tyne: 201-222.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- MacFarquhar, P. D., Jack C. R. (1983). On dictionaries and definitions. *RELC Journal*, 14, pp. 111-124.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. In *Applied Linguistics*. 24(2): pp. 223-242.
- Oxford Collocations Dictionary for Students of English* (2012). Oxford: Oxford University Press.
- Piotrowski, T. (1989). Monolingual and bilingual dictionaries: Fundamental differences. In Makhan L. Tickoo (ed.), *Learner's Dictionaries: State of the Art*. Singapore: Seamo Relc, pp. 72-83.
- Il Ragazzini 2011. Dizionario inglese-italiano, italiano-inglese*, Bologna: Zanichelli.
- Rundell, M. (1999). Dictionary use in production. In *International Journal of Lexicography*, 12, (1): pp. 35-54.
- Rundell, M. (2010). *Macmillan Collocations Dictionary for Learners of English*, Oxford: Macmillan Publishers Limited.
- Scholfield, P. J. (1999). Dictionary use in reception. In *International Journal of Lexicography* 12, (1): pp. 13-35.
- Sinclair, J. (2003). *Reading concordances. An introduction*. Harlow: Pearson Education Limited.
- Tiberii, P. (2012). *Dizionario delle Collocazioni. La Combinazione delle Parole in Italiano*, Bologna: Zanichelli.

Websites

- BNC: *British National Corpus*, accessible at <http://www.natcorp.ox.ac.uk>.
- WebCorp, accessible at <http://www.webcorp.org.uk/live/>.

