

A Corpus Based Description of English Lexical Collocations Used by Italian Advanced Learners

Aurelia Martelli

University of Torino

Dipartimento di Scienze del Linguaggio e Letterature Moderne Comparate

Via S.Ottavio 20

10124 Torino

Italy

Abstract

The present work is a pilot study of English lexical collocations in written productions by advanced Italian students of English. The aim is to describe errors made by learners in the production of word combinations. First a working definition of lexical collocation will be given in order to identify those word combinations that will be taken into consideration. Then, after introducing the data used and the methodology for its classification, collocational errors will be analysed from a quantitative and qualitative point of view by considering factors such as L1 influence, structural properties and degree of fixedness of wrong collocations. It will emerge that the L1 has a relevant role in the generation of wrong lexical collocations and that certain types of collocations are more error prone than others. Finally the implications of these observations are discussed with reference to the need for adequate treatment of collocations in lexicographic works.

1 Introduction

The term collocation is used to indicate the repeated co-occurrence of lexical items. The main assumption behind the idea of collocation is that some words attract each other thus forming naturally sounding combinations while other combinations, although acceptable from a syntactic and semantic point of view, are not considered acceptable and are not used. Native speakers of a language are usually aware of these limitations on the combinability of items and avoid producing what would be considered strange-sounding combinations. Collocational restrictions become particularly evident when non-native speakers, who do not possess interior knowledge of such limitations, produce word combinations which sound unnatural.

The study of collocational patterns has revealed itself useful in studies of second language vocabulary acquisition. Knowing a word means knowing the type of relations it contracts with other words, of both paradigmatic and syntagmatic nature. This ability is particularly important in the productive/encoding phase. Collocational errors derive from the fact that inadequate combinations of items are created, not misunderstood. A foreign learner who encounters what to him or her is a new collocation is usually able to derive its meaning and process it correctly, while it is more frequent that, in order to express a certain concept, a for-

foreign student will recur to lexical sequences that are not acceptable in the second language. Research in second language acquisition has recently focused on designing teaching materials that provide students with exhaustive information on the way in which a certain word may combine with other words. As far as lexicographic works are concerned, all the most recent editions of bilingual as well as monolingual learners' dictionaries include detailed information on the most frequent collocates of entry words and indicate limitations on the combinability of lexical items.

One of the most problematic aspects in dealing with collocations in a foreign language is selecting those combinations which should be the object of description and teaching. An important step to be taken in order to answer this question is to study the collocation mistakes made by learners, in order to identify the difficulties learners have in using collocations.

The present paper explores the problems of Italian advanced learners of English in the production of collocations. It presents a pilot study carried out on a small portion of the Italian component of the "International Corpus of Learner English, ICLE".¹ First a working definition of lexical collocation will be given in order to identify those word combinations that will be taken into consideration and to exclude all other lexical combinations (idiomatic word combinations, colligation patterns, phrasal verbs) which will not be part of the study. Secondly, the data analysed will be presented and the methodology for tagging and classifying collocational errors will be explained. Finally, the collocational errors retrieved from the corpus will be classified and described and some general observations will be made regarding the nature of collocation mistakes in learner English.

2 Defining collocations

It is difficult to provide a precise definition of the term "collocation". Within the field of phraseology terminology has been proliferating and in the last decades a number of terms have been introduced to describe all sorts of phraseological phenomena, including collocations.² The result is a set of similar terms that describe different types of word combinations (e.g. idioms, lexical phrases, prefabs and pragmatic formulae). In particular, many different definitions have been provided for "collocation", and it is clear that not all scholars use the term to designate the same phenomenon. As Bahns points out (1993:57) "regrettably, collocation is a term which is used and understood in many different ways".

The term lexical collocation is used in the present paper to indicate a combination of items belonging to open classes (this excludes all grammatical combinations) which seem to 'attract' each other and establish a relation of co-occurrence along the syntagmatic axis. A word combination will be considered a collocation if the items which constitute it are limited

¹ For general information on the ICLE project see Granger et al. (2002). For detailed information on the Italian component of the ICLE see Prat Zagrebelsky (2004).

² Among the most important studies in the area of phraseology see Yorio (1980), Alexander (1984), Cowie (1998), Nattinger and DeCarrico (1992), Lewis (1993), Fernando (1996), Moon (1997).

in their commutability, subject to syntactic variation (therefore not necessarily adjacent to each other) and semantically transparent (this excludes all idiomatic combinations).

3 Studying learner collocations

The analysis of collocational errors can provide interesting insights for the study of learner interlanguage as well as applications in the compilation of reference works for students, especially bilingual and monolingual learners' dictionaries. The present study attempts to contribute to the discussion of learner difficulties in the production of collocations by analysing collocational errors made by Italian advanced learners of English. Whereas previous research on non-native collocation production has relied almost exclusively on elicitation tests and translation exercises (Nesselhauf 2003), this study analyses authentic natural production data.

3.1 Data collection

The data analysed consists in 30 essays taken from the Italian component of "The International Corpus of Learner English" (henceforth ICLE-IT)³ for a total of about 20,000 words. Essays are argumentative in nature and have been written by university students in their final year of study. In order to ensure uniformity in the data chosen for the analysis, only essays having the following characteristics have been chosen: all essays have been written by students whose mother tongue is Italian; all essays have been written with the aid of reference tools; all essays have been written under exam conditions.

3.2 Classification procedure

All the essays were read and tagged manually for collocational errors. In the first phase of the tagging procedure all the collocations considered wrong or strange-sounding were added the tag [W]. This included collocations whose acceptability was in doubt. In the second phase of the procedure a second tag was added according to the part of speech involved in the generation of the incorrect combination. The following tagset was the result:

- [W][Adj] wrong choice of adjective for a noun, i.e. *animated discussion*.
- [W][V] wrong choice of verb for an object or subject noun, i.e. *obtain a goal*.
- [W][Adv] wrong choice of adverb for a verb, i.e. *openly manifest*.
- [W][Ex] the entire expression, realised through an "apparent" collocation is wrong, i.e. *drawing the threads*.

There were cases in which the boundary between strange-sounding collocations and incorrect collocations appeared subtle and determining the acceptability of a word combination proved rather problematic. In these cases two strategies were adopted to verify whether a

³ ICLE-IT consists in 392 essays for a total of about 227,000 words. Most essays (252) are produced in exam conditions, therefore with a time limitation. Reference tools have been used in 388 essays. The topics vary and are usually controversial issues such as gun ownership, artificial insemination, juvenile crime, crime punishment.

collocation should be tagged as wrong or not. First, native speakers were asked whether they considered a specific collocation acceptable or not, and in case of a negative answer they were asked to suggest a correct version.⁴ Secondly, word combinations whose acceptability was in doubt were checked in a large corpus of native English, the British National Corpus.

4 Results

In the thirty essays tagged for this pilot study the total number of collocational errors is 105. Of these 0 are adverb + verb collocations, 6 are collocations in which the entire expression is involved, 50 are adjective + noun collocations, 45 are verb + noun collocations. In the following sections the individual error categories will be described.

4.1 Adverb + verb collocations

Initially some adverb + verb combination (6) were found which were perceived as being unusual, for example *openly manifest*, *lawfully help*. However, both native speaker judgment and corpus evidence revealed that such combinations, although not very frequent and familiar sounding, cannot be considered unacceptable and erroneous. The result is that no wrong adverb + verb combinations are present in the 30 essays under examination.

This observation may be misleading and should not be misinterpreted. The absence of this type of error may be due to an avoidance strategy: students, feeling unsure about how to combine a verb with an adverb in an appropriate way, may choose to avoid expressing a certain concept though this type of word combination and use other linguistic strategies, for example prepositional phrases.

Thus, no relevant conclusion can be drawn about adverb + verb collocations in learner writing, but some suggestions can be given for further research. First of all, the entire corpus should be tagged for collocational errors to verify whether this collocational category is really less error prone than others. Secondly, the number of this type of error should be analysed also in relation to the different strategies for verb modification in order to verify whether students prefer to modify verbs by recurring to other linguistic devices.

4.2 Collocation errors involving whole expressions

This type of error occurs when a student "invents" a collocation which to him or her sounds plausible thus creating a word combination which does not exist in English. Substituting one of the two constituents therefore does not correct the combination which must be substituted by a completely different expression. Following are some examples (the underlined parts indicate the erroneous collocation): *The program is stopped for commercial informations...*; *In order to sell more copies or rise television shares...*; *Drawing the threads of this discussion together...*

⁴ The native speakers consulted were: Shan Hirst, Flo Ullmann, Chris Owen and Fiona MacWilliam, lecturers at the Faculty of Foreign Languages, University of Torino.

In the first example the correct version would be simply the word *commercials* or *advertisements*. In the second example the correct version would be *to increase television audience*. In the third example the expression *in conclusion* would be more appropriate. In all these examples the influence of the first language (Italian) is very strong. *Commercial informations* is a direct translation of the Italian *informazioni commerciali*. The combination *television shares* is directly taken from Italian in which the word *share* is a widely accepted and frequently used Anglicism to indicate the portion of audience that watches a certain TV show. *Drawing the threads* is an obvious literal translation of the Italian expression *tirare le fila*.

4.3 Adjective + noun collocations

There are 50 adjective + noun collocations considered wrong in the subcorpus analysed. Two important features emerge from the observation of the data. First of all, most erroneous adjective + noun collocations are lexical combinations which involve a medium degree of restriction. Following are some examples (the underlined items represent the wrong collocation): ...*a stable relationship can grant the right affectional environment ...*; *Juvenile crime has become a world problem...*; ...*unable to give his child an appropriate education ...*; *Another long debated reason is...*; ...*if there would be tight conditions for obtaining a gun licence...*

All the nouns in the examples above have a medium to low degree of restriction, meaning that they can combine with a relatively large range of adjectives. However, the adjectives chosen by the students do not belong to such range. The choice of the wrong adjective is probably caused by the fact that this type of restriction is arbitrary, i.e. not motivated semantically. The reason why the combination *tight conditions* is considered unacceptable while *strict conditions* is considered acceptable is not the result of the semantic features of the two adjectives, but of a somewhat arbitrary convention of the language. It seems therefore that a problematic area for students is represented not by fixed or almost fixed combinations (*auburn hair, avid reader*) but by those word combinations which involve a medium degree of restriction.

Another observation is that in many cases the mistake can be attributed to L1 influence. Following are some examples: *The most dramatic problems of nowadays society...*; ...*a maturation of the collective conscience...*; ...*if I personally analyse the nowaday situation...*; ...*has been regarded as a step toward an eventual disarmament...*; ...*the presence of gratuitous violence...*; ...*this gave rise to a quite animated discussion...* In generating these adjective+ noun collocations students have been influenced by the Italian expressions *società di oggi/odierna, coscienza collettiva, situazione odierna, l'eventuale disarmo, violenza gratuita, discussione animata*.

4.4 Verb + noun collocations

Word combinations belonging to this category are composed of a verb and a noun as either direct object complement or subject complement. There are 46 wrong verb+noun collocations in the subcorpus analysed. Similarly to adjective + noun collocations, many verb +

noun collocational errors involve combinations in which there is a limited degree of restriction, i.e. combinations in which the noun can be combined with a somewhat wide (although not unlimited) set of verbs. Following are some examples: *A 65 year old married woman undertook artificial insemination...*; *...cloning opens good opportunities for human beings...*; *...This power which is carried out by institutions...*; The correct choice of the verb to combine with the noun in the collocation would have been: *to undergo artificial insemination, to offer good opportunities, to exert power.*

Generally speaking, we can say that there are two trends identifiable in verb + noun collocations. First of all, most wrong verb + noun collocations belong to the category of creation/activation collocations (Benson et al. 1986). This term indicates collocations composed of a verb indicating a process of creation or activation and of a noun, for example *compose music, reach a verdict, set an alarm*. Following are some examples of errors in the production of creation/activation collocations: *...the government can create more severe punishment; Artificial insemination has created astonishment...*; *This situation has arisen many polemics...*; *...the use of guns provokes a high rate of death...*; The correct verbs to choose would have been *inflict punishment, generate astonishment, give life to a polemic, cause a high rate of death.*

The second observation regards the fact that the learner's L1 plays an important role in determining the production of wrong collocations. Following are some examples of collocational errors influenced by Italian: *Parents must take precautionary actions, following their children in an age...*; *They have whatever they want so they do not prove emotions...*; *...A solution to contain juvenile criminality...*; *...top models would like to conduct a more common life...*; *...the decision has to be weighted up...*; *...to do something else like to order the room...*; *...to help these women realize their hopes...*; *...social institutions can give successful results...*All these verb + noun combinations have been generated with the Italian counterpart clearly in mind: *seguire i bambini* (in the sense of "to look after children's growth"), *provare emozioni, condurre una vita, soppesare una decisione, ordinare una stanza, realizzare i propri sogni, dare dei buoni risultati.*

5 Conclusions

Two areas that seem particularly prone to errors are adjective + noun collocations and verb + noun collocations. In both cases the collocations more frequently misproduced are combinations involving a low degree of restriction. It seems therefore that learners are more aware of word combinations in which a verb or a noun can be combined with a limited set of items (fixed combination). This is probably due to the fact that, in virtue of their highly limited combinability, such collocations are signalled to students explicitly in teaching materials and during class work. Another feature which is common to both categories of error is that L1 influence plays an important role in generating wrong word combinations. It is interesting to compare these results with data regarding students of different mother tongue. For instance, research on verb + object noun collocations by German mother tongue learners has shown that, also in this case, the degree of restriction and the influence of the L1 play an important role in collocation production (Nesselhauf, 2003).

The present study is only an exploratory study, but some implications for language teaching and reference material can be discussed. Research on the difficulties of producing correct collocations in L2 needs to be carried out mainly in two directions. First of all, attention has to be paid to all those word combinations which show a low degree of restrictions and which are not perceived by students as fixed, pre-constructed phrases to be learned as single items. Thus, dictionaries for learners, both bilingual and monolingual, should list as many potential collocates as possible for each entry word. Secondly, the role of the L1 in producing word combinations needs to be acknowledged. In selecting which collocations to teach, it is necessary to consider collocations which are highly frequent in the L1 and which students will be tempted to translate literally thus generating unnatural collocations in the L2. Learners have to be made aware of L1-L2 differences in order to avoid producing collocations which are calques of L1 existing collocations.

References

- Alexander, R.J. (1984), 'Fixed Expressions in English: Reference Books and the Teacher' *English Language Teaching Journal* 38/2, pp. 127-134.
- Bahns, J. (1993), 'Lexical Collocations: Contrastive View', *English Language Teaching Journal* 47/1, pp. 56-63.
- Benson, M., Benson, E., Ilson, R. (1986), *A Lexicographic Description of English*, Philadelphia, John Benjamins.
- Cowie, A.P. (1998), *Phraseology: Theory, Analysis and Applications*, Oxford, Clarendon Press.
- Fernando, C. (1996), *Idioms and Idiomaticity*, Oxford, Oxford University Press.
- Granger, S., Dagneux, E., Meunier, F. (eds) (2002), *International Corpus of Learner English*, Université Catholique de Louvain, Presses Universitaires de Louvain (CD-ROM and handbook).
- Lewis, M. (1993), *The Lexical Approach*, Hove, Language Teaching Publications.
- Moon, R. (1997), 'Vocabulary Connections: Multi-Word Items in English', in Schmitt, N., McCarthy, M. (eds), *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge, Cambridge University Press.
- Nattinger, J., DeCarrico, J. (1992), *Lexical Phrases and Language Teaching*, Oxford, Oxford University Press.
- Nesselhauf, N. (2003), 'The Use of Collocations by Advanced Learners of English and Some Implications for Teaching', *Applied Linguistics* 24/2, pp. 223-242.
- Prat Zagrebelsky, M.T. (2004), *Computer Learner Corpora. Theoretical Issues and Empirical Case Studies of Italian Advanced EFL Learners's Interlanguage*, Alessandria, edizioni dell'Orso.
- Yorio, C.A. (1980), 'Conventionalised Language Forms and the Development of Communicative Competence', *TESOL Quarterly* 14/4, pp. 433-442.