



**HAL**  
open science

# A corpus-based study of the phraseological behaviour of abstract nouns in medical English

Natàlia Judith Laso Martín

► **To cite this version:**

Natàlia Judith Laso Martín. A corpus-based study of the phraseological behaviour of abstract nouns in medical English. 5th Corpus Linguistics (CL'09), Jul 2009, Liverpool, United Kingdom. Paper #319. hal-01108375

**HAL Id: hal-01108375**

**<https://hal.science/hal-01108375>**

Submitted on 15 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **A corpus-based study of the phraseological behaviour of abstract nouns in medical English**

*Natalia Judith Laso Martín*

University of Barcelona

*njlaso@ub.edu*

## **Abstract**

It has been long acknowledged (Carter 1998; Williams 1998; Biber 2006; Hyland 2008) that writing a text not only entails the accurate selection of correct terms and grammatical constructions but also a good command of appropriate lexical combinations and phraseological expressions. This assumption becomes especially apparent in scientific discourse, where a precise expression of ideas and description of results is expected. Several scholars (Gledhill 2000; Flowerdew 2003; Hyland 2008) have pointed to the importance of mastering the prototypical formulaic patterns of scientific discourse so as to produce phraseologically competent scientific texts.

Research on specific-domain phraseology has demonstrated that acquiring the appropriate phraseological knowledge (i.e. mastering the prototypical lexico-grammatical patterns in which multiword units occur) is particularly difficult for non-native speakers, who must gain control of the conventions of native-like discourse (Howarth 1996/1998; Wray 1999; Oakey 2002; Williams 2005; Granger & Meunier 2008).

This paper aims to analyse native speakers' usage of abstract nouns in medical English, which will contribute to the linguistic characterisation of the discourse of medical science. More precisely, this research study intends to explore native speakers' prototypical lexico-grammatical patterns around abstract nouns. This analysis is based entirely on corpus evidence, since all collocational patterns discussed have been extracted from the *Health Science Corpus (HSC)*, which consists of a 4 million word collection of health science (i.e. medicine, biomedicine, biology and biochemistry) texts, specifically compiled for the current research study. The exploration of the collocational behaviour of abstract nouns in medical English will serve as a benchmark against which to measure non-native speakers' production.

## **Introduction**

Corpus-based studies have drawn the attention to the study of the lexicon as the central principle in language and have also emphasised the interconnections between lexis and syntax (Francis 1993; Hunston & Francis 2000; Wray 2002, among others). Linguistic investigation of naturally-occurring data has revealed that language is organised in terms of a lexico-grammar and, thus, it consists of recurrent patterns of words (Renouf & Sinclair 1991; Sinclair 1991; Altenberg & Tapper 1998; Stubbs 2001). The study of how words are used to make meanings; in other words, how meaning maps onto use, is one of the key concerns in current research in phraseology.

Phraseological empirical studies have confirmed the important role of prefabricated expressions in the textual development of meaning (Gledhill 2000b; Kaszubski 2000) and have also highlighted the need for further research on the phraseological conventions characteristic of specialist genres. As Kaszubski (2000) points out:

Word combinations are inextricably related to the layer of style –the appropriateness and/or naturalness of selection and co-occurrence of items, subject to genre-sensitive restrictions and conventions. Thus, in order to compare aspects of lexical use, one is bound to focus attention on phraseology.” (Kaszubski 2000:2)

Specifically, in the field of Language for Specific Purposes (LSP), corpus linguistics has established itself as a fundamental methodological tool for determining the defining linguistic features of different discourse communities (Swales 1990; Lee 2001; Gledhill 2000; Lee & Swales 2006).

This research study intends to explore native speakers' prototypical lexico-grammatical patterns around abstract nouns, illustrated by a case study on the phraseological behaviour of the noun *contribution* in medical research articles.

### **Phraseology in specialised registers**

[T]he picture we have of the research article is far from complete. That picture suggests that there are certain characteristics of RAs [research articles] which, by and large, tend to occur and recur in samples drawn from an extensive range of disciplines (...) However, it remains the case that RAs [research articles] are rarely simple narratives of investigations. Instead, they are complexly distanced reconstructions of research activities, at least part of this reconstructive process deriving from a need to anticipate and discountenance negative reactions to the knowledge claims being made." (Swales 1990:174-175)

Swales' account of the genre of the scientific article points to the existence of a number of conventions which contribute to define and characterise the scientific discourse. Textual analyses of the genre are hence extremely important so as to identify the phraseological structures characteristic of scientific English.

Genre analysis is concerned with a particular subset of language; that is, a specific language practice, characterised by a number of linguistic features and phraseological conventions. It can be therefore claimed that genres make use of different ways of expressing meaning (Hunston 2002:178). This assumption is intimately linked with the concept of *local grammar* (Gross 1993; Barnbrook & Sinclair 1995; Hunston & Sinclair 2000), which consists of a description of particular areas of language (e.g. the analysis of the phraseology characteristic of medical discourse), rather than the language as a whole (Bednarek 2007).

The current treatment of phraseology in specialised registers acknowledges the need for corpus-based studies of the prototypical lexico-grammatical patternings and discourse functions of lexical phrases across disciplines (cf. Carter 1998; Oakey 2002a/b; Hyland 2008)<sup>1</sup>. According to Hyland (2008),

Gaining control of a new language or register requires a sensitivity to expert users' preferences for certain sequences of words over others that might seem equally possible. (Hyland 2008:5)

Thus, it seems that getting familiar with the specific phraseology of the register of a discourse community will imply not only a better knowledge of the genre but also an enhanced competence in the process of writing and reading in specialised registers. As Williams (2002) claims:

In order to understand texts, we must look at them closely to find the lexico-grammatical strategies that they adopt to assist communication within a specialised community. (Williams 2002:60)

Studies in genre analysis, such as Swales' (1990) investigation of academic and research settings; Bhatia's (1993) exploration of genre in professional contexts and Gledhill's (2000b) work on collocations in science writing, suggest that there are significant textual variations in different specific-domains and genres<sup>2</sup>. These findings have underlined the convenience of analysing specialist corpora so as to find out the "*kinds of language data which particular communities of users might encounter and which will inform their use.*" (Hyland 2008:8).

Despite the abovementioned growing interest in the formulaic aspects of language knowledge in specialised registers, Gledhill (2000b) observes that in comparison with linguistic analyses based on general English corpora, less work has been conducted on specialised language to date. More specifically, he claims that there is a noticeable shortage of linguistic corpus-based studies in the field of phraseology in scientific discourse.

There are, however, some remarkable exceptions. Several worth mentioning studies focusing on scientific articles as a whole are, for example, Myers' (1989) account of the pragmatics of politeness involved in scientific papers; Master's<sup>3</sup> (1991) study of active verbs with inanimate subjects in scientific English and Banks' (1994) analysis of the organisation of different clause types in the scientific journal article.

Other studies on the phraseology characteristic of scientific discourse, on the contrary, have centred their investigation either on a specific domain within the field of scientific English or, to a lesser extent, on the different (sub)sections of the scientific article. (Adams-Smith 1984; Salager-Meyer 1994; Gledhill 1995/1996 and Williams 1996, to name but a few).

All the above literature has proven extremely useful in the characterisation of science writing, and has also confirmed Swales' (1990) assertion of the complexity of the scientific research article: "*the RA [research article] is anything but a simple genre*" (Swales 1990:128). Several textual properties of the scientific discourse such as modality (Huddleston 1971; Widdowson 1979; Adams-Smith 1984; Salager-Meyer 1992; Banks 1994; Gledhill 2000a/b), hedging (Myers 1989; Swales 1990; Salager-Meyer 1994; Banks 1994; Varttala 1999; Gledhill 2000), the use of the passive and the anticipatory *it*-pattern so as to disguise authorial interpretations (Huddleston 1971; Swales 1990; Banks 1994; Biber et al. 1998/1999; Hyland 2008), an attested tendency to use grammatical metaphor (Salager-Meyer 1992; Banks 1994; Halliday 1998; Gledhill 2000a/b) and a high use of abstract nouns in the expression of processes and methods (Halliday 1993; Flowerdew 2003) have been identified as defining rhetorical devices which contribute to a great extent to the development of scientific discourse (Luzón 2000; Gledhill 2000b; Noguchi 2006; Hyland 2008).

If, as already discussed in the reported literature, much of the language involved in scientific discourse is "*highly stereotypical in nature*" (Gledhill 2000a:116), it seems of paramount importance that members of that discourse community become familiar with the collocational expressions considered to be "good scientific style", since conforming to those conventions will provide scientists with the phraseological competence necessary for effective and accurate communication.

### **Methodology (Corpus data)**

The main corpus analysed in this study is the *Health Science Corpus (HSC)*, which is a representative sample of texts specifically assembled for the current investigation of the use of abstract nouns by the health science community. The data corpus was compiled as part of a research project, 'Creation of a Database of Lexical Combinations in Scientific English', financed by the Spanish Ministry of Science and Education and FEDER (BFF2001-2988). Within this project, co-ordinated by Dr. Isabel Verdager (University of Barcelona), the GReLiC<sup>4</sup> research team at the University of Barcelona has developed *SciE-Lex*, a dictionary that provides contextual information on usage as well as the combinatorial potential of words commonly used in scientific registers.

Interested as we were and still are in the lexico-grammatical patterns of non-technical terms in scientific English and the conventionalized phraseological characteristics of that genre, and bearing in mind that there was no corpus of scientific English publicly available<sup>5</sup>, the GReLiC research group decided to compile their own micro-corpus, now consisting of approximately 4 million words of scientific research articles from prestige online journals<sup>6</sup> that cover different disciplines such as medicine, biology, biochemistry and biomedicine.<sup>7</sup>

Computerised corpora and linguistic software tools are essential for linguistic data management. The processing of a corpus by means of computerised methods has proven to be a very useful tool for the researcher to process in real time large quantities of texts, which had been otherwise completely impossible. There are several software programmes at the lexicographer's disposal to store a corpus electronically. The software used in retrieving data from the *HSC* and refining the results further was version 3.0 of the concordancing program *WordSmith Tools*<sup>8</sup>. This program provided a list of words, which allowed us to find out what general terms are most frequently used in scientific English. Taking such a list as the starting point, the abstract noun *contribution* was selected for the present research paper.

Among the various tools available, *WordSmith* was extremely useful not only to identify collocates and their frequency, but also to classify such collocates in terms of grammatical position, word class, semantic category, etc., to analyse word clusters so as to see the patterns of repeated phraseology in the concordance lines analysed and to make generalisations from the observation of repeated language events.

As pointed out above, one of the main aims when compiling the *HSC* was trying to make a representative selection of naturally-occurring language in a very specific type of genre, the health science discourse, so as to analyse the collocations and syntagmatic structures associated with abstract nouns in that particular register. To this respect, it seems worth-recalling the notion of "local grammar" (Barnbrook & Sinclair 1995; Hunston & Sinclair 2000; Hunston 2002), which refers to descriptions of particular areas of language (rather than the language as a whole): "*the connection between pattern and meaning opens the possibility of quantifying ways of expressing meanings in different registers via the concept of 'local grammar.'*" (Hunston 2002:178).

The compilation of the *HSC* thus understood as "*an authoritative body of linguistic evidence which can support generalizations and against which hypotheses can be tested*" (Sinclair 1987:2) has facilitated the exploration of the phraseological behaviour of abstract nouns in medical English, with respect to patterning by means of corpus evidence. For many research and pedagogical purposes, I do believe that the larger the corpus is, the more reliable conclusions can be drawn from the careful examination of the language shown. Nevertheless, it must be stressed that I am fully aware of the fact that the *HSC* constitutes a sample<sup>9</sup>, a cross-section of the health science discourse<sup>10</sup>, so claims will be just based on the results obtained from the in-depth analysis of the *HSC* data. As Partington (1998) observes: "*a corpus, no matter how large and varied, is only representative of itself and claims made about the behaviour of linguistic items after studying corpus data should bear this in mind*" (Partington 1998:146). Following Partington's consideration, Hunston (2002) notes that all observations made from a particular collection of texts "*must be dealt with as deductions rather than as facts*". (Hunston 2002:23)

### **The use of abstract nouns in medical English. Some preliminary considerations.**

It must be noted that the decision to analyse abstract nouns was based on the findings from a study<sup>11</sup> conducted on the behaviour of the noun *conclusion* and its restricted collocations in scientific register. While focusing on that abstract noun, it became apparent that there was a frequent list of comparable nouns, etymologically related to a verb (*conclusion* ~ *conclude*;

*agreement* ~ *agree*; *comparison* ~ *compare*; *contribution* ~ *contribute* and *decision* ~ *decide*, to name but a few), which needed more thorough investigation.

Abstract nouns in combination with other parts of speech are frequently used in academic register to refer to scientific processes, methods, evidence and findings. A detailed study of the patterns of abstract nouns in the *HSC* has revealed a vast amount of phraseological units whose overall meaning is the result of the interaction among its various elements. As will be shown later, some of the collocates these frequent abstract nouns co-occur with have undergone a process of ‘delexicalisation’ and ‘semantic bleaching’<sup>12</sup>, which means that these nouns mostly provide the semantic content to the whole unit. In most cases, such abstract nouns collocate with delexicalised or support verbs which contribute very little to the meaning of the whole unit. It is thus the larger unit that is the complete unit of meaning, rather than the individual lexical items.

This phenomenon has already been pointed out in innumerable corpus studies which highlight the importance of multiword units in linguistic production (Sinclair 1991; Gledhill 2000; Altenberg & Granger 2001, Oakey 2002; Stubbs 2001; Simpson 2004). In line with Sinclair’s idea that meanings are clustered into lexico-grammatical patternings and not in isolated terms, the following section focuses on the recurrent sequences of words<sup>13</sup> in combination with the abstract noun *contribution*, selected for the present study, that commonly co-occur in the health science discourse.

In Partington’s words, the process of ‘delexicalisation’ involves “*the reduction of the independent lexical content of a word, or group of words, so that it comes to fulfil a particular function but it has no meaning apart from this to contribute to the phrase in which it occurs.*” (Partington 1993:183). In this respect, expressions such as *make a conclusion*, *reach an agreement*, *make a comparison*, *make a contribution* and *make a decision* must be analysed as multiword units, where the abstract noun provides the semantic content to the extended lexical string of words to the detriment of the semantic content of the lexical verb. Such restricted verbs have adopted a more grammatical role and simply perform a verbal function.

As Partington (1993) observes, the notion of ‘delexicalisation’ is closely related to Sinclair’s concept of ‘shared meaning’, “*a distribution of meaning across a number of words*” (Sinclair 1987b:110), which accounts for the fact that single words and their context of appearance are mutually co-selected: “*words in English do not normally constitute independent selections (...)* *The item and the environment are ultimately not separable.*” (Sinclair 1992:15).

According to Sinclair (1997:323), ‘semantic depletion’ is common with high frequency words, which tend to lose their independent meaning and adopt a more grammatical role. The analysis of **V + Noun** periphrastic structures in the corpus data reinforces this statement. In this view, the boundaries between a lexical item and its environment become fuzzy. Evidence from the *HSC* has revealed that delexical uses of verbs co-occurring with abstract nouns are usual. The conclusion to be drawn from this fact is that the more delexicalised a unit is, the more widely it collocates (Partington 1993:183). All these lexico-grammatical issues will be illustrated in the pages to follow by means of a case study on the combinatorial patterns in which the noun *contribution* occurs.

### **The collocational patterning of the noun *contribution* in the *HSC***

This section describes the collocational patterning of the noun *contribution* in the *HSC*, which will be followed by a discussion of the overall results, with a view to exploring the particular phraseology of the noun under study in the health science discourse. Unlike other abstract nouns, such as *conclusion*, *agreement* and *decision*, the noun *contribution* shows fewer entries (249), whereas its corresponding verb, *contribute (to)* [1071 occurrences] is more overtly used in the *HSC*. The examination of the morphology of the noun *contribution* indicates that this is a countable noun used to refer to an abstraction. In Biber’s et al. (1999) view,

“countability is not a simple reflection of things observed in the external world (...) with reference to discrete concrete objects, but also to abstractions which do not so obviously or naturally come as distinct entities.” (1999:242)

Taking into account that an obvious feature of countable nouns is their variation in number, the abstract noun *contribution* must be regarded as a fully countable entity. From a grammatical perspective, *contribution* is characterised by number variation (i.e. it inflects for the plural) and its co-occurrence with determiners, mainly central determiners: definite and indefinite articles, demonstrative and possessive determiners. However, it should be mentioned that this noun shows a preference for its base form (166 occurrences) over its inflected counterpart, *contributions* (83 occurrences). The next subsections will focus on the different patterns this noun collocates with.

#### a) verb + *contribution* (to)

Amongst the various verbs combining with this abstract noun, there is a limited group of verbs that stand out as being semantically equivalent to *contribute*. This group of restricted collocates consists of the following verbs: *make*, *provide* and *produce*. Consider some examples<sup>14</sup>:

- (1a) Bruce Weir **has made** many **important contributions to** population genetic inference theory.
- (1b) Bruce Weir **contributed to** population genetic inference theory.
  
- (2a) This book should **make a significant contribution to** the reemergence of the field.
- (2b) This book should **contribute significantly to** the reemergence of the field.
  
- (3a) Embryos were organized like cellular jigsaw puzzles, each cell of which was prespecified to **produce its own precisely delimited contribution to** the mosaic that was the developing organism.
- (3b) Embryos were organized like cellular jigsaw puzzles, each cell of which was prespecified to **contribute to** the mosaic that was the developing organism.
  
- (4a) It **provides only marginal contributions to** binding as judged by inhibition studies.
- (4b) It **contributes marginally to** binding as judged by inhibition studies.
  
- (5a) These multivitamins did not appear to **provide significant contributions to** the parameters stated.
- (5b) These multivitamins did not appear to **contribute significantly to** the parameters stated.

As can be seen in the examples above, the periphrastic structures of the type **V+ *contribution*** equal the verb *contribute*, given the fact that they convey the same meaning. In this respect, it is particularly relevant the fact that *make*, *provide* and *produce* have undergone a process of ‘delexicalisation’, by which they have gradually lost their primary sense of “making, creating something” in favour of the meaning provided by the abstract noun they collocate with.

From a semantic point of view, there arise two key questions which this analysis attempts to answer: what do the periphrastic structures *make a contribution to*, *provide a contribution to*, *produce a contribution to* mean? Are they as polysemic as their equivalent, *contribute to*? The answers to these two questions are of central importance as one of the main fundamentals lying behind ‘pattern grammar’ is that the meaning of one of the items in a collocation is tied to its co-occurrence with the other item. Thus, for instance, the meaning of *make* in *make a contribution to* is semantically constrained by the collocation it occurs with.

Likewise, the full verb *contribute to* is highly polysemic in general English. *WordNet*<sup>15</sup> identifies four different senses of such a verb: 1) bestow a quality on; 2) provide money, time, knowledge, assistance, etc. along with others to a common supply, fund, etc.; 3) be conducive to and 4) contribute to some cause –for instance, furnish works for publication.

On the contrary, the occurrences of both *contribute to* and “**V + contribution to**” found in the *HSC* show a narrower field of use. All the examples refer to a more figurative sense of *contribute ~ contribution* rather than being associated with money, time and the like. A possible paraphrase of these units in the *HSC* could be “play a significant part / help cause something”. With this sense, the restricted collocates of the type **V + contribution to** and its related cognate verb *contribute to* are semantically interchangeable. This fact goes reasonably well with the phenomenon of ‘delexicalisation’ observed in the verbs *make*, *provide* and *produce*, since the meaning of “playing a significant part” is mainly conveyed by the abstract noun *contribution* in these periphrastic structures and by the verb *contribute to* in synthetic uses.

Moving now on to syntax, there are a few points that should be considered as well. It has been long acknowledged that in scientific writing, writers often make use of the passive voice in order to avoid the recurrent repetition of personal references (i.e., *I / my / me; we / our / us*) and to make the text look more impersonal, neutral and objective (Swales 1990; Biber et al. 1998/1999; Hyland 2008). This syntactic feature (**form**) contributes to a great extent to placing the emphasis of a given message on processes and experimental procedures (**meaning**), which is a widely used device in academic writing.

The use of *make a contribution*, however, differs from what has been stated above. With a total of 25 occurrences in the corpus, there has only been found one instance of this sequence in the passive:

- (6) First, mediastinal tissue analysis of AIDS patients from autopsy revealed that five of seven (71%) patients had either no thymus or no areas of thymopoiesis, demonstrating that **no contribution to the peripheral T-cell pool was being made by the thymus** (...)

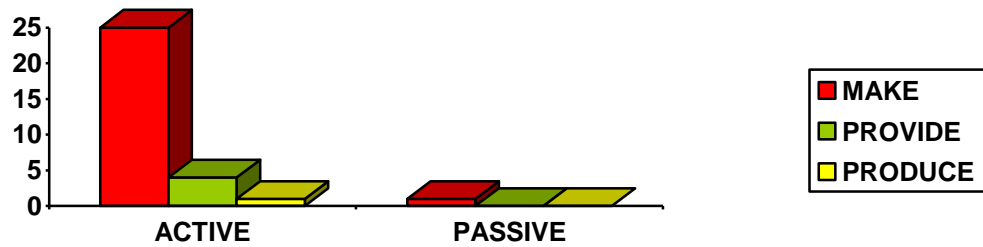
Such a preference for active constructions seems to be semantically motivated. The sense of this pattern, which could be paraphrased as “X plays an important part in Y” requires an explicit specification of the ‘agent’ (i.e., the maker / causer of the action described); it can, by no means, demote the Subject because it represents an important focus of attention. Below are some examples that illustrate the gist of this argument:

- (7) **This amplification makes too small a contribution to the total amount of lac DNA** to be detected by (...)
- (8) The challenge for the future will be to determine not simply that such altered **cell biology** could have an effect but that **their effects** are large enough to **make a significant contribution to age-related changes**.
- (9) **Additional transcripts of abundance class D make the largest and decisive contribution to the colon-cancer phenotype**.

Last but not least, notice that in the only example of *make a contribution* (see example 6 above) in the passive found in the corpus (see Figure 1 below), the prepositional phrase (agent ‘by’-PP) of passive sentences has not been omitted (“[...] *by the thymus*”) Again this underlines the fact that this typically optional element in passive sentences is considered to be relevant in this case.



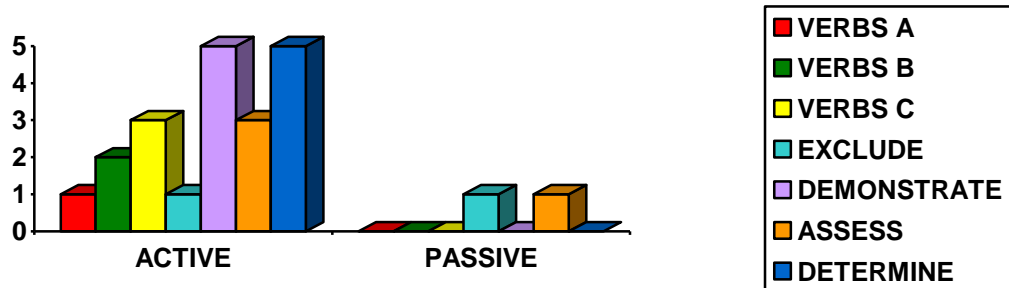
### RESTRICTED COLLOCATES



**Figure 1** Active and passive constructions with the pattern **restricted verb + contribution**

In sharp contrast, there is a wide group of verbs that combine with *contribution* but do not equal semantically the verb *contribute*. Most of these free collocates only appear once in the *HSC* (see Figure 2 and Table 1 below for frequency rates), but they are used to convey a really wide range of meanings.

### FREE COLLOCATES



**Figure 2** Active and passive constructions with the pattern **free verb + contribution**

VERBS A (1 occurrence)	VERBS B (2 occurrences)	VERBS C (3 occurrences)
CONSIDER	SUGGEST	EXAMINE
ELIMINATE	INDICATE	EVALUATE
SHED LIGHT ON	MODEL	INVESTIGATE
SUSTAIN	ESTIMATE	
REFLECT	DISSECT (OUT)	
SCORE	ENABLE	
QUANTIFY	LIMIT	
APPRAISE	TEST	
OVERLOOK		
DISCERN		
DISCUSS		
ELUCIDATE		
DEFINE		
REVEAL		
ACKNOWLEDGE		
DISENTANGLE		
MINIMIZE		

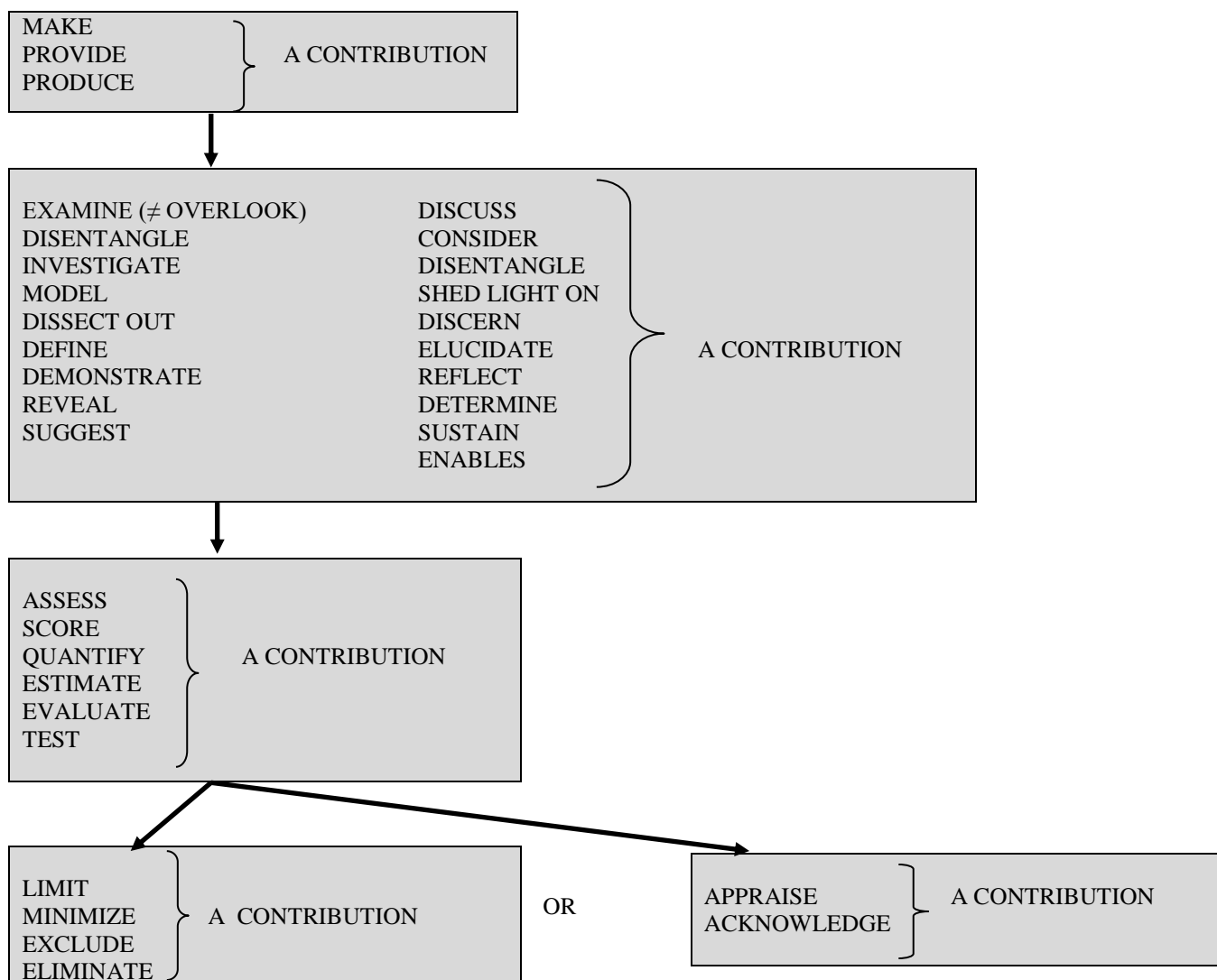
**Table 1** Frequency rates of the pattern **free V + contribution**

Especially noticeable is the use of active structures (i.e. only the verbs *exclude* and *assess* are used once in the passive voice) as well as a common preference for verbs connected with scientific procedures: *assess, demonstrate, determine, examine, evaluate, investigate*, etc.

For the present study, these free collocates have been grouped into different semantic fields, consisting of comparable verbs that convey similar meanings and, consequently, can be encompassed under the same category. From the very beginning, all these semantic fields seem to follow a logical line of thought that can be summed up as follows:

- 1) X **makes** a contribution to Y.
  - 2) Z **examines** that contribution of / to something.
  - 3) Z **evaluates** that contribution of / to something.
- 
- 4) Z **excludes** that contribution of / to something.
  - 5) Z **appraises** that contribution of / to something.

Such a sequence of actions can be expressed by a variety of verbs and described as part of a logical sequence of events, as illustrated in Figure 3:



**Figure 3** Restricted and free collocates of the pattern **V+ contribution** grouped into semantic fields and described as part of a logical sequence of events

**b) adjective + contribution**

There is a wide range of attributive adjectives occurring to the left of the abstract noun *contribution*. Although both **descriptors** and **classifiers**<sup>16</sup> can be found as modifiers of *contribution*, **descriptors** are outnumbered by **relational** and **topical classifiers** (see Table 2 for frequency counts of modifiers of this noun). As for the former, there are adjectives covering the semantic domains of **size**, **quantity** and **extent** (*small, minor, minimal, heavy, lower, massive*), **time** (*new, early*) and **evaluation** (*significant, important, outstanding, biased, decisive, favorable, functional, sympathetic*). The following Table shows the most frequent adjectives in combination with the noun *contribution* in the *HSC* corpus:

ADJECTIVES + <i>contribution</i>					
DESCRIPTORS				CLASSIFIERS	
<b><u>EVALUATIVE</u></b>	<b><u>SIZE</u></b>	<b><u>EXTENT</u></b>	<b><u>TIME</u></b>	<b><u>RELATIONAL</u></b>	<b><u>TOPICAL</u></b>
SIGNIFICANT (10)	MAJOR (4)	MINIMAL	NEW	RELATIVE (23)	CALORIC (2)
IMPORTANT (7)	MINOR (3)	HEAVY	EARLY	FUNCTIONAL (4)	TECHNICAL
SUBSTANTIAL (3)	SMALL (2)	LOWER		MATERNAL (4)	BIOLOGICAL
OUTSTANDING	MASSIVE			PARTICULAR (3)	PHYSIOLOGIC
BIASED				FRACTIONAL (2)	PLACODAL
FAVORABLE				INDIVIDUAL (2)	CELLULAR
SYMPATHETIC				SPECIFIC (2)	NUTRIENT
				UNEQUAL (2)	GENETIC
				UNIQUE	
				OVERALL	
				ADDITIONAL	
				DELIMITED	
				DETECTABLE	
				DIFFERENTIAL	
				EXCLUSIVE	
				INDEPENDENT	
				INTRINSIC	
				QUANTITATIVE	
				SECONDARY	
				SIMILAR	
				SEPARATE	

**Table 2** Frequency rates of **descriptors** and **classifiers** + *contribution*

Although descriptive adjectives do not usually collocate with the noun *contribution*, it should be highlighted that the list of descriptors seems to be limited to the constraints of the genre in question. Scientific writing appears to rely more on **relational/classifying** (*relative, functional, maternal*); **affiliative** (*African*) and **topical adjectives** (*technical, physiologic, caloric, cellular, nutrient, genetic*).

This phenomenon is what could be referred to as “stylistic preference”; the most common adjectives in this type of genre are **classifiers** because academic writing is concerned with delimiting, defining, classifying and focusing on demonstrable data rather than on making judgements or personal evaluations, which tend to be more common of fiction and literary writing.

It is wellknown that many adjectives by a process of derivational affixation become adverbs by suffixing *-ly* to the base form of an adjective. This universal truth in grammar plays an important part when analysing the most frequent adjectives in combination with *contribution*: *significant, substantial, important*. Examining the environment of the verb *contribute* in the *HSC* shows that it usually collocates with adverbs derived from the most common adjectives combining with *contribution*: *significantly, substantially, importantly*.

The following concordance lines extracted from *WordSmith Tools* show enlightening examples for *significant contribution* and *contribute significantly* in context:

```
WordSmith Tools -- 07/05/2008 17:42:26

133 et al., 1989). To evaluate the relative contribution of embryonic gene
134 T. Menes, and E. Hanski (1997) Relative contributions of
135 es and an assessment of their relative contributions. The success of
136 "selective adhesion" makes a secondary contribution: "In consequence
137 lexly interrelated that their separate contributions cannot be
138 1984] clearly demonstrate a significant contribution of ancestral
139 oa. Given the acknowledged significant contribution of 'male factor'
140 e. This book should make a significant contribution to the reemergence
141 70/85 kDa S6 kinases make significant contributions to events
142 cells but could make a more significant contribution to assays of
143 ment for proteinuria made a significant contribution to the model, but
144 are large enough to make a significant contribution to age-related
145 e. This book should make a significant contribution to the reemergence
146 did not appear to provide significant contributions to the parameters
147 athione is likely to make a significant contribution to the mechanisms
148 1992; Braun et al., 1995). A similar contribution to the
149 crossover loop makes a relatively small contribution to the energetics
150 rons on the O3 atom also to make some contribution to the
151 impossible to dissect out the specific contributions of individual
152 c al proteins was due to non-specific contributions of the linker or
153 rs at SCI, with a possible substantial contribution from the
154 ipogenesis did not make a substantial contribution of fatty acids
155 s thus unlikely to make a substantial contribution to rejection of
156 ng of dense SS RBC (13). Despite such contributions, trans-species
157 romoting Th2 cell growth and survival. Contributions from non-CD41
158 gh ratios would indicate a sympathetic contribution. A similar approach
```

**Figure 4** Concordance lines of the node word *contribution*

```
WordSmith Tools -- 07/05/2008 17:49:39

N Concordance
1 ked. We believe that these traits will contribute significantly to the
2 ccess exhibited by T. giacomellii will contribute significantly to its
3 and neutralising antibodies, that will contribute substantially to our
4 receptors on these T cell subsets also contributes significantly to the
5 , with transient placenta, could also contribute significantly. Gene and
6 DNA bend toward the major groove, and contributes significantly to a
7 other than NEFA and de novo lipogenesis contributed significantly to VLDL-TG
8 (at KI's 3135, 3335, and 3555) did not contribute significantly to the
9 herited MLC-2, if such exists, does not contribute significantly to the
10 ing individual components that did not contribute significantly (at P 5 0.15)
11 nce that the host inflammatory response contributes significantly to the
12 ccess exhibited by T. giacomellii will contribute significantly to its
13 ked. We believe that these traits will contribute significantly to the
14 Parasitism by wasps and tachinids also contributed substantially to
15 to as 'male factor' infertility) contributes substantially to
```

**Figure 5** Concordance lines of the node word *contribute*

Compare the *HSC* examples below:

- (10) Food transfers seem to be an *important contribution* to the total food obtained by young.
- (11) All these mechanisms *contribute importantly* to the shaping of animal embryos.
- (12) It is thus unlikely to make a *substantial contribution* to rejection of cytosine.
- (13) That will *contribute substantially* to our understanding of the overall biological significance.
- (14) Abnormalities in potassium channel function are also unlikely to *substantially contribute* to conduction disturbances.

With regard to the semantic classification of these adverbs, three of them refer to adverbs of manner which give account of how an action (*contribute*) is performed. As can be seen from the examples above this general sense may present specific connotations depending on the context they appear in. For instance, the adverb *substantially* in example (13) refers to the overall importance of a given *contribution*, whereas in example (14) the same adverb is used in its more specific contextualised statistical / quantitative sense.

### c) noun + *contribution*

Not only do attributive adjectives premodify the abstract noun *contribution*. Premodification by means of another noun also occurs in the *HSC* corpus. However, it should be pointed out that nouns as premodifiers of *contribution* are not as common as adjectives. Among the most frequent N+N sequences, the following chunks could be identified: *call contributions*, *cell contributions*, *gene contributions*, *group contributions*, *mutation contributions*.

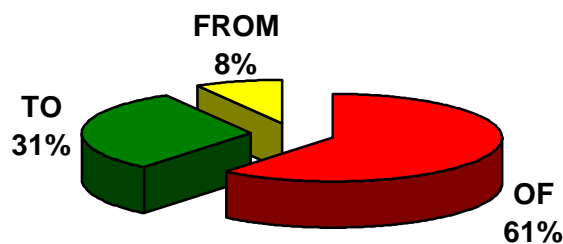
All these sequences contain content words and express a ‘source relationship’ between the modifying noun (on the left) and the head noun *contribution* (on the right). In other words, the modifying noun expresses the source of the head: *contributions* are provided by cells, genes, a group, etc., as is clearly observable in the examples below:

Source relationship	N+N sequence	Meaning / paraphrase
	<i>call contributions</i>	→ contributions provided by the calls
<i>HSC</i> example:		“Because our subjects differentially contributed to the number of calls recorded in each session, it is possible that our analyses were biased by calls from particularly vocal contributors. However, the extremely high percentages of correctly classified syllables suggests that unequal <b>call contributions</b> played little role in our discriminant function classification results.”
	<i>cell contributions</i>	→ contributions from cells
<i>HSC</i> example:		“Glucose phosphate isomerase (GPI) analysis (Papaioannou and Johnson, 1993) was performed on spleen samples from the two mice with ALL to estimate the contribution originating from the CCB-derived ES cells, injected into the C57Bl/6 blastocysts. Both mice had significant CCB ES <b>cell contribution</b> in the spleen (data not shown) suggesting that the lymphoblastic tumours are of ES cell origin.”
	<i>gene contributions</i>	→ contributions from genes
<i>HSC</i> example:		“Unlike these HOXA and HOXD genes, the HOXC <b>gene contribution</b> is not restricted so closely to the 58 end of the cluster.”
	<i>group contributions</i>	→ contributions of / from a group
<i>HSC</i> example:		“Prior probabilities options were utilized to weight unequal <b>group contributions</b> , and syllables were analyzed and classified separately.”

**Table 3** Noun + *contribution* collocation pattern

### d) *contribution* + preposition

This pattern has proven to be frequent in the *HSC*. It is particularly noticeable the fact that 83% of the occurrences of *contribution* are postmodified by a prepositional phrase that delimits the meaning of the abstract noun. *Contribution* is postmodified by the prepositions *of* (126 occurrences), *to* (64 occurrences) and *from* (17 occurrences).



**Figure 6** Prepositions following the noun *contribution*

It is important to mention that they are not simply function words (dummy prepositions) but, on the contrary, they are subcategorised by the abstract noun *contribution* in order to convey different meanings. In this context, each of them is used to refer to the source / origin / contributor (*of / from*) or the goal / purpose (*to*) of a *contribution*.

What seems to be rather remarkable is the fact that the prepositions *of* and *from* are equally interchangeable; that is, both of them refer to the origin / source of the *contribution*. There is no difference in between them and, in fact, they collocate with the same noun phrases. Consider some illustrating examples:

- (15) (...) with a significant *contribution from* surrounding cells.
- (16) (...) they did enable the extent of the *contribution of* the donor cell(s) within different tissues.
- (17) (...) that enabled the relative *contributions from* maternal and embryonic gene expression.
- (18) (...) it provides an opportunity to test the *contributions of* specific genes to neurobiological processes.
- (19) JGW wrote the paper with core *contributions from* all authors.
- (20) Barrow specifically discussed the *contributions of* women to the field of ornithology.

Despite being semantically equivalent, there are two worth mentioning aspects. Firstly, frequency rates show a preference for the use of the preposition *of* (88%, 126 out of 143 occurrences) over *from* (12%, 17 out of 143 occurrences). Secondly, there seems to be a recurrent tendency to use *of* in correlation with the preposition *to*. In other words, the combination “contribution *from* X *to* Y” does appear (e.g. “*In the present study, a contribution from* unlabeled hepatic lipid stores *to* TG synthesis may be less likely, because the subjects had been fasted for 24 hours by the end of the infusion test, which should substantially reduce hepatic lipid stores”) but less frequently than the sequence “contribution *of* X *to* Y”. Here are some examples:

- (21) *Contributions of* blastocyst micromanipulation *to* the study of mammalian development (...)
- (22) *Contribution of* prostaglandin EP2 receptors *to* renal microvascular reactivity (...)
- (23) We wished to determine more systematically the intrinsic *contributions of* individual Bcd-binding sites *to* transcriptional activation by Bcd.
- (24) Our ability to assess the relative *contributions of* genic and chromosomal factors *to* the genetic barrier (...)

Regarding the noun phrases that play the role of Prepositional complement of these sequences, they correspond to both animate [+/- volitional] (e.g., *authors, cells, parasitoids*) and inanimate entities (e.g., *genes, processes, enzymes, metabolism*). They all refer to scientific processes, substances and mechanisms which are characteristic of this specific genre. The following Table shows frequent collocates found in combination with the pattern *contribution + of / to / from* (with the number of occurrences in brackets<sup>17</sup>).

<i>contribution</i> + <i>OF</i> (126 occ)		<i>contribution</i> + <i>TO</i> (64 occ.)		<i>contribution</i> + <i>FROM</i> (17 occ.)
genes (6)	studies	gene activation (4)	differences	cells (2)
site (5)	age	phenotype (4)	<b>mutation</b>	stores
factors (5)	lineages	<b>cell/gene pool</b> (2)	the total	<b>mechanisms</b>
cells (4)	reactions	project (2)	catalysis	parasitoids
receptors (4)	routes	the host (2)	properties	cell pool
proper names (3)	<b>binding</b>	interface (2)	anaphase	authors
proteins (2)	activities	<b>process</b> (2)	<b>binding</b>	biscuits
effects (2)	transcription	stabilization (2)	positioning	helicases
thymus (2)	<b>mechanisms</b>	<b>binding affinity</b>	synthesis	nitrosyl (heme)hemoglobin
<b>mutation</b> (2)	metabolisms	total amount of	specificity	<b>diet</b>
<b>diet</b> (2)	infertility	values	understanding	de novo lipogenesis
fatty acids (2)	species	mosaic	genetic	neighbours
fat accumulation	<b>process</b>	work	isolation	
box	interactions	assembly	development	
membrane	experience	theory	cancer	
substances	pathway	<b>metabolism</b>	progression	
gluconeogenesis	<b>model</b>	problem	blood pressure	
enzymes	rearrangement	strength	degenerations	
component	errors	regulation	<b>mechanisms</b>	
plasma	agents	<b>proteins</b>	parameters	
males & females	source		events	
	variables		assays	
	individuals		<b>model</b>	
			rejection of	

**Table 4** Noun Phrases as Prepositional complements of the sequence *contribution of / to / from*

### Discussion of overall results

The main goal of the previous section was to provide a detailed account of the colligational and collocational patterns of the abstract noun *contribution* in the *HSC*. After having investigated the various patterns the noun *contribution* collocates with, several conclusions can be drawn from these findings.

Three processes need to be highlighted from the overall results: a) various degrees of delexicalisation in restricted verb collocates, b) preference for the use of periphrastic structures (i.e. verb + abstract noun) in active forms and c) a general tendency for adjective collocates to appear in attributive position. Each of these processes deserves further discussion.

Firstly, corpus data have revealed that *contribution* appears in restricted combinations of the type **V + Noun**; in particular, the patterning associated with the lexical verbs *make*, *provide* and *produce*. Of the several senses related to these polysemic verbs, it has been proved that their core original meanings are not the most frequently used. On the contrary, the most prominent use of the above verbs points to the general meaning of “doing”; that is, the lexical verb has become a grammatical form with very little semantic content on its own. Thus, it can be stated that these verbs have undergone a process of what scholars (Sinclair 1992; Partington 1993; Howarth 1996, Hunston & Francis 2000, among others) refer to ‘delexicalisation’ and/or ‘semantic depletion’ and, as a consequence, their original meaning is no longer the most prominent.

Howarth (1996:94) further claims that such delexical senses are being identified as semantically depleted not only on the basis of the semantic nature of the verb itself but also on the kind of noun in combination with it and the correspondence of the collocation to a morphologically related verb of equivalent meaning. These last two tests for semantic depletion will now be further explored.

Following Howarth’s (1996) theory, the nature of the noun may contribute to a certain extent to the fact that the **V + Noun** pattern be regarded as a restricted collocation. In his point,

abstract nouns are more likely to be combined with semantically empty verbs; in other words, abstract nouns can activate the semantic weakening of the verb they collocate with: “*the less concrete the noun is, the less possible it is for the primary literal sense of the verb to be activated*” (Howarth 1996:97). In *make chemical products* the noun (i.e. [chemical] products) is concrete and the verb (i.e. *make*) keeps its literal meaning of “creating, producing”, whereas in *make a contribution* the noun (i.e. *contribution*) is abstract and the sense of the verb has faded away and, consequently, does not refer to the notion of “producing” anymore.

It seems thus that one of the main tests so as to label a given verb sense as delexical is the possible transformation from a **V + Noun** construction (i.e. *make a contribution*) to a lexical verb (i.e. *contribute*). In *make a contribution*, the verb could be replaced by *provide* and/or *produce* and such a transformation would keep working perfectly. However, if the noun (i.e. *contribution*) were substituted in the periphrastic structure, the sense of the whole multiword unit would be significantly altered. As Howarth (1996) states, “*if the delexical verb is regarded as contributing little or no independent meaning to the whole collocation (having the role of a grammatical prop for the noun), then one is justified in considering that no substitute for a noun can exist for the verb in that sense.*” (Howarth 1996:112-113).

Concerning the equivalence of the periphrastic structure to a cognate verb (i.e. *make a contribution* ~ *contribute*), it is worth noting that this is an outstanding feature of the abstract noun selected for the present analysis. In this respect, Howarth asserts that “*collocations involving these verbs demonstrate a strong tendency to collocate with nouns that have cognate lexical verbs*” (Howarth 1996:98).

It has been proved that periphrastic structures consisting of **restricted Verb collocates + Noun** are semantically equivalent to the abstract noun’s cognate lexical verb. The main question is then raised as to what syntactic features distinguish the periphrastic uses (i.e. **delexical verb + Noun**) from the full lexical synthetic constructions. Taking into account that “*syntax is driven by lexis*” (Francis 1993:142), it has been observed that there is a preference for the use of the periphrastic structure (i.e. delexical verb + *contribution*) in active constructions

Free collocates have also been analysed. On average, the abstract noun under study collocates with a wide range of free combination lexical verbs. Unlike restricted collocates, the absolute frequency of this pattern is much lower. Interestingly, it seems that the wider range of free collocates an abstract noun combines with, the less frequently such verbs are used. Figures have revealed that the absolute frequency of free combination verbs is hardly significant as in most cases it consists of no more than two or three instances with a given noun (e.g. *examine a contribution* [3 occurrences]).

A further feature that is striking regarding the collocational patterns of the abstract noun *contribution* in the *HSC* is that there is a general tendency for adjective collocates to be used as internal pre-head modifiers to the following abstract noun; that is, there seems to be a preference for the use of adjectives in attributive position.

In addition, the typology of adjectives found in combination with *contribution* in the *HSC* deserves special attention. As pointed out in the previous section, the kind of adjectives used to denote properties of abstract nouns fall into two main semantic groups: **descriptors** and **classifiers**. The properties conveyed by the former may relate to opinions (**evaluative** adjectives) and **size**, **quantity**, **extent** or **time** descriptors, whereas the latter refer to **relational** or **topical** adjectives.

The analysis of the abstract noun *contribution* reveals that it shows a strong preference for **relational** (e.g. *relative contribution*, *functional contribution*) and **topical** (e.g. *caloric contribution*) classifying adjectives. These findings corroborate Biber’s et al. (1999) assertion of the high use of classifiers in academic prose. Due to the fact that scientific writing is more concerned with defining and describing demonstrable data from an objective point of view,



both topical and relational adjectives showing the subject area or a relation with the noun described are likely to be more frequently used.

## Conclusions

The above processes have made it evident that the use of the abstract noun *contribution* tends to be associated with specific grammatical patterns. From the point of view of medical discourse analysis and English for Specific Purposes (ESP), there is much to be said about the phraseological behaviour of abstract nouns in medical English.

Firstly, it has been shown that *contribution* in the *HSC* is usually combined with semantically empty support verbs (e.g. *make*), whose original meaning has faded away. Through a process of delexicalisation, these verbs have gained a more grammatical role and simply perform a verbal function, whereas the abstract nouns they collocate with can be said to carry the semantic content of the whole multiword unit. This redistribution of meaning across lexical items underlines the close interrelationship between the verb and the abstract noun associated in collocation with it (e.g. *make a contribution*). In addition, such periphrastic use has been found to correspond to a morphologically related verb of equivalent meaning (i.e. *contribute*), which reinforces the delexicalised use of the restricted verb collocates in verb-noun combinations.

Other noteworthy lexico-grammatical features observed lie in the fact that the restricted verb + *contribution* construction is frequently used in the active voice. To this respect, some remarkable differences have been observed with other restricted **V + Noun** collocations (i.e. *draw/reach a conclusion, reach an agreement, make a decision*), which are often passivised in the *HSC* corpus.

Finally, the investigation of the most salient patterns of the abstract noun *contribution* has also informed about the nature of the adjectives associated in combination with it. Data drawn from the *HSC* have addressed two main issues regarding adjective usage. The first issue relates to the fact that the adjective collocates of the selected noun analysed show a clear preference for attributive position. The second general observation refers to the kind of adjectives being associated with the abstract noun under study. Evidence from the *HSC* has revealed that the noun *contribution* is usually premodified by relational and topical classifying adjectives. As already pointed out, due to the fact that the publications examined in the *HSC* corpus are related to medical topics (in the form of observable, demonstrable data), classifiers were expected to outnumber descriptors as the most common premodifiers of *contribution* in the native corpora.

The findings drawn from the analysis of the *HSC* with regard to the phraseological behaviour of the word *contribution* in medical English have also given rise to several implications for the teaching and learning of phraseology in specific genres.

Further corpus-based studies on collocations of both native and non-native production in scientific English are needed so as to contribute to a thorough characterisation of the discourse community as well as pinpoint non-native speakers' difficulties and design efficient collocation teaching materials. For a better understanding of non-native speakers' main problems as far as their written production is concerned, the examination of a large text corpus consisting of their published articles in English would also be beneficial in order to evaluate their collocational performance against native corpus evidence in the light of greater amounts of data.

There is also an urgent need to explore the materials used in EFL and ESP contexts for the teaching of collocations in scientific English. Very little is known about the materials, in the form of coursebooks, writing manuals, workbooks and grammars, among others, used for the teaching of English collocations in health sciences. Future work should be applied to the analysis of these sources for further insights into the way collocations are taught. As collocational patterns are essential to the construction of knowledge in the disciplinary area, it is firmly believed that research on lexico-grammar will provide the framework for the creation

of improved collocation teaching materials. Thus, as can be inferred from all these considerations, there is clearly much work to be done in the area of phraseology and lexical patterning of general vocabulary in specialist registers.

## References

- Adams Smith, D. E. (1984). "Medical discourse: aspect of author's comment". *The ESP Journal*, 3, 25-36.
- Aitchison, J. and D. M., Lewis (2003). "Polysemy and bleaching". In B. Nerlich, Z. Todd, V. Herman and D. Clarke (eds.) *Polysemy: Flexible Patterns in the Mind*. Berlin: Mouton De Gruyter, 253-65.
- Altenberg, B. and M. Tapper (1998). "The use of adverbial connectors in advanced Swedish learners' written English". In S. Granger (ed.) *Learner English on Computer*. London & New York: Addison Wesley Longman, 80-93.
- Altenberg, B. and S. Granger (2001). "The grammatical and lexical patterning of MAKE in native and non-native student writing". *Applied Linguistics*, 22/2, 173-94.
- Banks, D. (1994). "Clause Organization in the Scientific Journal Article". *Alsed-Lsp Newsletter*, Vol 17/2, 4-16.
- Barnbrook, G. and J. M. Sinclair (1995). "Parsing COBUILD entries". In J. M. Sinclair, M. Hoelter and C. Peters (eds.) *The Languages of Definition: The Formalization of Dictionary Definitions for Natural Language Processing* Luxembourg: Office for Official Publications of the European Community, 13-58.
- Bednarek, M. (2007). "Local Grammar and Register Variation: Explorations in Broadsheet and Tabloid Newspaper Discourse". *ELR Journal* 1.1. Available at: <http://www.ejournals.org.uk/ELR> (accessed: 28 June 2009)
- Bhatia, V. (1993). *Analysing Genre: Language in Use in Professional Settings*. London: Longman.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., S. Conrad and R. Reppen (1998). *Corpus linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Carter, R. (1998) (2<sup>nd</sup> edition). *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.
- Cowie, A. P. (1988). "Stable and Creative Aspects of Vocabulary use". In R. Carter and M. J. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman, 126-37.
- Francis, G. (1993). "A Corpus-Driven Approach to Grammar. Principles, Methods and Examples". In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology. In Honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 137-56.
- Flowerdew, J. (2003). "Signalling nouns in discourse" *English for Specific Purposes* 22: 329-46.
- Gabelentz, G. (1891). *Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: Weigel.
- Gledhill, C. (1995). "Collocation and genre analysis. The discourse function of collocation in cancer research abstracts and articles". *Zeitschrift für Anglistik und Amerikanistik*, 1, 1-26.
- Gledhill, C. (1996). "Science as a collocation. Phraseology in cancer research articles". In S. Botley, J. Glass, T. McEnery and A. Wilson (eds.) *Proceedings of teaching and language corpora 1996: UCREL technical papers* Vol. 9, The University of Lancaster, UK, 108-26.
- Gledhill, C. (2000a). "The discourse function of collocation in research article introductions". *English for Specific Purposes*, 19/2, 115-35.

- Gledhill, C. (2000b). *Collocations in science writing*. Gunter Narr: Tübingen.
- Granger, S. And F. Meunier (eds.) (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins.
- Gross, M. (1993). "Local grammars and their representation by finite automata". In M. Hoey (ed.) *Data, Description, Discourse*. London: Harper Collins.
- Halliday, M. A. K. (1993). "Language in a changing world". Occasional paper no. 13 *Applied Linguistics Association of Australia*.
- Halliday, M. A. K. (1998). "Things and Relations. Regrammaticising Experience as Technical Knowledge". In J. R. Martin and R. Veel (eds.) *Reading Science: Critical and Functional Perspectives on Discourses of Science*. London: Routledge, 185-235.
- Howarth, P. A. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer Verlag.
- Howarth, P. A. (1998a). "The phraseology of learners' academic writing". In A. P. Cowie (ed.) *Phraseology: Theory, analysis and applications*. Oxford: Clarendon Press, 161-86.
- Howarth, P. A. (1998b). "Phraseology and second language proficiency". *Applied Linguistics*, 19/1, 24-44.
- Huddleston, R. (1971). *The Sentence in Written English. A Syntactic Study Based on an Analysis of Scientific Texts*. Cambridge: Cambridge University Press.
- Hunston, S. and G. Francis (2000). *Pattern Grammar*. Amsterdam: John Benjamins.
- Hunston, S and J. M. Sinclair (2000). "A local grammar of evaluation". In S. Hunston and G. Thompson (eds.) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 75-101.
- Hunston, S. (2002). "Pattern grammar, language teaching, and linguistic variation. Applications of a corpus-driven grammar". In R. Reppen, S. M. Fitzmaurice and D. Biber (eds.) *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, 167-83.
- Hyland, K. (2008). "As can be seen: Lexical bundles and disciplinary variation". *English for Specific Purposes*, 27, 4-21.
- Kaszubski, P. (2000). *Selected Aspects of Lexicon, Phraseology and Style in the Writing of Polish Advanced Learners of English: A Contrastive, Corpus-Based Approach*. Available at: <http://www.staff.amu.edu.pl/~przemka/rsearch.html#PhD> (accessed: 28 June 2009)
- Laso, N. J. and I. Verdaguer (2005). "An invitation to explore conclusions and its restricted collocations". *Specific*, 2, 47-54.
- Lee, D. (2001). "Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle". *Language Learning & Technology*, 5/3, 37-72.
- Lee, D. and J. Swales (2006). "A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora". *English for Specific Purposes*, 25/1, 56-75.
- Lewis, M. (1996). "Implications of a lexical view of language". In D. Willis and J. Willis (Eds.) *Challenge and Change in Language Teaching*. Oxford: Heinemann, 10-16.
- Luzón, M. J. (2000). "Collocational frameworks in medical research papers: A genre-based study". *English for Specific Purposes*, 19, 63-86.
- Master, P. (1991). "Active verbs with inanimate subjects in scientific prose". *English for Specific Purposes*, 10, 15-33.
- Myers, G. (1989). "The Pragmatics of Politeness in Scientific Articles". *Applied Linguistics*, 10/1, 1-35.
- Noguchi, J. (2006). *The Science Review Article: An Opportune Genre in the Construction of Science*. Bern: Peter Lang.
- Oakey, D. (2002a). "A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines". In R. Reppen, S. M. Fitzmaurice and D. Biber

- (eds.) *Using corpora to Explore Linguistic Variation*. Philadelphia: John Benjamins, 111-29.
- Oakey, D. (2002b). "Lexical Phrases for Teaching Academic Writing in English: Corpus Evidence". In S. Nuccorini (ed.) *Phrases and phraseology –data and descriptions*. Bern: Peter Lang, 85-105.
- Partington, A. (1993). "Corpus Evidence of Language Change. The Case of the Intensifier". In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology. In Honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins, 177-92.
- Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- Pawley, A. (1985). "On speech formulas and linguistic competence". *Lenguas Modernas*, 12, 84-104.
- Renouf, A. and J. M. Sinclair (1991). "Collocational frameworks in English". In K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics*. London: Longman, 128-44.
- Salager-Meyer, F. (1992). "A text-type and move analysis study of verb, tense and modality distribution in medical English abstracts". *English for Specific Purposes*, 11, 93-113.
- Salager-Meyer, F. (1994). "Hedges and textual communicative function in medical English written discourse". *English for Specific Purposes*, 13, 149-70.
- Scott, M. (1996). *Wordsmith Tools 3*. Oxford: Oxford University Press.
- Scott, M. & Tribble, C. (2006). *Textual patterns*. Amsterdam: John Benjamins.
- Simpson, R. C. (2004). "Stylistic Features of Academic Speech: The Role of Formulaic Expressions". In U. Connor and T. A. Upton (eds.) *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 37-64.
- Sinclair, J. M. (1987a). "Collocation: a progress report". In R. Steele and T. Threadgold (eds.) *Language Topics: Essays in Honour of Michael Halliday*. Amsterdam: John Benjamins, 319-32.
- Sinclair, J. M. (ed.) (1987b). *Looking Up: An Account of the Collins COBUILD Project*. London: Collins ELT.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. (1992a). "Trust the Text: the Implications are Daunting". In M. Davies and L. Ravelli (eds.) *Advances in Systemic Linguistics. Recent Theory and Practice*. London and New York: Pinter Publishers, 5-19.
- Sinclair, J. M. (1992b). "The Automatic Analysis of Corpora". In J. Svartvik (ed.) *Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82, Stockholm 4-8 August 1991*, Berlin and New York: Mouton de Gruyter, 379-97.
- Sinclair, J. M. (1997). "Corpus Linguistics at the Millenium". In J. Kohn, B. Ruschoff & D. Wolf (eds.) *New Horizons in CALL*. Szombathely: Bersenyi Daniel College.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. New York: Blackwell.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Sweetser, E. (1988). "Grammaticalization and semantic bleaching". In S. Axmaker, A. Jaisser and H. Singmaster (eds.) *Proceedings of the 14<sup>th</sup> annual meeting of the Berkeley Linguistics Society*. Berkeley, California: Berkeley Linguistics Society, 389-404.
- Teubert, W. and A. Čermáková (2004). "Directions in corpus linguistics". In M. A. K. Halliday, W. Teubert, C. Yallop and A. Čermáková (eds.) *Lexicology and Corpus Linguistics*. London: Continuum, 113-65.
- Varttala, T. (1999). "Remarks on the Communicative Function of Hedges in Popular Scientific English and Specialist Research Articles". *English for Specific Purposes*, 18/2, 177-200.

- Widdowson, H. G. (1979). *Explorations in Applied Linguistics*. Oxford: Oxford University Press.
- Williams, I. A. (1996). "A Contextual Study of Lexical Verbs in Two Types of Medical Research Article". In *English for Specific Purposes*, 15/3, 175-97.
- Williams, G. (1998). "Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles". *International Journal of Corpus Linguistics*, 3/1, 151-71.
- Williams, G. (1999). "Looking in before looking out: Internal selection criteria in a corpus of plant biology". *Papers in Computational Lexicography. Proceedings of Complex '99*. Hungary: Budapest, 195-204.
- Williams, G. (2002). "In search of representativity in specialised corpora: Categorisation through collocation". *International Journal of Corpus Linguistics*, 7/1, 43-64.
- Williams, G. (2005). "Challenging the native-speaker norm: a corpus-driven analysis of scientific usage". In G. Barnbrook, P. Danielsson and M. Mahlberg (eds.) *Meaningful Texts. The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. London/ New York: Continuum, 115-27.
- Wray, A. (1999). "Formulaic language in learners and native speakers". *Language Teaching*, 32, 213-31.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

<sup>1</sup> Other linguists who have stressed the importance of analysing the prototypical phraseological expressions in different genres and registers are Lewis 1996; Williams 1998/1999/2002; Biber 2006 and Scott & Tribble 2006.

<sup>2</sup> Further genre analyses can be found in Biber et al. 1999; Lee 2001; Lee & Swales 2006 and Scott & Tribble 2006.

<sup>3</sup> In his analysis of scientific prose, Master (1991) pays special attention to writers with an Asian background.

<sup>4</sup> *Grup de Recerca en Lexicologia i Lingüística de Corpus* (GRELiC) [Lexicology and Corpus Linguistics Research Group] ; University of Barcelona.

<sup>5</sup> The Professional English Research Consortium (PERC) has developed a 100 million word corpus of Professional English, which has been recently released.

<sup>6</sup> All these articles correspond to the years 1998 and 1999.

<sup>7</sup> Biology: *Genes and Development* (40 articles), *Genetics* (54 articles), *Journal of Cell Biology* (26 articles), *The American Journal of Primatology* (20 articles), *Biological Control* (97 articles), *The Journal of Experimental Zoology* (32 articles), *BioEssays* (99 articles), *Integrative Biology* (18 articles), *Zoo Biology* (65 articles); biochemistry: *Biochemical Journal* (53 articles), *The Embo Journal* (64 articles); medicine: *Journal of Clinical Investigation* (53 articles), *British Medical Journal* (58 articles), *Journal of Bacteriology* (39 articles).

<sup>8</sup> © Mike Scott & Oxford University Press, 1998 (<http://www.lexically.net/wordsmith/>).

<sup>9</sup> This is a central idea in corpus linguistics.

<sup>10</sup> In Teubert's & Čermáková's words: "we are only justified in claiming that a given corpus is representative of a discourse, however we have defined it, if we have, at least in principle, access to all the texts the discourse consists of." (Teubert & Čermáková 2004: 117).

<sup>11</sup> See Laso, N. J. & Verdaguer, I. 2005. "An invitation to explore conclusions and its restricted collocations", *Specific*, 2: 47-54.

<sup>12</sup> The term 'semantic bleaching' was coined by the German neogrammarian Georg von der Gabelentz in 1891 (see section 'References'). Some other linguists (Sweetser 1988; Aitchinson and Lewis 2003) refer to that concept when discussing polysemy and grammaticalisation.

<sup>13</sup> These sequences have been referred to with different labels such as "speech formulas" (Pawley 1985), "multiword units" (Cowie 1988), "prefabricated routines" (Cowie 1988), "collocations" (Sinclair 1991), "formulae" (Moon 1992) and lexical bundles (Biber et al. 1999; Cortes 2002), among others.

<sup>14</sup> Examples (1a), (2a), (3a), (4a) and (5a) have been taken from the *HSC*, whereas their counterpart (b) examples have been produced *ad hoc* so as to compare the periphrastic structures (i.e. **V + contribution**) against the lexical verb *contribute*.

<sup>15</sup> <http://wordnet.princeton.edu/perl/webwn>

<sup>16</sup> This classification of adjectives can be found in Biber et al. (1999: 508-509).

<sup>17</sup> Notice that **coloured** noun phrases appear as Prepositional complements of more than one of these prepositions.