



Exploring atypical verb+noun combinations in learner technical writing

MARÍA JOSÉ LUZÓN MARCO*
Universidad de Zaragoza

Received: 4 April 2011 / Accepted: 21 July 2011

ABSTRACT

Professional and academic discourse is characterised by a specific phraseology, which usually poses problems for students. This paper investigates atypical verb+noun collocations in a corpus of English technical writing of Spanish students. I focus on the type of verbs that most frequently occurred in these awkward or questionable combinations and attempt to explore the reasons why the learners deviate from NS's norms. The analysis indicates that these learners tend to have problems with a set of sub-technical and high-frequency verbs. Deviant combinations involving these verbs are frequently the result of a deficient knowledge of the phraseology of academic and technical discourse. The unawareness of collocations that are typical of this discourse often leads students to create V+N combinations by relying on the "Open Choice Principle" (Sinclair, 1991) or by using patterns from their mother tongue.

KEYWORDS: learner corpora, technical writing, collocation, sub-technical vocabulary, high-frequency verbs.

RESUMEN

El discurso profesional y académico se caracteriza por una fraseología específica, que suele plantear problemas a los estudiantes. Este artículo investiga colocaciones de verbo+nombre atípicas en un corpus de textos técnicos en inglés escritos por estudiantes españoles. El estudio se centra en los verbos que más frecuentemente aparecen en estas combinaciones atípicas y explora las razones por las que los estudiantes se desvían de la norma. El análisis indica que estos estudiantes suelen tener problemas con un grupo de verbos sub-técnicos y verbos de alta frecuencia. Las combinaciones atípicas en las que estos verbos aparecen son frecuentemente el resultado de un conocimiento deficiente de la fraseología del discurso académico y técnico. El desconocimiento de colocaciones que son típicas de este discurso a menudo lleva a los estudiantes a crear combinaciones basándose en el "principio de opción abierta" (Sinclair, 1991) o a usar colocaciones prestadas de su lengua materna.

PALABRAS CLAVE: corpus de aprendices, escritura técnica, colocación, vocabulario sub-técnico, verbos de alta frecuencia

**Address for correspondence:* Universidad de Zaragoza, Centro Politécnico Superior. María de Luna, 3, 50018 ZARAGOZA. Email: mjluzon@unizar.es

1. INTRODUCTION

Corpus-based studies of professional and academic discourse have revealed the existence of a highly conventionalised phraseology (e.g. Biber, Conrad & Cortés, 2004; Charles, 2006; Gledhill, 2000; Groom, 2005; Luzón 2000). However, although these studies provide useful information on the discursive features that students should eventually master, they are not enough to inform the design of ESP teaching materials and must be complemented with studies on learner corpora (Aston, 2000; Granger, 2002), which help to address interlanguage development and the relative difficulty of particular features to be taught.

Studies based on learner corpora have shown that collocation is an aspect of language problematic for L2 students. Research on academic writing by non-native students has revealed frequent errors involving the collocational patterning of words, phraseological infelicities and overreliance on a limited set of linguistic items (Flowerdew, 2000; Gilquin, Granger & Paquot, 2007). A number of studies have focused on the verb+noun miscollations in the writing of university students of English (e.g. Howarth, 1998; Nesselhauf, 2004; Zinkgraf, 2008). Part of this research is concerned with a pre-determined set of collocations. For instance, Nesselhauf (2004) analyses support verb construction involving the verbs *make*, *have*, *take* and *give* and Altenberg and Granger (2001) study collocations with the delexical verb *make*. Other studies take a broader approach and examine the phraseological features of complete texts produced by students. Howarth (1998), for instance, focused on the language of advanced EFL students from different mother tongues in the field of social sciences. Zingraf (2008) analysed non-standard V+N collocations in a corpus of texts written by Spanish speaking students (with a high-intermediate to advanced level of English) in an English language course in Teacher and Translation training programs.

In this paper I present the results of the analysis of a computerised corpus of technical English texts, written by Spanish Engineering students. The purpose was to identify and analyse atypical V+N combinations and to explore the reasons for these collocational infelicities. The analysis is intended to provide information that helps to improve the teaching of technical writing and to identify some collocational aspects that should be focused on when teaching writing to Engineering students.

2. BACKGROUND: COLLOCATIONS IN EXPERT AND LEARNER WRITING

Although there are various approaches to collocation (see Durrant & Schmitt, 2009: 159) I adopt here Hoey's definition of collocation as "the relationship that a lexical item has with items that appear with greater than random probability in its textual context" (Hoey, 2005: 3). Hoey goes on to claim that collocations indicate "a psychological association between words" (Hoey 2005: 5): words are mentally primed to occur with particular other words. The way we

use words is shaped by our learning experience, by the way we have seen it used in similar texts.

Corpus-based research of academic and professional discourse has revealed the pervasiveness of collocation in this type of discourse (Biber, 2006; Biber et al., 1999, 2004; Cortés, 2004; Coxhead, 2008; Gledhill, 2000, Luzón, 2000, Ward, 2007) and has thrown some light on the nature of academic phraseology. Some researchers have focused on extended collocations that frequently co-occur in a register, referred to as “lexical bundles” (Biber et al., 1999; Cortés, 2004) or “clusters” (Hyland, 2008), e.g. *as a result of, it has been noted that*. Research on these lexical bundles has shown that many of them are discipline bound, that their frequency and use varies across text types and that they are rarely used in student academic prose (Cortés, 2004; Hyland, 2008; Hyland & Tse, 2009). Similarly, research on collocations deriving from complex noun phrase formation (e.g. *reaction time, critical value, stable system*) has shown that these collocations are also highly discipline specific (Ward, 2007).

The competent use of phraseology has been considered an important part of fluent language use (Nattinger & DeCarrico 1992; Pawley & Syder 1983; Schmitt 2004; Wray 2002). In the field of EAP, it is generally agreed that lexical bundles are central for academic discourse (Cortés, 2004; Coxhead & Byrd, 2007; Hyland, 2008) and discipline-specific collocations have been presented as a “kind of threshold” to specialised disciplinary discourse at the undergraduate level (Ward, 2007). However, the increasing body of research on learner corpora (both spoken and written) has unveiled learners’ problems with L2 collocational use, e.g. learners’ overuse of the formulaic sequences they know well, underuse of more restricted collocations, collocational errors (e.g. De Cock, 2003; Cortés, 2004; Durrant & Schmitt, 2009; Granger, 1998; Howarth, 1998; Nesselhauf, 2004). Cortés (2004) found that in the few cases in which students used certain lexical bundles, their use was different from that by professional authors. The overuse of some items or word combinations is sometimes a result of the influence of the mother tongue, as shown by Granger (1998) in her study of phrases which function as macro-organisers with a pragmatic function. She discovered that French learners massively overused the frame “we/one/you can/cannot/may/could/might say that”.

Since there is general agreement that students need to acquire the high-frequency collocations used in their discipline, some researchers are currently engaged in the creation of listing of such collocations in academic English (Durrant, 2009; Simpson & Ellis, 2010), comparable with the Academic Word List (Coxhead, 2000). The Academic Formulas List (AFL), created by Simpson and Ellis (2010), for instance, includes “formulaic sequences identified as (i) frequent recurrent patterns in corpora of written and spoken language, which (ii) occur significantly more often in academic than in non-academic discourse, and (iii) inhabit a wide range of academic genres”. Since some researchers suggest, however, that collocations and lexical phrases are discipline related (Hyland & Tse, 2009), it seems logical to focus students’ attention on the collocations used in their discipline.

3. THE STUDY

3.1. Corpus

The learner corpus for this research was made up of 80 student assignments of approximately 1,900-2,100 words each, totalling 160,613 words. 35 texts were written by Computer Engineering students, 29 by Chemical Engineering students and 16 by Industrial Engineering students. Students wrote their assignments as response to a real-like task, where they were asked to produce a text reporting their research. Although most learner corpora are compiled from the writing of high-intermediate level students, this corpus was made up by the texts written by all the students taking the course “Technical English” in Computer Engineering, Industrial Engineering and Chemical Engineering (high-intermediate and low-intermediate level), since the main purpose of the research was to identify the errors that our students (no matter what level) made. As the corpus was compiled from course assignments, all the texts were read and corrected manually, which made it easy to get a preliminary idea of the problems that students had with V+N collocations.

3.2. Method

The corpus was analysed with the Worldlist and Concordance tools of the program *WordSmith Tools*. The Sketch Engine System (Kilgariff et al., 2008) was also used to explore the British National Corpus (*BNC*) and the *BAWE* (British Academic Written English Corpus) for comparison purposes. The first step of the analysis involved producing a wordlist, ordered by frequency. The list was analysed manually in order to identify verbs that could occur in the structure verb+noun (direct object)¹. In the second step I resorted to concordancing and qualitative analysis of the different verbs to find information on collocations. Concordances of individual verb lemmas were computed and each concordance line was scrutinised to identify instances of possible inappropriate collocations. The analysis of verb concordances revealed that while students had little problem with the nominal complements of most verbs, there was a set of verbs whose collocational behaviour posed problems. All the instances of possible inappropriate collocations were recorded with the number of occurrences.

The following step was to analyse the acceptability of such collocations. For that purpose, I first consulted three dictionaries, i.e. the *Collins Cobuild English Dictionary* (Sinclair et al., 1987/1995), the *Oxford Collocations Dictionary* (McIntosh et al., 2009) and the *BBI Dictionary of English Word Combinations* (Benson et al., 1997). However, it is difficult to assess the acceptability of a specific collocation (unless it is a restricted collocation) using the information in dictionaries, and even more difficult to assess the acceptability of a collocation in a specific register. The following step was to resort to evidence from the *British National Corpus (BNC)* and *BAWE (British Academic Written English Corpus)*. I tried to determine the acceptability of a combination in technical writing

by checking its frequency and salience in the sub-corpus of the *BNC* dealing with applied, pure and physical science- henceforth referred to as *BNC(Sc)*²- and in the sub-corpus of the *BAWE* dealing with Physical Science- henceforth referred to as *BAWE(PS)*³. The Sketch Engine was used for this stage. This system can provide a word sketch of the word. A word sketch is an automatic one-page corpus-derived summary of a word's grammatical and collocational behaviour: it “provides one list of collocates for each grammatical relation the word participates in” (Kilgariff et al., 2008: 298). Following this procedure I got a list of atypical verb+noun combinations.

Although there were some clear cases of miscollocations, especially those involving restricted collocations with a delexical verb, acceptability seems to be a matter of degree, since there are combinations that are not frequent and sound non-native, but we cannot categorically state that they are unacceptable. Even native speakers disagree about the acceptability of certain collocations, as Nesselhauf (2005) reported. Therefore, I decided not to produce a list of miscollocations, but rather focus on the verbs that most frequently pose problems and are involved in non-standard or questionable V+N combinations, i.e. combinations that do not occur in the *BNC(Sc)* and the *BAWE(PS)* or which occur with a very low frequency.

4. RESULTS AND DISCUSSION

Table 1 provides an alphabetically ordered list of the verbs that most frequently occur in non-standard V+N combinations in the learner corpus. I will use these verbs to analyse and discuss students' problem with V+N collocations in technical writing.

| |
|----------|
| Achieve |
| Cause |
| Do |
| Explain |
| Generate |
| Give |
| Have |
| Produce |
| Make |

Table 1. Verbs that most frequently occur in non-standard V+N

In order to discuss the type of verbs that pose problems for students, it is useful to consider the different types of vocabulary in academic discourse. Coxhead and Nation (2001) divide vocabulary in academic texts into four categories: high frequency words, academic vocabulary, technical vocabulary, and low frequency words. High frequency words, the most frequent 2,000 words of English, account for around 80% of the running words of academic texts (Nation, 2001). A few of these high frequency verbs can be used in delexical structures, i.e. structures whose meaning derives from the words and phrases that co-occur with the verb

(e.g. *make noise*). Although the meaning of a delexical structure is transparent, there are strict restrictions on the range of nouns which can combine with specific delexical verbs (e.g. to *make noise*, but not * to *make a reaction*). Therefore, in this case it is relatively easy to assess collocational acceptability.

Academic vocabulary, which covers about 8.5% of academic text, has also been referred to as sub-technical vocabulary (Baker, 1988; Cowan, 1974). Several definitions have been provided for this concept, most of them including the idea that it is vocabulary common to several scientific disciplines. Baker (1988: 91) defines sub-technical vocabulary as “a whole range of items that are neither highly technical and specific to a certain field of knowledge nor obviously general in the sense of being everyday words which are not used in a distinctive way in specialised texts”. Baker (1988) brings together all the previous definitions of sub-technical vocabulary in a list with six categories. Some of the verbs which are problematic in the learner corpus belong to Baker’s categories (v) and (vi): (v) “General language items which are used, in preference to other semantically equivalent items, to describe or comment on technical processes and functions”, e.g. *take place* instead of *happen*; (vi) “Items which are used in specialised texts to perform specific rhetorical functions”, e.g. “It has been *pointed out* by...” (Baker, 1988: 92). The verbs *achieve*, *cause*, *explain*, *generate* and *produce* in Table 1 belong to the category of words defined by Baker as sub-technical vocabulary. It should be pointed out, however, that *cause* and *produce* are not included by Coxhead in the AWL but in the GSL (General Service List). This list includes words that are in the most frequent 2,000 words of English, but they are not further included in the AWL.

As we can see from Table 1, the verbs that most often occur in non-standard V+N combinations in the learner corpus are delexical (*do*, *give*, *have*, *make*) and sub-technical verbs (*achieve*, *cause*, *explain*, *generate*, *produce*), while there seems to be little problem with technical verbs. Howarth (1998), who also found that technical senses posed the least problems for the students in the corpus he analysed, suggests that this could be result of a greater degree of lexicalisation and therefore familiarity of collocations. However, he also points out that technical verbs are by far the least frequent, which could also help to explain the lower frequency of atypical combinations involving these verbs. It should also be noted that most high-frequency and sub-technical verbs occurring in the structure verb+noun (direct object) were correctly used as well. For instance, students did not make mistakes with verbs such as *reduce*, *increase*, *decrease*, *provide*, *avoid*, *need*, *use*, etc. which are frequently used in the corpus. This makes it even more interesting to study why the verbs in Table 1 pose problems.

4.1. Subtechnical vocabulary used in atypical combinations

4.1.1. Cause

Most subtechnical verbs that occurred in awkward collocations in the learner technical corpus are used to express cause/ effect relationships, which supports Flowerdew's (1998) finding that students have problems with cause/ effect markers.

The main problem with the verb *cause* derives from the fact that students are not aware that this verb has a negative semantic prosody (i.e. negative connotations associated with a particular word). Stubbs (1995) found that more than 90% of the words that collocate with *cause* are negative (e.g. *crisis, accident, delay, death, damage, trouble*). Interestingly, Wei's (2002) analysis of *cause* in a corpus of academic English texts reveals that this item has a stronger negative prosody in academic English texts than in general English texts, thus suggesting that semantic prosody has specific features in specialised texts. The highly negative semantic prosody of *cause* is confirmed by its word sketch from the *BNC(Sc)* and the *BAWE(PS)*. All the most frequent and salient collocates as direct object of the verb *cause* are negative: *BNC(Sc)* (*damage, pollution, disease, problem, cancer, symptom, increase, difficulty, confusion, reduction, death, loss, concern*), *BAWE(PS)* (*problems, error, change, damage, decrease, noise, concern, pressure, interference, reduction, drag, failure, reaction, harm, distortion, negligence, resistance, instability*). Although in the learner corpus we do find *cause* in combination with negative words (e.g. *problems, noise, impact, pollution, accidents, erosion, damage, confusion, diseases, reduction*), it also occurs with positive words (e.g. *cause an improvement in mobility, cause profits*). Since *cause* is a very frequent verb in the learner corpus, the analysis of its collocates helps to reveal clearly that students are unaware of its semantic prosody. This is not the only verb whose semantic prosody poses problems for students. For instance, the verb *suffer* occurs in two cases with positive nouns (e.g. *suffer an improvement*).

4.1.2. Produce

In other cases, the awkwardness derives from the fact that some verbs are overused in combinations where other verbs would be more frequent or more appropriate. An example is the verb *produce*. The verb (lemma) *produce* occurs 1,414 times in the *BAWE(PS)* (0.1% of the words in the corpus), 7,454 in the *BNC(Sc)* (0.06% of the words in the corpus), and 142 in the learner corpus (0.08% of the words in the corpus). Just by looking at these figures, it would seem that the verb *produce* is in fact used less frequently in the learner corpus than in the proficient student corpus (*BAWE*). However, if we look at the collocates we can get a more accurate picture.

The most frequent collocates with the word *produce* in the learner corpus are *energy* (15), *gas* (9) and *electricity* (8). The word *produce* is also very often used in the corpus with

negative words, mainly to refer to negative consequences, e.g. *emissions, noise, carbon dioxide, spillages, pollution, damages, difficulty, problems, impact, vibrations, waste*.

- (1) The construction of the tram will *produce* more *noise and vibration* than ...
- (2) The warming can *produce* an *explosion or failures* in the supply

However, in the *BNC(Sc)* there are no negative words among its most frequent and most salient object collocates. Evidence from the *OCD* and from the *BNC(Sc)* suggests that *produce* collocates with some of these negative nouns (e.g. *emissions, noise, pollution*), although in some cases the frequency of the collocation in the *BNC(Sc)* is very low. For instance, the frequency and salience of the collocation *produce pollution* are 5/1.65 (frequency/salience), while the frequency and salience of *cause pollution* are 48/5.22. Similarly, the frequency and salience of *produce damage* are 4/1.26, while those of *cause damage* are 201/7.22; the frequency and salience of *produce noise* are 2/0.29, while those of *make noise* are 45/2.1 and those of *generate noise* are 4/3.4. This suggests that even if the combinations *produce pollution, produce damage* or *produce noise* do occur in the *BNC(Sc)*, they are uncommon. In addition, *produce* does not occur in the *BNC(Sc)* corpus in combinations such as *produce impact, produce problem, produce explosion, or produce difficulty*. When comparing the sketch of *produce* and *cause* in the *BNC(Sc)* *problem, difficulty* and *explosion* always occur with *cause*, and the word *pollution* occurs much more frequently with *cause* than with *produce*. *Produce* is therefore used in V+N combinations where other verbs would be chosen by native speakers, e.g. *produce problem* instead of *pose/cause/present problems; produce impact* instead of *have/make an impact*. Interestingly, when looking at the word sketch of *producir* in the Spanish Web Corpus (a corpus also available in the Sketch Engine), there are many negative words among its most salient object collocates (e.g. *alteración, daño, lesión, escalofrío, explosión, enfrentamiento, pérdida, quemadura, disminución, adición*), which suggests that the awkward collocations of *produce* are in part due to the influence of the mother tongue.

But this is not the whole story. In the learner corpus we do not find collocations which are common in science writing- the *BNC(Sc)*- or in science writing by students- the *BAWE(PS)*. In the *BNC(Sc)* and the *BAWE(PS)* *produce* collocates frequently with *results, model, graph, output, pattern, effects*. In the learner corpus, there are only 3 occurrence of *produce effect*, 1 occurrence of *produce output*, and there are no occurrences of *produce* with the other nouns.

The use of the verb *produce* by Spanish speakers seems to support Warren's (2005) claim that non-native speakers are likely to construct a generalised meaning of an L2 word by equating it with some core meaning in L1, i.e. a translation equivalent, while native speakers construct generalised meaning of words by abstracting semantic commonalities from the uses of the word. Knowing a word also involves knowing its collocates and, as Hoey (2005) remarks, the way we use words is shaped by the way we have seen them used in similar texts.

Therefore, it seems likely that, given that non-native learners have had less exposure to target texts from which to learn how to use words, they tend to transfer their collocational knowledge of the L1 word to the L2 word. Warren (2005: 41) also points out that in addition to constructing a generalised meaning, the native speaker memorises some “more or less fixed phrases which represent language-specific uses”, e.g. in the case of transitive *drop*, combinations such as *drop bombs*, and therefore these phrases and not only generalised meanings should be the object of study.

4.1.3. Generate

Another common sub-technical verb in the learner corpus which posed problems for students is *generate*. This verb, included by Coxhead (2000) in the AWL, has a general meaning but it is very frequently used in Engineering and Environmental Science. In the learner corpus it tends to collocate with words referring to sources of energy, but also with words referring to different types or aspects of pollution (e.g. *emissions, wastewater, residues, pollution, sewage*). This collocation pattern is supported by the *OCD*. In the *BAWE (PS)*, however, *generate* collocates with *power, profit, results, models, output, noise, revenue, values, energy, electricity, heat, reaction*, and in the *BNC(Sc)* it collocates with *electricity, pulse, heat, signal, revenue*. Although there are few negative words among the most salient and most frequent collocates of *generate* in these corpora, it should be taken into account that the learner corpus included many texts on environmental issues and it could be logical to expect this kind of combinations. Since Hyland and Tse (2009: 111) have shown that individual lexical items on the AWL “often occur and behave in different ways across disciplines and that words commonly contribute to “lexical bundles” which also reflect disciplinary preferences”, the difference in results could be due to the specific collocational behaviour of *generate* in some Engineering disciplines.

Although *generate* may collocate with negative words referring to different aspects of pollution, it certainly does not collocate with other negative words like *risk*. Information both from the *OCD* and from the word sketch of *risk* in the *BNC(Sc)* shows that the verbs that most frequently collocate with *risk* in this sense are *pose, involve* and *present*.

4.2.4. Achieve

The *Cobuild Dictionary* provides the following definition of the verb *achieve*: “If you achieve a particular aim or effect, you succeed in doing it or in causing it to happen, usually after a lot of effort”. In the *BNC(Sc)* and the *BAWE(PS)* sub-corpora the most frequent object collocates are: *objective, goal, aim, result, balance, success, product, reduction, feat, target, fusion, performance, improvement, equilibrium, independence, accuracy, efficiency*. In the learner corpus *achieve* collocates with some of these words, e.g. *objective* (4 occurrences), *speed* (4), *accuracy* (3), *goal* (2), *target* (2), *results* (1), *solution* (1), *aim* (1). But it also collocates with some words that do not collocate with *achieve* in the other corpora: *achieve requirements/*

expectations/ demands (e.g. achieve the demands of the Kyoto protocol) instead of *meet requirements/ expectations/ demands*; *achieve an agreement* instead of *reach an agreement*; *achieve offers/ discount* instead of *get offers/ discounts*.

A comparison of the word sketch of *achieve* and *reach* in the *BNC* shows that, although they have a few common object collocates, e.g. *target, consensus, level, standard*, the word *agreement* collocates with *reach* but not with *achieve*. Similarly, although *achieve* and *meet* have a few common collocates (fewer than *achieve* and *reach*), such as *target* or *standard*, the words *requirement* and *demand* only collocate with *meet*. The verb *meet* seems to be underused in the learner corpus. There are only 5 occurrence of the lemma *meet* in the learner corpus: the object in these cases are *needs* (2 occurrences), *requirements* (1), *condition* (1), *limitations* (1) (see example 3). This is a low frequency if we consider that some of the essays were recommendation reports where student had to evaluate whether some products/ techniques, etc. met specific criteria.

- (3) The *limitations* we will *meet* due to the nature of these communications will be: delays in communication...(instead of “The limitations we will face...”)

The reasons for the low frequency of *meet* with objects such as *requirement, demand*, etc. could be that in most of the students’ encounters with the verb *meet*, the verb occurred in the structure *meet + people*, with the sense “come together with at the same place and time” and many of the students do not seem to be familiar with the sense “fulfil or satisfy (a requirement)”, where the verb collocates with a different set of nouns.

An interesting feature of the collocational behaviour of the verb *achieve*, which is revealed from its word sketch, is that it collocates frequently with signalling nouns (Flowerdew, 2006) indicating goal or target (e.g. *objectives, result, target*). A “signalling noun” is defined by Flowerdew (2006: 345) as “any abstract noun the full meaning of which can only be made specific by reference to its context”. In his analysis of signalling nouns in a corpus of argumentative essays written by L1 learners of English, he found that these nouns are problematic for learners. The second most frequent category of errors by learners was the incorrect choice of signalling noun. The problems with signalling nouns reported in Flowerdew’s study help to explain some miscollocations involving the verb *achieve*, as in the examples below:

- (4) This reduction made it easier to increase the running and walking speed of the robot.
This is an *aspect* that Honda tried to *achieve* ...
- (5) Banning the disposal of industrial residues. If this *point* is *achieved* ...

In example (4) the realisation (underlined fragment) cannot be labelled as “aspect”, as the writer has done. A more appropriate signalling noun, which would collocate with *achieve*, might be something like *goal* or *target*. Similarly, in example (5), *point* is not an appropriate

word to refer to “banning the disposal of industrial residues”. In this case, example (5) is part of a list that the writer has previously labelled as “recommendations”. Therefore, “if this point is achieved” could be replaced with something like “if this recommendation is followed”. The first problem here is that students find it difficult to know which label (signalling noun) to use for specific stretches of text. This is also reflected in examples where the signalling noun is omitted, which is one of the errors pointed out by Flowerdew (2006). In example (6), although the anaphoric pronoun “this” is used, the meaning would be clearer with a signalling noun like *objective* or *goal* (e.g. *achieve this objective*):

- (6) Nowadays a big part of research is focused on making life easier for people with disabilities. One way to achieve this is through the development of brain controlled devices.

An additional problem is that students are unaware that *point* and *aspect* do not collocate with the verb *achieve*. The most salient verb collocates of *point* and *aspect* in the BNC(Sc) are as follows: *point* (*illustrate, reach, locate, emphasise, indicate, miss*), *aspect* (*emphasise, quantify, neglect, clarify, cover, investigate, highlight, illustrate, discuss, refine, explore*). Both nouns tend to collocate with discourse and research verbs, but not with *achieve* or verbs in the same semantic class.

4.1.5. Explain

Another type of verbs that are problematic for students are those used to refer to discourse acts within the paper. The verb *explain*, for instance, is used in V+N combinations where other verbs would be more suitable (*explain a method, explain a system, explain a protocol, explain an algorithm, explain an experiment*). The fragments below provide further examples:

- (7) a. Explain the main treatment of water
 b. Thanks to the sensors that we will explain later
 c. We are going to explain our robot...
 d. There are a lot of reasons for explaining the importance of water...

When comparing the word sketch of *describe* and *explain* in the BNC(Sc), words such as *method* or *approach* are mostly used with *describe*. The combination *describe method* occurs 111 times in the BNC(Sc) with a salience of 6.42. Other discourse verbs that collocate with *method* in this sub-corpus are *illustrate, outline, discuss, propose, evaluate, compare*, but not *explain*. Similarly, the words *experiment, algorithm, object* are used as objects with *describe* and not with *explain*. As for the noun *importance*, it occurs in the corpus BNC(Sc) in the structure V+N with the verbs *emphasise, stress, highlight, underline, acknowledge*. The misuse of *explain* seems to be related to its overuse in the corpus when compared with other

discourse verbs. In the learner corpus there are 110 occurrences of *explain*, 37 of *describe*, 24 of *discuss*, 19 of *propose* and 5 of *illustrate*. There are no occurrences of collocations like *emphasise/ stress/ highlight the importance*. Just by comparing the frequency of *explain* and *describe* in the learner corpus, the *BNC(Sc)* and the *BAWE(PS)*, we can see that while *explain* is much more frequent than *describe* in the learner corpus, it is less frequent than *describe* in the *BNC(Sc)* and they display roughly the same frequency in the *BAWE(PS)*.

| | Learner corpus | BNC(Sc) | BAWE(PS) |
|-----------------|-----------------------|----------------|-----------------|
| explain | 0.06% | 0.02% | 0.03% |
| describe | 0.02% | 0.04% | 0.03% |

Table 2. Frequency of *explain* and *describe* in the learner corpus, the *BNC(Sc)* and the *BAWE(PS)*

These results suggest that learners have difficulties in choosing the right discourse verbs. Since V+N combinations involving discourse verbs play important discourse functions in expert technical and academic writing, they should be paid attention to when teaching this type of writing.

4.2. Delexical verbs

Since many collocations typical of the academic register involve delexical verbs (Biber et al., 1999: 1027-29), students need to know how these verbs are used in academic discourse. However, several studies have revealed that students have problems with delexical constructions (Altenberg & Granger, 2001; Barfield, 2003; Howarth, 1998; Nesselhauf, 2004; Shirato & Stapleton, 2007; Zingraf, 2008). According to Sinclair (1991: 147), “many learners avoid the common verbs, especially when they occur in idiomatic phrases. Instead of using them, they rely on larger, rarer, and clumsier words making their language sound stilted and awkward”. The learner corpus includes several examples of wrong delexical constructions, e.g.

- *make a task, make a process, make a search, make work, make an operation, make disinfection, make an experiment, make energy, make a function*
- *do changes, do functions, do a movement, do a study, do an analysis*
- *give a discussion*

The delexical verbs most frequently involved in miscollocations in the learner corpus are *make* and *do*. While in some cases one of these verbs is used instead of the other (e.g. *do changes* instead of *make changes*), probably due to the fact that learners find it difficult to differentiate between these two verbs, which are rendered in Spanish by a single word, i.e. *hacer*, in most cases it would be more appropriate to use a sub-technical verb. In the *BNC(Sc)* the most salient verb collocates of some of the nouns that in the learner corpus miscollocate

with *make* or *do* are verbs like *perform*, *complete* or *conduct*: *task* (*perform*, *accomplish*, *complete*), *search* (*perform*, *conduct*, *complete*), *process* (*complete*, *develop*), *work* (*undertake*, *carry*, *complete*), *operation* (*perform*, *conduct*, *complete*, *carry out*), *function* (*perform*, *fulfil*), *study* (*carry out*, *perform*, *conduct*, *make*, *undertake*, *complete*), *analysis* (*perform*, *conduct*, *undertake*), *experiment* (*perform*, *conduct*, *carry out*, *undertake*). In the *BNC(Sc)* *energy* also tends to collocate as an object with verbs like *generate* (33/6.34) and *produce* (61/5.2), rather than with *make*.

These data could suggest that students resort to delexical verbs when they are not sure about the verbs that collocate with a specific noun. This is in agreement with Zingraf's (2008: 108) statement that in students' eyes delexical verbs seem to be accompanied by an unrestricted number of nouns: "Under the assumption that these verbs lack a specific meaning, learners over-generalise and combine them with any noun under the illusion that there is no restriction to the way they can be used". The results also provide support for Gilquin's (2007) claim that the problem with collocations (in her study, *make* collocations) is not only the errors that students make, but also that they underuse some collocations and that they limit themselves to those collocations they are more sure about. In our learner corpus there is a clear underuse of collocations involving verbs like *complete* or *perform*.

I also found that students tend to use a delexical verb plus a nominalisation in cases where native speakers would tend to use a full lexical verb, a tendency that was also revealed by other studies of learner writing (Barfield, 2003; Juknevičienė, 2008), e.g. *make installation and maintenance* vs. *install and maintain*, *make an investment* vs. *invest*, *make an identification* vs. *identify*). Sometimes the form delexical verb+nominalisation replaces a discourse verb (e.g. 10).

- (8) The disinfection is made to eliminate any microorganism vs. X is disinfected to eliminate microorganism.
- (9) Access to the database can be done by using the user account vs. the database can be accessed by...
- (10) Before making any evaluation of the data, we should do some explanations

Again, as Zingraf (2008) points out, the use of delexical structures instead of lexical verbs can be considered as evidence of the students' perceptions of delexical verbs as having unrestricted collocations. These results provide support for other studies which have shown that if the learners' knowledge of English collocations is incomplete they construct meaning by relying on the "open choice principle", e.g. by adding up meaning of individual words, rather than on the "idiom principle" (Zingraf, 2008; Juknevičienė, 2008).

The high frequency lexical verb *have* also occurs in collocations that either would not be used by native speakers or are infrequent, giving rise to awkward expressions. In the

examples below, a deficient knowledge of English leads learners to produce incorrect collocations:

- (11) Current laws don't *have a restriction* about..
- (12) Fossil fuel which *have other problems* like the shortage and variations in prices

In example (11) *have* is used instead of verbs that frequently collocate with *restrictions*, such as *impose* or *place*. In example (12), although *have* can be easily used with *problems*, that use would need an animate subject. In this case *pose problems* or *present problems* would be the right collocations to use. However, the learner is probably unfamiliar with these collocations.

5. DISCUSSION AND CONCLUSION

The analysis of V+N combinations in the corpus of technical writing by Spanish speakers has shown that students have problems with a set of high frequency and sub-technical verbs, while technical verbs and other high-frequency and sub-technical verbs (e.g. *take*, *increase*, *decrease*, *avoid*) seem to pose little problem.

Previous studies have revealed that deviant V+N combinations are sometimes the result of transfer from the mother tongue (e.g. Fan, 2009; Zinkgraf 2008) and that collocational use is adversely affected by a deficient knowledge of L2 grammar and lexis (Fan, 2009). The analysis carried out here confirms these results. For instance, as we have seen, some of the deviant collocations involving *produce* are translations from collocations in Spanish. The verbs in table 1 are sometimes used in combination where native speakers would more likely use other verbs (e.g. *produce problems* vs. *cause problems*, *achieve demands* vs. *meet demands*, *make tasks* vs. *perform tasks*). Thus, many of the awkward or deviant collocations seem to be used by Spanish learners because they do not know the right collocation. This is especially the case with the delexical verbs *do* and *make*, whose strict collocational restrictions seem to be ignored by some learners.

The study reveals that students are unaware of collocations that are typical of technical writing, e.g. *perform a task*, *achieve an objective*, *discuss a point*, and which often have very specific rhetorical functions. This suggests that students' knowledge of the sub-technical vocabulary of a discipline may be incomplete: it is mainly based on a generalised meaning of the words or on equating the word with an L2 equivalent, but they lack collocational knowledge of the words, and, in general, they are unaware of the phraseology of academic and technical discourse. This research has also confirmed previous studies which show that inappropriate word choice by ESL/EFL students sometimes derives from an unawareness of semantic prosodies (Wang & Wang, 2005; Wei, 2006).

The results of this study have several implications for the teaching of technical writing. Since many of the awkward combinations derive from a deficient knowledge of academic phraseology and, more specifically, of the phraseology of the students' discipline, this research lends support to recent studies on academic vocabulary which argue for the need to provide students with a repertoire of academic lexical phrases and to teach these phrases together with their rhetorical or organisational function (Granger, 2011; Simpson-Vlach & Ellis, 2010). Verbs typically used in academic discourse should not be taught on their own, but together with their collocational patterns. A feature of academic discourse that seems to be especially difficult for students is the combination of verbs with signalling nouns, partly due to the complex nature of signalling nouns as cohesive devices. However, since they have such important cohesive functions in academic discourse, these combinations should be explicitly taught. In addition, when teaching items of academic vocabulary with a clear semantic prosody, awkward combinations can only be avoided if attention is focused not only on the denotational meaning of the word, but also on its semantic prosody.

Finally, I would like to point out a limitation of the study, which affects the assessment of the acceptability of some combinations. This limitation derives from the difficulty to identify native texts that are equivalent in type and discipline to the learner corpus, and that can therefore be used to compare the presence and frequency of specific collocations in both corpora. To minimise this limitation I have used two sub-corpora that could help to provide information on how language is used in expert and native student writing in science and technical disciplines. However, there still could be examples where the overuse of some collocations in the learner corpus when compared to the other corpora could be due to the more restricted topics of the students' writing (e.g. *generate emissions*).

ACKNOWLEDGEMENTS

The work reported in this paper was funded by the University of Zaragoza (Project Code: UZ2007-HUM-03).

NOTES

1. Within this structure I considered not only V+(Det)+(Adj)+Noun, but also any other grammatical realisations of the pattern, such as passive voice.
2. *BNC(Sc)* includes applied science, natural and pure science and has 12,312,723 words
3. The *BAWE* corpus (6.5 million words) contains 2,761 proficient student assignments, most of them (1,953) written by L1 speakers of English. The Physical Science subcorpus includes Engineering, Chemistry, Computer Science, Physics, Mathematics, Meteorology, Cybernetics & Electronics, Planning, Architecture and has 1,381,356 words.

REFERENCES

- Altenberg, B. & Granger, S. (2001). The gramatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics*, 22:2, 173-194.
- Aston, G. (2000). Corpora and language teaching. In L. Burnard, & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 7-17). Hamburg: Peter Lang.
- Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, 4(2), 91-105.
- Barfield, A. (2003). *Collocation recognition and production: Research insights*. Tokyo: Chuo University.
- Benson, M. & Ilson, R. (1997). *The BBI Dictionary of English Word Combinations*. John Benjamins Pub. Co., Amsterdam, Philadelphia.
- Biber, D. (2004). Lexical bundles in academic speech and writing. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language corpora (PALC 2003)* (pp. 165-178). Hamburg: Peter Lang.
- Biber, D. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S., & Cortés, V. (2004). *If you look at ...: Lexical bundles in university teaching and textbooks*. *Applied Linguistics*, 25(3), 71-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (Eds.). (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310–331.
- Cortés, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), 389-400.
- Coxhead, A. (2000). A new academic wordlist. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A. (2008). Phraseology and English for academic purposes. In F. Meunier & S. Granger (Eds.), *Phraseology in language learning and teaching* (pp. 149-161). Amsterdam: John Benjamins.
- Coxhead, A. & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129-147.
- Coxhead, A. & Nation, P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (pp. 257-267). Cambridge: Cambridge University Press.

- De Cock (2003). Recurrent sequences of words in native speaker and advanced learner spoken and written English: A corpus-driven approach. Doctoral dissertation, Université catholique de Louvain, Louvain-la-Neuve, Belgium, Unpublished.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for Academic Purposes. *Journal of English for Specific Purposes*, 28(3), 157-169.
- Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, 47(2), 157-177.
- Fan, M. 2009. An exploratory study of collocational use by ESL students – A task based approach. *System*, 37(1), 110-123.
- Flowerdew, J. (2006). Use of signalling nouns in a learner corpus. *International Journal of Corpus Linguistics*, 11(3), 345-362.
- Flowerdew, L. (1998). Integrating expert and interlanguage computer corpora findings on causality: Discoveries for teachers and students. *ESP Journal*, 17(4), 329-345.
- Flowerdew, L. (2000). Investigating referential and pragmatic errors in a learner corpus. In L. Burnard, & T. McEnery (Eds), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 145-154). Frankfurt am Main: Peter Lang.
- Gilquin, G. (2007). To err is not all: What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift Fur Anglistik Und Amerikanistik*, 55(3), 273-291.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: the missing link in EAP pedagogy. In P. Thompson (Ed.), *Corpus-based EAP Pedagogy*. Special issue of *Journal of English for Academic Purposes*, 6(4), 319-335.
- Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, 19, 115-135.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam and Philadelphia: Benjamins.
- Granger, S. (2011). From phraseology to pedagogy: Challenges and prospects. In T. Herbst, P. Uhrig & S. Schller (Eds.), *Chunks in the Description of Language. A tribute to John Sinclair* (pp. 123-146). Berlin & New York: Mouton de Gruyter.
- Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, 4(3), 257-277.
- Hoey, M. (2005). *Lexical Priming: A new theory of words and language*. London: Routledge.
- Howarth, P. (1988). The phraseology of learners' academic writing. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.

- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, K. & P. Tse. (2009). Academic lexis and disciplinary practice: corpus evidence for specificity. *International Journal of English Studies*, 9(2), 111-129.
- Juknevičienė, R. (2008). Collocations with high-frequency verbs in learner English: Lithuanian learners vs. native speakers. *Kalbotyra*, 59(3), 119-127.
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2008). The Sketch Engine. In T. Fontenelle (Ed). *Practical Lexicography: a reader* (pp. 297-306). Oxford: Oxford University Press.
- Luzón, M.J. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, 19(1), 63-86.
- McIntosh, C., Francis, C. & Poole, R. (2009). *The Oxford Collocations Dictionary*. Oxford University Press, Oxford.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nattinger, J. & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford University Press: Oxford.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125-152). Amsterdam: John Benjamins.
- Pawley, A. & Syder, F.H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (Eds.), *Language and Communication* (pp. 191-225). Longman: London.
- Schmitt, N. (Ed.) (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam & Philadelphia: John Benjamins.
- Shirato, J. & Stapleton, P. (2007). Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan. *Language Teaching Research*, 11(4), 393-412.
- Simpson-Vlach, R. & Ellis, N.C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J, Hanks, P., Fox, G., Moon, R., & Stock, P. (Eds.) (1987 1st ed.) (1995). *Collins Cobuild English Dictionary*. London: HarperCollins Publisher.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 23-55.
- Wang, H., & Wang, T. (2005). A contrastive study on the semantic prosody of CAUSE. *Modern Foreign Language*, 28(3), 297-307.
- Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, 6(1), 18-35.

- Warren, B. (2005). A Model of Idiomaticity. *Nordic Journal of English Studies*, 4(1), 35-54.
- Wei, N.X. (2002). A corpus-driven study of semantic prosodies in specialized texts. *Modern Foreign Language*, 25(2), 165-175.
- Wei, N.X. (2006). A corpus-based contrastive study of semantic prosodies in learner English. *Foreign Language Research*, 132, 50-54.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zinkgräf, M. (2008). V+N miscollocations in the written production of university level students. *ELIA (Estudios de Linguística Inglesa Aplicada)*, 8, 91-116.