Journal of Social Sciences

Original Research Paper

How Trainee Translators Analyse Lexico-Grammatical Patterns

Chris Gledhill and Natalie Kübler

CLILLAC-ARP, University Paris Diderot, Sorbonne Paris Cité, France

Article history
Received: 11-03-2015
Revised: 08-05-2015
Accepted: 08-05-2015

Corresponding Author:
Chris Gledhill
CLILLAC-ARP, University Paris Diderot, Sorbonne Paris Cité, France
Email: cgl@eila.univ-paris-diderot.fr

Abstract: In this study, we examine the ability of advanced students of specialised translation to identify and analyse 'generic collocations' in a corpus of specialised multilingual texts (mostly technical or scientific texts in English, French and German). In general, we find that our students attach much importance to frequently-occurring 'clusters' or 'n-grams'. However the students find it difficult to see these fragments as productive patterns of wording, or to assign a rhetorical function to them. This rather fixed view of phraseology suggests that there may be shortcomings in the way that we as teachers conceptualise and problematise the concept of the 'lexico-grammatical pattern' for our students. In the second part of this study, we suggest a different way of identifying and conceptualising phraseological phenomena using the metalanguage of Systemic Functional Grammar (SFG).

Keywords: Discourse Function, Generic Collocation, Lexico-Grammatical Pattern, Phraseology, Systemic Functional Grammar

Introduction

For many years now, an increasing number of linguists and educationalists have argued the case for a 'corpus-informed' approach to language learning, based on the analysis of large-scale text archives. Following this trend, a more central place has been given to the study of idiomatic phrases, collocational patterns and other regularities of wording, as encapsulated by the term 'phraseodidactics' (González Rey, 2005; 2008). Although phraseology as a subject has not yet become a central feature of the language syllabus, it is certainly the case that some key notions such as 'collocation' have become very widespread, most notably in areas such as second-language acquisition (Hasselgren, 2002), text-based terminology (Pavel, 1993) and Natural Language Processing (Williams, 1998; 2003). One of the reasons for the relative success of 'collocation' in these areas must lie in the way the notion is conceptualised. We would suggest that there are essentially two approaches to the notion within the field of phraseology: the phrase-oriented and the pattern-oriented approach. For the **phrase-oriented approach** (adopted by many mainstream phraseologists, lexicologists and analysts who are interested in language as system), the key units of analysis are 'idiomatic phrases,' 'proverbs' or 'stereotypes', that is to say exceptional or idiosyncratic sequences which are pragmatically 'marked' can thus often be easily identified in a given text. Such phraseological units clearly have an important rhetorical role to play in a variety of text-types belonging to the general language (horoscopes, popular journalism, film titles, etc.). In contrast, for the **pattern-oriented approach** (adopted by many corpus linguists, lexicographers and those concerned with language as discourse), the key units of analysis are 'collocational frameworks', 'lexical patterns', 'clusters' and so on. In contrast to idiomatic expressions, these constructions are often routine formulae and fragments of expressions which often attract little attention, but which often make up the most typical wording of a particular genre or text-type (for a review of this approach, see for example Hunston and Francis, 2000; Frath and Gledhill, 2005; Legallois and François, 2006).

We would suggest that it is this second, pattern-oriented, approach to phraseology has particular relevance to a range of applied and academic contexts, most notably in courses on language engineering, terminology, technical communication and so on. Let us take a specific example from our own teaching context: at the Université Paris Diderot, we teach phraseology and corpus linguistics to first year and second students who are for the most part being trained to work in the language industries (including document design, text mining, technical writing, specialised translation, etc.). (The course we are referring to here is usually known by its French abbreviation: M2 ILTS, *Industries de la langue et traduction spécialisée* = 'Language industries and specialised translation').

As part of their course, these students are asked to work on a terminological and phraseological database

(ARTES) and to write a brief analysis of what we call 'generic collocations'. We have reported elsewhere on how ARTES is used in order to teach and conduct research on terminological and phraseological units in French and one other language (often English or German) in order to promote awareness of these phenomena in the process of translation (Pecman, 2015; Kübler and Pecman, 2012). In this study, we examine the particular problem of how the same students identify and analyse generic collocations as part of the terminology dissertation they have to write at the end of their year of study. As we see below, our students are often confronted with problems of analysis, which even we as experienced linguists have difficulty disentangling. For instance, which part of a given sequence is 'terminological' and which part of the same sequence is 'collocational'? And is it possible to assign a regular meaning or 'discourse function' to this sequence? Can a pattern have more than one discourse function? Are similar patterns related and if so do they form a hierarchy?

Questions such as these have prompted us to re-evaluate some of the core assumptions of our own approach to phraseology. Those who promote a pattern-oriented view of phraseology often concentrate on the most statistically frequent or recurrent sequences of language, as can be seen in the many studies of 'bundles', 'clusters' or 'n-grams' that have been published over the years (Luzón-Marco, 1999; Biber *et al*., 2004; Scott and Tribble, 2006; Cheng *et al*., 2008; Lee and Xiao, 2008; Hyland, 2008). But perhaps the very ease of this approach and sheer quantity of data that it produces have obscured something about the way we conceive of language? As we shall see below, our students have no difficulty in finding useful lexico-grammatical patterns. But they also need help in analysing these fragments out of context. Also and more specifically, they need help in identifying and classifying underlying patterns of expression that are not obvious when one looks at de-contextualised, disembodied sequences derived from a concordancer.

Generic Collocations and Lexico-Grammatical Patterns

Before looking at our students' analysis, it is worth setting out our own general approach to phraseology and to phraseological patterns in language. One starting point which we adopted in previous work (Gledhill; 1995; 2000) is to look at the regularities of expression associated with high-frequency items, such as grammatical words. In these studies, we examined the distribution and co-occurrence patterns of grammatical items within a corpus of texts representing one genre (the scientific research article), with a particular focus on how these items are used across the different sub-sections of the research article. This research was principally inspired by the Birmingham school of corpus linguistics, particularly the notion of the 'collocational framework' (Renouf and Sinclair, 1991). Since this time, there have been several studies on the collocational behaviour of high-frequency items, notably in the form of lexical bundles, chains, clusters, n-grams and so on. Many of these studies propose an analysis of continuous sequences of signs which are *n* words long, hence the term 'n-grams'. The significance of these studies is that they allow for a systematic and replicable method of identifying the most frequent and typical expressions which occur in a given corpus of texts. This approach has proved to be a highly productive, as can be seen in some recent research (Mhedbhi, 2014). However, for reasons which we set out below, this approach has its drawbacks. In particular, we would suggest the term 'fragment' to describe these sequences, because it occurs to us that frequently-occurring n-grams can only be fragmentary and incomplete: they do not include key information about the structural productivity or semantic consistency of longer, more contextualised patterns of language.

We would suggest that there is a crucial difference between the 'n-gram' approach and the methodology we are developing with our students. According to our approach, **generic collocations** are discontinuous sequences of grammatical items which involve a productive, meaningful pattern. These collocations are 'generic' because they are associated with a specific text-type or 'genre'. But they are also 'generic' because they stand in contrast to 'specific collocations', that is to say the co-selection of two or more specific lexical items which may or may not belong to a specific text-type (Gledhill, 2000: 206). For example, one pattern of this type involves the sequence *(a/n) + Noun Group + is + Verb Group + to + Verb Group*. Without the help of a corpus and a concordancer, it would be difficult to specify what kind of sequence this might correspond to. But in a specialised corpus (such as the corpus we used in our early work, the Pharmaceutical Sciences Corpus: 500 000 words), the pattern is quite clear. In this corpus, the sequence corresponds to a very regular pattern of expression associated with explanations, in which the authors set out tentative (hedged, or modalised) propositions of specific (biochemical) processes:

1a. **NG** [Disease: HPV 16 E6, hyperphasia, leukaemia) **is VG** (Research-oriented verb: thought, likely, known) **to VG** (Biochemical process: act via many cells, attract factors, differentiate...)...

Let us take another example. The 'reporting' clause *VG + that + CLAUSE* is a well-known structure in academic writing. But once again, it is necessary to consult a corpus to have a more precise view of this sequence as a productive and meaningful pattern. In the Pharmaceutical corpus, this sequence typically includes a modal verb and a statement of how efficiently 'a

particular drug' + *did/did not/may* + 'treat' (*block*, *delay*, *prevent*) + 'a disease'. The following example *In this + NG + we + VG + that + VG + NG* shows how extended and regular this pattern can be, at the same time as allowing for variations which are semantically related:

1b. **In this NG** (paper, report, study) **we VG** (conclude, show, suggest) **that** (Drug: dextran-coated charcoal, ICI, TBCI) (did/did not/may) **VG** (Empirical verb: attenuate, block, delay, decrease, prevent) **NG** (Gene: BRCA1, IL2, M202)-(associated, induced, related) (Disease-related item: breast cancer, morphological changes, oxidation pattern...)

What can be said of these two examples? We would argue that this type of sequence is a prime example of the typical wording or phraseology of scientific writing. This claim supposes that by 'phraseology' we mean 'the typical way of expressing meaning within a particular type of discourse'. We would also suggest that we are dealing here with an extended phraseological unit and that all patterns of this type in English involve the co-selection of several different sub-patterns (Firth, 1957 refers to this as 'colligation'). As we can see in example 1b, the sub-pattern *we +VG + that* typically colligates with the sub-pattern *in this + NG*. As far as semantics is concerned, this overall pattern displays a high degree of specialisation or 'lexicalisation' (in the sense of Brinton and Traugott, 2005), that is to say a sequence of signs which displays a more restricted set of structural and semantic preferences than we would find if we looked only at *we + VG + that* out of context or in a less specialised corpus of texts.

More, recently we have proposed the term **lexico-grammatical pattern** to refer to both generic as well as specific collocations (Gledhill, 2011, 2012, Kübler and Volanschi 2012). There are several reasons for this change of emphasis. In the first place, the term 'pattern' is useful because it refers to a regularity of expression, a sequence of signs which is used habitually to refer to a single, consistent meaning within the same social context (as we see later on, in the Systemic Functional model, this abstract kind of meaning is referred to as 'discourse function' or 'rhetorical function'). Similarly, we should point out here that our use of the term 'pattern' is not original: it is an explicit reference to the work of John Sinclair and the Birmingham school of corpus linguists (c.f. Hunston and Francis, 2000; Hunston 2008, Groom, 2005; 2010). However, although Hunston and others use the term 'lexical pattern' or just 'pattern' for short, we use 'lexico-grammatical pattern', because this happens to be the name of the central stratum in our preferred model of linguistic description, Systemic Functional Grammar (SFG, Halliday and Matthiessen, 2014). From this point of view, the term 'lexico-grammatical' emphasises the fact that any pattern must contain information not only about lexis (a pattern is made up of one or more paradigms of semantically related, co-selected words), but also grammar (a pattern is built around at least one pivotal grammatical structure: this may serve as a linking item or 'hinge' between one pattern and other patterns).

So far then, we have described two related notions, which on paper at least seem fairly clear to us: 'generic collocation' and 'lexico-grammatical pattern'. However, there are clearly many problems with these notions. For example, how are lexico-grammatical patterns related to each other? Where are the borders or limits of a pattern? Which item in a given example is pivotal (i.e. invariable) and which is a part of a paradigm (variable)? We shall discuss some of these issues in the final sections of this paper. However, it is perhaps worth noting here that from our perspective, a 'pattern' should have a clearly identifiable rhetorical or discursive function. This definition stands in contrast to other terms in mainstream linguistics such as 'phrase', or more recent notions such as 'construction' (Goldberg, 1995) and 'collostruction' (Stefanowitsch and Gries, 2003). As far as we are concerned, these terms are useful when talking about abstract de-contextualised units ('noun phrase', 'the passive', 'the caused-movement construction', etc.), but by definition these terms do not include any one specific lexical or grammatical item as part of their structure and – importantly – do not depend for their definition on their rhetorical meaning or discourse function.

To what extent does a lexico-grammatical pattern have a particular 'discourse function'? Both of the examples cited above involve reporting verbs (*X is thought to, we conclude that Y*). This kind of pattern has often been seen as typical of academic discourse and the research article genre in particular. It is clear that a sophisticated text type such as the research article will be made up of many hundreds of these (and other equally productive) patterns. But this point raises a number of other interesting issues. What are the key patterns in a particular text type? Do these different patterns have the same status and if so what is their function in the eco-system of a particular genre? Are there macro-patterns, which subsume other more specific ones? Is it possible to find patterns which are unique to a particular genre or Language for Specific Purposes (LSP)?

In order to address some of these questions, in the rest of this section we examine how one lexico-grammatical pattern relates to its various sub-patterns and how this pattern can have different rhetorical functions in the LSP as opposed to the general Language (LGP). Let us examine the sequence *NG + Reduced relative Clause + Preposition + NG*. Because we wish to analyse this as a lexico-grammatical pattern, we need to specify at least one lexical item and one grammatical structure and we should include a more extended co-text. In this particular case, we therefore analyse the following sequence of signs < , *NG VG-ed/-en in NG*> (we put the sequence in triangular brackets in order to include the all-important comma at the beginning). If we look for

this sequence in the British National Corpus (BNC: 100 million words of British English), we find 32 examples (including various forms of the past participle). Most of these belong to three major variants of the same overall pattern. The first variant (examples 2a-e) involves the definition of a *journal, magazine* or *organisation* which was *based, founded* or *established* in a particular location, or at a particular time:

2a. She has built a reputation for herself as a specialist in the area and is not involved as photographic director for Al-Wasat**, a magazine based in** London, which is launched this month for distribution throughout the Arab world
2b. Balcon could reasonably feel defensive about the sort of criticism of British cinema that began to appear from the Film Society milieu, particularly in the pages of Close Up**, a journal founded in** 1927.
2c. But perhaps the easiest way to follow the new pop story is in the pages of The Face**, a magazine founded in** 1980 at the conjuncture of music, fashion, art and design.
2d. Pantell S.A. distributed a series of advertisements in order to persuade persons in the United Kingdom to purchase shares in European American Corporation Inc.**, a company incorporated in** Utah, U.S.A.
2e. He also stated his intention to resign his seat in the House ... and to accept the presidency of the United Negro College Fund (UNCF)**, an organization established in** 1944 to help blacks gain access to a college education.

The second variant (examples 2f-j) involves a *definition, name* or *term* which was *adopted, coined* or *used* at a particular time, location or context:

2f. The town's appearance today is that of a compact, prosperous market centre still clinging to its medieval street plan and to memories of the Franciscan priory once known as the "Lamp of Lothian"**, a designation transferred in** our own century to the church of St Mary.
2g. Legislative power in the State of Cambodia (SOC)**, a name adopted in** April 1989 by what had been since 1979 the People's Republic of Kampuchea…
2h. ..."the quality of life"**, a term coined in** this context by US Secretary of State George Schultz.
2i. [Author, Date] prefers to call this approach "output budgeting"**, a term used in** the UK civil service (see below).
2j. The lifestyle of a Cyrenaean aristocrat was centred on his *pyrgos* (Hdt. iv. 164; Strabo 836)**, a word found in** Asia Minor and Attica.

Finally, a third variant (examples 2k-o) involves a *chemical, compound*, or *substance* which is either found in a particular location or *involved, implicated* or *used* in a particular (biochemical) process:

2k. Hundreds of sheep farmers have been struck by the side effects of organophosphorus**, a chemical used in** animal dip.
2l. For instance, frequencies at 2-2.5 Hz stimulate the production of the body's own pain-controlling substances (the endorphins) and frequencies at around 80 Hz stimulate the production of 5 hydroxy tryptamine (serotonin)**, a compound involved in** brain and nerve function.
2m. The team says it has found no evidence of aluminium**, a cytokine characteristically found in** the brains of Alzheimer's patients in brain tissue samples tested under a new method of "nuclear microscopy"
2n. The constitutive and oxidant induced activity of Adenosine Diphosphate Ribosyl Transferase (ADPRT)**, an enzyme involved in** DNA repair, is reduced in patients with inflammatory bowel disease and also in those with colon cancer.
2o. An American study has found that theaflavin-2**, a substance found only in** black and oolong teas, was able to induce aptosis (cell death) in...

In all of the above examples (2a-2o), a noun (often a proper noun) is typically placed in capitals, speech marks or italics and then reformulated by a superordinate word (*Al Wasat > a magazine, Close Up > a journal*, *'output budgeting' > a term, pyrgos > a word*), or in examples (2k-2o), a technical noun is defined in relation to a superordinate term: *Aluminium > a cytokine, 5 hydroxy tryptamine > a compound, ADPRT > an enzyme, organophosphorus > a chemical*. This superordinate noun is then post-modified by a reduced relative clause.

It is clear that we are dealing here with the same overall lexico-grammatical pattern. However, it is also clear that the third variant of this pattern (examples 2k-o) is somewhat different from the other two. Whereas the first two sub-patterns are taken from texts dealing with biography, general culture and politics, the third pattern comes from scientific journalism (2k, 2m, 2o) or scientific research (2l, 2n). In addition, whereas in the first two sub-patterns the preposition *in* introduces noun groups referring to locations and dates, in the third pattern, *in* is associated with a biochemical process (expressing cause: *implicated in, involved in*), which takes place either in a specifically biological or chemical location (*brains, animal dip*) or as part of a nominalised biochemical process (*development of a response, brain and nerve function, DNA repair*). Not surprisingly, if we look for the same pattern in a specialised corpus of research articles (the Pharmaceutical Sciences Corpus, as mentioned above), we find a similar set of examples (32 instances). However and perhaps most importantly, the relative clause in the PSC is built almost exclusively around the verbs *implicated* and *involved* or (in a small minority of cases) *expressed.* In other

words, verbs which express material or biochemical processes of causation:

2p. Finally, BAL fluid contains significant levels of IL-16**, a cytokine implicated in** T lymphocyte recruitment, following antigen challenge of allergic mice [...]
2q. Cloning of Schistosoma mansoni Seven in Absentia (SmSINA)(+) homologue cDNA**, a gene involved in** ubiquitination of SmRXR1 and SmRXR2.
2r. Clustering of the BCR has also been reported to trigger an indirect association with invariant chain (Ii)**, a molecule involved in** the trafficking of newly synthesized MHC II. Since Ii resides
2s. Moreover Op18**, a protein involved in** the regulation of microtubule dynamics and Myosin Light Chain (MLC) are phosphorylated upon NKG2D cross-linking.
2t. In an earlier study [20], inducible Nitric Oxide Synthase (iNOS)**, an enzyme expressed in** cytokine induced macrophages [...] showed the same staining pattern as macrophage staining.

What can we conclude from this? Generally speaking, the pattern that we have been looking at corresponds to what terminologists call a 'definitional context' (Pearson, 1998). This is one of many such constructions that have been studied in research on the automatic extraction of definitions and neologisms (Humbley, 2001). We would claim that the basic meaning of explanation and definition corresponds to the 'rhetorical function' or 'discourse function' for this pattern. It is this broad, abstract meaning that allows us to see this as a pattern and not just the purely syntactic sequence *NG + Reduced Relative Clause + in + NG*. In addition, we would claim that this pattern has at least three different rhetorical functions depending on the type of discourse it is used in:

- The definition of institutions in terms of their original locations or dates of establishment,
- The definition of proper nouns and neologisms in terms of their original locations or dates of creation,
- The definition of biochemical substances in terms of their causal role in biochemical processes.

The overall meaning of this pattern is something like '*here is an explanation of the term we have just mentioned: it is an example of* (*superordinate term*) *X and it originated in* (*time or space*) *Y*'. This is how the pattern operates in most forms of the English language and as we have seen, the pattern typically occurs in elaborate academic or journalistic English. However, we can also observe how the parameters of the expression have evolved to express something rather more specific in the language of the pharmaceutical sciences. As we move from general English in which the pattern is used to define a term in relation to its origins (with verbs such as *based in, found in*), we enter into a more technical discourse in which the pattern is used to define a specific chemical entity in relation to its biochemical role (with verbs such as *implicated in, involved in*).

Lexico-Grammatical Patterns in Students' Reports

In the previous section, we set out our general approach to the analysis of phraseology. The approach consists of two basic steps: a) the observation of regularities of wording in a representative corpus of texts, b) the association of particular patterns of wording with particular discourse / rhetorical functions. In other words, after the direct observation of specific examples, we interpret any regularities of expression in terms of a more general lexico-grammatical pattern. From our point of view, this is a tried-and-tested methodology. However, it is salutary to observe how other observers go about this type of analysis, especially when they are our own students. In this section, therefore, we examine the choices made by our students when they identify their own examples and analyse them as lexico-grammatical patterns.

As mentioned above, the subjects in this study are students in applied foreign languages (speakers of French, English and typically German or Spanish) who are following a second-year Masters course in specialised translation (M2 ILTS, Université Paris Diderot). Since this course takes place at a French university, our students often write their reports and work on the ARTES database in French. Some of the examples cited later on in this section are therefore in French.

As part of the main assessment of this course, each of the M2 ILTS students is asked to write three end-of-year reports: (1) documentary research, (2) specialised translation and (3) terminology. The main part of the terminology report is dedicated to the description and analysis of LSP corpora (reported in Kübler and Pecman, 2012; Pecman, 2012); part of the translation report consists in describing examples of phraseological phenomena, such as collocations, colligations, semantic preference and semantic prosody. The students are asked to look at these phenomena and to examine how they may cause difficulties in the translation process, which they then attempt to solve by querying comparable (bilingual) corpora. As part of their preparation for these reports, each student is required to build a specialised comparable corpus of texts in French and one other working language. Typically the students choose to analyse scientific research articles or, more rarely, another specialised/technical genre (such as technical

manuals). The students then analyse their LSP corpora for terminology and phraseology using concordancers and other software. The main results of this analysis are then entered in the ARTES database and also written up as part of the terminology report and the translation report. Various teaching courses contribute to the students' skill-set for this particular project, notably: 'Tools for Corpus Analysis', 'Corpus Linguistics' and 'Terminology and Phraseology' (taught by either our colleagues or ourselves).

However not all of the terminology report is dedicated to the discussion of terminology. As part of their report, each student is also asked to analyse 5 different 'generic collocations', in other words 5 examples of lexico-grammatical patterns (as mentioned in section 2 above). These sequences are generally presented to the students by our colleagues in functional terms (i.e., they are regular patterns of expression in the LSP which have specific discourse functions, but they are not terminological in nature and are not specific to an individual text or a domain, but belong to academic English in general). The students are also taught to use corpus-query software, such as the 'concordance', 'collocate', 'clusters' and 'n-grams' functions in AntConc (Anthony, 2002), in order to extract a sample of these sequences systematically from their LSP corpora. This is an important methodological point, because although the students encounter this software at various stages of the M2 ILTS course, not all of them choose to use the 'clusters' and 'n-gram' functions when writing up their final dissertation. It is also important to note that each dissertation is supervised and evaluated by different teachers on the M2 ILTS course. As far as our own supervision is concerned, we allow our students free choice in the matter, with the instruction that they should demonstrate that they have used corpus-informed techniques in order to arrive at their choice of collocational patterns.

In the following section, we examine a sample of patterns analysed by 10 students whose terminology reports we supervised over the period 2012-2014 (thus a total of 50 different patterns). Each of the 50 patterns chosen by the students and the basic results of their analysis are set out in the form of a table in the Appendix at the end of this paper (Appendix 1). Note that in this Appendix 1, each pattern is presented in decreasing order of structural complexity (full clauses being the longest, most elaborate patterns at the top of the list and single words being the shortest, simplest structures at the bottom of the list). For convenience, we have repeated three interesting (and rather typical) patterns identified by our students (#1, #8, #26) in Table 1 (see below).

Space precludes us from analysing each example presented in the Appendix. In the remaining parts of this paper, we merely provide a summary of our observations including comments on each of the main columns set out there (NB the final column is discussed in section 4 of this paper. This column sets out our analysis of each example in terms of Systemic Functional Grammar (SFG). It is important to note here that we are using SFG as a metalanguage of analysis, but at the current time we do not teach SFG explicitly to our students.).

Patterns

The first column of Appendix 1 presents the citation forms given by our students. In 8 out of 50 patterns, the students have included variables (a word such as 'that', a symbol such as 'x', a part of speech tag such as VG). We see this as a sign that these students have correctly understood that these sequences are predictable but also productive lexico-grammatical patterns. In addition, we would suggest that the examples at the top of the list are better examples of patterns than those at the bottom: the best examples of patterns involve longer, more complex grammatical structures and leave more room for variable paradigms. However, many of these patterns involve specific lexical items and structures which suggest that some students have confused specific examples with generic patterns (we are thinking here of #33 *A critical issue in handling...*, #36 *The assumption underpinning the concept...,* #37 *To remain an unsolved challenge for...*). Also, towards the bottom of the list, the patterns begin to look increasingly incomplete. In our view, examples such as #44 *Key insight,* #45 *safety precautions*, #*46 related work*, #47 *serious games for*, #49 *due to*, #50 *in parallel with* are too short to qualify as valid lexico-grammatical patterns. Also, out of context, it is difficult to assign discourse functions to them. Some of these short examples (often corresponding to recurrent 'n-grams') may be technical terms, while others are short collocations or colligations, or perhaps compound nouns belonging to academic English. It occurs to us, however, that in many cases our students may have had more specific contexts in mind. To give just one example, the noun group in #44 *key insight* was chosen because it is a pivotal lexical item in a 'projected' complement clause (here in brackets):

> #44 Context: Our **key insight** is [[that a high degree of correlation exists between battery usage and a user's movements.]]

Examples such as these suggest that our students are good at identifying statistically significant fixed sequences, but they are generally less successful in identifying the more extended and therefore more meaningful patterns within which these fragments are embedded.

Table 1. (Extract). Lexico-grammatical patterns in 10 student reports

	Pattern	Structure[4]	Discourse function	Discourse system
#1	Failure to + VG (observe, follow, do so, heed, comply) + N G+ will result in + NG	Effective clause (Subject includes embedded expansion clause)	(Not supplied)	Ideational: Material, logical ('so') Interpersonal: Engagement: Source (projection 'failure to', modality 'will') Textual: Identification ('so')
#8	In this paper we outline...	Effective clause with Prepositional phrase in Marked Theme position	Referential 'introductory phrase'	Ideational: Mental, communicative Interpersonal: Engagement: Source Textual: Periodicity and identification
#26	It should be noted that..	Receptive (passive) Projecting clause with postponed clause	'Impersonal introductory phrase' 'Talking about characteristics, properties, specificities' 'Expressing a necessity'	Ideational: Mental Interpersonal: Engagement: Source: Projection Textual: N/A

[4]The analysis in columns 2 and 4 follows the conventions of Systemic Functional Grammar (SFG, c.f. Halliday and Matthiessen 2014; Martin and Rose, 2003)

Lexico-Grammatical Structure

The second column of Appendix 1 sets out the basic grammatical structure of each pattern using the terminology of Systemic Functional Grammar (SFG). For example, whether the pattern is active (effective) or passive (receptive), whether the pattern involves a projecting clause (as in #7 *it should be noted that*) or an expansion clause (as in #29 *As Figure X shows*).

Overall, 11 out of 50 patterns chosen by our students are full clauses (i.e. full sentences in traditional grammar). We consider that these are all legitimate examples of lexico-grammatical patterns. In addition, 28 out of 50 patterns are clause complexes, that is to say incomplete sequences, which require or predict the presence of another clause. Some of these are legitimate examples of lexico-grammatical patterns, as in #2 *We reserve the right [[to VG]].* Others are less clearly examples of patterns as in #36 *The assumption [[underpinning the concept]]*. Finally, at the lower end of the scale, 9 out of 50 patterns are phrases (i.e. preposition + noun group, such as #40 *to this end*) or groups (such as #44 *Key insight*) and 2 out of the 50 patterns are word-sized units (#49 *due to* and #50 *in parallel with*). As mentioned above, all of the examples chosen by the students are valid collocations or n-grams, since they have often been identified as statistically salient by the AntConc software. However, we would suggest that the most useful patterns chosen by our students (around two-thirds of the sample) involve full or partial clauses, while many of the other examples chosen by the students are only fragments of a more extended lexico-grammatical pattern.

Discourse Function

The third column of Table 1 presents the 'discourse function(s)' of each pattern as defined by each of our student sample. Although the students are often successful in identifying valid lexico-grammatical patterns, it is clear that almost all of them encounter problems when it comes to assigning rhetorical functions. One question which we were interested in is the range of discourse functions which our students were able to identify in the sample of 50 patterns. Column 3 shows that our students typically formulate rhetorical functions in terms of the basic structure: *Verb Group + Noun Group*. In addition, the students typically use one of 3 VGs *(Explaining / Presenting / Talking about*) and 12 NGs (*additional data*, *anteriority, compatibility, decisions, non-textual data, methods and tools, necessity, one's position, one's research goals, restrictions, the specifics of the study, the subject of the study*). This is an interesting range of functions and also happens to identify many of the most common patterns to be found in the ARTES database. However, there is often a mismatch between these functions and the collocational patterns to which they are associated. For example, one of the most common lexico-grammatical patterns identified by our students typically involves an active or passive reporting verb with the structure *'it + (active/passive V expressing a mental/communicative process) + that'*, as in the following examples (including many examples in French):

#12 **On estime que**... 'it is thought that'
#13 **On observe que**... 'it is observed that'
#14 **X shows that**...
#15 **Les** travaux ont **démontré que**... 'Studies have shown that'
#21 **It is** worth **noting** (**that**)...
#24 **It is** widely **known that**...
#25 **It** can **be seen that**...
#26 **It** should be **noted that**..

Structures such as these (called 'projecting' clauses in SFG) have been widely reported in the

literature and it is unsurprising that our students should have picked up on them in their own data. Interestingly, however, the semantic analysis of these patterns is often highly varied, even when we are looking at several examples of the same pattern. Thus, as can be seen in the Appendix, each of these patterns has been assigned very different discourse functions ('Impersonal introductory phrase', 'Talking about characteristics, properties, specificities', 'Expressing a necessity' and so on). To be fair, it is in fact rather difficult to assign one particular rhetorical function to examples such as these and we would not penalise students for assigning several functions to the same pattern.

One example of a legitimate lexico-grammatical pattern which has been correctly analysed in our sample is #20 *the intent of this article is to...* This pattern is identified as having the discourse function *'Presenting one's research goals'*. It would seem that this pattern has been identified by several other students in ARTES, because there are several examples of this in the database. The following sample gives a picture of the grammatical regularity of this pattern, but also its range of internal lexical variation:

- **L'objectif** de cette étude **est de**... 'the objective of this study is to....'
- **Le but** de cette étude **est** triple. 'the aim of this study is triple'
- Our **goal is** to...
- The **aim** of the present study **is** to do...
- The **purpose** of this work **is** to...

These examples are also related to the general pattern which we mentioned in the introduction to this paper, namely *the NG (aim, objective, goal) of this NG (study, paper, etc.) is to VG.*

A second correctly-identified example is #40 *To this end.* This happens to correspond to a pattern that can be found very frequently in the ARTES database. As we see below, this type of 'locution' or prepositional phrase has a textual signalling function, a pattern that our students are generally able to identify and categorise very successfully:

- Dans **ce but**... 'to this end'
- A **cette fin**... 'to this end'
- Avec **cet objectif** en vue... 'with this objective in sight'

A third example of a pattern that has been correctly analysed by one of our students is #10 *The analysis is based on.…* This pattern has the structure: *Attribution Clause (passive) + Preposition + NG* and is described accurately as *'Présenter ses méthodes, outils, ses approches, ses techniques' (Presenting one's methods, tools, approach, techniques').* There is some evidence in the student's report that this analysis has been based on the observation of several contexts and not just on the basis of frequency or a decontextualised example. Once again, many different students have identified examples of this structure in ARTES and so there is a good case to be made for seeing this as a prototypical example of a productive lexico-grammatical pattern in this type of text. Here are some examples:

- The analysis **is based on**... (context: The analysis is based on 3,263 observations from 821 MFIs in 91 countries reporting data.)
- The measurements **were made with** sth... (context: "The measurements were made with an 8 mm diameter diaphragm inset with optical glass."
- Les mesures **ont été réalisées avec** qch... 'the mesures were made using'... (context: Les mesures ont été réalisées avec un spectrocolorimètre Minolta...)
- Les résultats **sont exprimés conformément à**... 'results were expressed in accordance with' (context;: Les résultats sont exprimés conformément au système CIELAB …)

More generally speaking, however, we would suggest that in approximately half of our sample, our students have assigned discourse functions which are doubtful or simply incorrect. The most obvious examples of these in column 3 in the Appendix involve the use of lexico-grammatical structures rather than discourse functions (12 examples, such as *Adjectival, adverbial, nominal, verbal construction*). There are also several examples of vague categorisation (14 examples, including *demonstrative discourse, partial phrasal pattern, impersonal introductory phrase, referential introductory phrase, technical discourse*). In addition, in 7 out of 50 cases, no discourse function has been supplied. Perhaps more interestingly (because they are not incorrect), there are several cases where several discourse functions have been assigned to the same pattern (for example #21 *It is worth noting that...* which has been assigned five functions and #26 *it should be noted that...* which has been assigned three functions).

What general conclusions can be made about our students' analysis of phraseological patterns? On the positive side, we find that:

- The students identify some of the most important lexico-grammatical patterns in science writing, most notably structures involving active and passive reporting verbs ('projecting clauses').
- The students are adept at identifying patterns that are phraseological in nature rather than

terminological (they rarely confuse multiword terms with lexico-grammatical patterns, apart from some examples at the lower end of the scale).

On the negative side, we find that:

- the students often erroneously confuse a specific instance or 'example' with a generic 'pattern': as mentioned below, we suggest that our students require a more standardised way of recognising variability within more general patterns.
- the students often have difficulty assigning a 'discourse function' to the patterns they identify. We discuss this problem in more detail below.

But looking at these problems in detail, in the following section, we examine the extent to which the patterns identified by our students may be analysed using an alternative model of analysis (referred to here in the fourth column of Appendix 1 as 'Discourse System').

The Systemic Functional Approach to Analysing Lexico-Grammatical Patterns

In the previous section, we saw that our students are relatively successful in finding valid collocational patterns, but rather less successful when it comes to recognising underlying lexico-grammatical structures or assigning discourse functions to these patterns. The difficulties encountered by our students have prompted us to look for an alternative system of analysis. As mentioned above, a number of studies have analysed n-grams and other recurring sequences in specialised (LSP) texts. Hyland (2008) for example proposes a very simple system of classifying lexical bundles in academic and technical texts. Hyland distinguishes between three types of discourse function for each bundle: 'research-oriented', 'participant-oriented' and 'text-oriented' functions. Ultimately, all of these categories can be related to the three 'metafunctions' of Systemic Functional Grammar (Ideational, Interpersonal and Textual). We were curious to see how such an analytical system might be used to classify our own students' examples. We therefore re-analysed the 50 patterns identified by our students using a five-part classificatory system of 'discourse systems' proposed by Martin and Rose (2003). This analysis is summarised in column 4 of the table set out in the Appendix. Table 2 below presents a summary of these findings.

In order to interpret the findings presented in Table 2, it has to be understood that more than one discourse system may be involved in the same pattern (even within the same 'metafunction', as can be seen in example #21 *It is worth noting that...*). It is also worth noting that the discourse systems presented here (Experiential, Logical, Appraisal, Periodicity, Identification) were originally designed for the analysis of whole texts and that other discourse systems, which we have not mentioned here, such as Information structure (belonging to the textual metafunction) or Exchange (belonging to the interpersonal metafunction), can only be found in more extended stretches of discourse, such as spoken interaction (for a fuller picture, see Halliday and Matthiessen, 2014). Returning to Table 2, although we would suggest that the proportions set out here are typical of any analysis using this system, a number of features stand out: these are discussed in the following three sub-sections.

Ideational Metafunction

In SFG, the 'Ideational metafunction' refers to expression of participants and processes (the latter being generally divided into material, relational, mental and other minor processes), as well as the expression of logical connections within a text. As can be seen in Table 2, our students tend to choose patterns which express either a mental process (including a full range of cognitive processes *#12 on estime que, 'it is thought that',* perceptive processes #13 *on observe que 'it is observed that'* and communicative processes #14 *show that,* #16 *Let us say that,* etc.) or a 'logical connection' (#30 *when considering,* #42 *due to the presence of...*) or a 'comparison' (#32 *in the same way that*). It is significant that mental processes make up over 50% of the process-types in our students' sample. The proportion of process types which are typically used in full academic research articles is rather difficult to compare, since different researchers use different categories. Generally speaking, however, it is generally assumed that material and relational processes are dominant in scientific research writing (Banks, 1994). Halliday and Matthiessen (2014: 215) suggest that in a registerially-mixed sample of texts (8425 clauses) material processes typically make up 37%, followed by relational processes 36% and then mental processes 10% (the remaining 17% is shared between various minor process types). The proportions presented in Table 2 therefore reflect a rather marked preference among our students for projecting (i.e. reporting) verbs. This is nevertheless a reasonable analysis, given that these verbs (and the projected clauses which they introduce) are highly relevant to the phraseology of academic/scientific discourse.

Table 2. Distribution of Metafunctions in Sample of 50 Patterns

Metafunction	Discourse system	Examples	Total (NB the same pattern may involve more than one discourse system)
Ideational	Experiential: Representing experience as 1) Material process, 2) Mental process, 3) Relational process.	#1 Failure to.... will result in … (material process) #21 It is worth[[noting that...]] (relational process [[mental process]])	1) Material: 4 2) Mental: 18 3) Relational: 6
	Logical: 1) Addition, 2) Comparison, 3) Time, 4) Consequence.	#17 Il s'ensuit que ('It follows that'...) (logical connection) #32 *in the same way that* (comparison)	4) Logical consequence: 7 Total (Ideational): 35
Interpersonal	Appraisal: 1) Attitude (affect, judgment, appreciation of value), 2) Engagement (modality, concession, projection, source of authority), 3) Graduation (force, focus).	#21 It is worth noting (that...) (attitude, value) #3 Previous works focus on... (engagement, source of authority) #26 It should be noted that... (engagement, modality plus projection) #44 Key insight... (graduation, force)	Attitude: evaluation: 6 Engagement: 22 Total (Interpersonal): 33
Textual	Identification: 1) Presenting/presuming, 2) Tracking.	#4 This paper describes... (presenting) #5 Figure X depicts... (tracking: Exophora) #11 Proceed as follows... (tracking: Cataphora)	Reference: 11
	Periodicity: 1) Marked theme (textual), 2) Marked theme (interpersonal).	#7 Next on our agenda is... (textual theme) #11 Proceed as follows... (interpersonal theme)	Marked themes: 5 Total (Textual): 16

Interpersonal Metafunction

The 'Interpersonal function' refers to tone, authorial voice and other manifestations of interaction within a text, as can be seen for example in the absence or presence of the imperative, impersonal constructions, subjective lexis and so on. As can be seen in Table 2, the most important systems of interpersonal expression found in our sample include 'engagement' (the positioning of the authors in relation to assertions of fact: #24 *it is widely known that, #27 have been shown to be,* #28 *is known to be...*) and 'source' (the citation of other legitimising sources of knowledge #3 *previous works focus on,* #34 *studies undertaken to date*). The Interpersonal metafunction also includes the expression of 'affect' or 'emotion'. One would expect few examples of this in the types of text analysed by our students. However, the students did find several examples of 'appraisal', that is to say the evaluation of ideas or propositions (#22 *il est intéressant de constater que* 'it is interesting to note that', #33 *a critical issue in the handling of,* #32 *to remain an unsolved challenge*). Also, the sample includes one expression of 'graduation' (evaluation by expressing graduated focus or force, as in in example #44 *(a) key insight...*) Generally speaking, however, although our students have identified a range of patterns which involve some degree of interpersonal expression, they appear to be less explicitly aware of this function. When the students assign discourse functions to expressions such as these, they often refer to register or rather vague style labels such as 'impersonal introductory expression' and so on.

Textual Metafunction

The 'Textual function' involves the explicit marking of cohesion within a text, either by 'tracking' referents as they progress within the text, signalling periodic changes in the flow of the text, presenting referents as 'new' or 'given' and so on. Generally speaking, explicit linking devices and conjunctive adjuncts are always present, but not necessarily very frequent in scientific research articles (the main text type analysed in this study). It is therefore interesting to see that these items are an important category of pattern identified by our students. The most frequent examples of this type involve 'tracking' and 'presentation', that is to say expressions which identify or present a key referent in the co-text, either indirectly (#4 *This paper describes*) or directly (*#8 In this paper, we, #5 Figure X depicts*). Another category includes 'periodicity' or explicit signals of ordering in the text, including what Halliday and Matthiessen (2014) call

'marked textual themes', such as #6 *Finally, we consider...*#7 *Next on our agenda is* and *#11 Proceed as follows* (exceptionally, this example, together with #1 was found by a student whose project was on technical instruction manuals). A final example involves 'anaphoric nouns' which assign a new informational value to an existing textual referent (e.g., #40 *to this end*). Overall, our students are good at spotting patterns with this type of function, as can be seen in the terminology they use, e.g., 'referential introductory phrase'.

Space prevents us from setting out a full SFG analysis of all 50 patterns identified by our students and we have only been able to provide a minimal analysis in the Appendix. As mentioned above, we do not currently teach SFG to our students. In addition, we would not suggest that this method of analysis is an appropriate alternative to the categories of discourse function used by our colleagues or in the ARTES database. Nevertheless, we believe that the SFG system does provide a relatively systematic way of categorising the very different types of pattern which we have been looking at in this study. There are of course limitations to this approach; the 'metafunctions' of systemic functional analysis are sometimes not as specific or as intuitive as the discourse functions assigned by our students. Nevertheless, our main intention in conducting this survey has been to see whether there are any regularities in the way that our students identify and classify these patterns. Overall, we find that there is a clear preference among our students for two types of discourse system:

- Patterns which express 'interpersonal engagement', in other words patterns with a projecting (reporting) verb or expressions which signal modality or the source of a given assertion
- Patterns which express 'textual identification', that is to say segments which involve an explicit metalinguistic reference to the co-text

Conclusion

In this study, we have explored a very specific exercise in phraseology (the identification and analysis of 'generic collocations') performed by a group of advanced Masters students of specialised translation. Although this exercise is designed for trainee-translators, we believe that this type of analysis is also relevant for the development of academic and technical writing skills in various domains (including abstracting and review writing), as well as other more generally transferable language skills (such as corpus-informed research). As mentioned above, we ask our students to use concordancing software to extract generic collocations from their own corpora of specialised texts. While our students are generally good at identifying what they feel to be regular patterns of wording, many identify incomplete fragments, often leaving out relevant items or elements of grammatical structure which would allow the pattern to be meaningful out of context and therefore truly 'generic'. Also, while some students correctly suggest that the particular instances they have identified belong to a much broader generic pattern, others seem to see a specific occurrence they had found as a generic pattern in itself, without considering the possibility that the sequence in question (often a recurrent, frequently-occurring fragment or 'n-gram') can in fact be seen as a more productive grammatical structure, or variable lexical paradigm. And while some students do identify valid lexico-grammatical patterns (including variable paradigms), many nevertheless have trouble in assigning systematic labels in order to describe the discourse functions of these sequences.

Before we conclude, it would be informative to briefly look at some of the comments our students made about this part of their terminology project. Although out of context, some of their choices appear to be mistaken, in reality there is a logic underlying even the most unusual patterns. For example, Student 1 chose to analyse a sequence which we would consider to be more terminological than phraseological (example #47 *serious games for*). Clearly, there can be no rhetorical function for this incomplete segment. Nevertheless, as this student states, it is an extract from a more extended pattern:

> "La collocation "serious games for" est retrouvée fréquemment dans le corpus anglophone. Elle détermine le but d'un serious game, à quoi il va servir. Voici quelques exemples: Serious games for learning, serious games for vocabulary education, serious game for training of collaboration skills" (Student 1)

> 'the collocation *serious games for* occurs frequently in the English corpus. It determines the objective of a *serious game,* what it is going to be used for. Here are a few examples: *Serious games for learning, serious games for vocabulary education, serious game for training of collaboration skills*" (Student 1 [our translation].)

Here, the pattern *serious games for* indicates a productive structure for creating subordinate terms (co-hyponyms) for *serious games*. This shows that, even if our students are training to become translators, it is necessary to help them with their handling of linguistic structures, in order to distinguish terminology from phraseology.

Other students are more clearly aware of the discourse functions of the sequences they have identified and are able to disentangle them efficiently from terminological issues. Student 2 for example discusses how logical connections can be expressed by a variety of short phrases:

> '...des tournures comme « due to the presence of », « plays an important role in » et « as a function of » sont employées pour exprimer des actions ou des liens fonctionnels entre les éléments du texte. Elles sont donc situées à diverses positions au sein des phrases. Ce type de collocation sert à la construction et à la cohérence du discours à un niveau plus interne que les collocations utilisées au niveau argumentatif.....' (Student 2)

> '...phrases like '*due to the presence of, plays an important role in*' and *'as a function of*' are used to express actions or functional links between different text elements. They are therefore found in different positions in the sentence. This type of collocation aids in the construction of coherence at a deeper level than collocations used at the argumentative level... (Student 2 [our translation])

And then there are very capable students who are perhaps justifiably confused by the idea of 'generic collocations'. From the following, it is clear that Student 3 (writing in English) does not feel that it is valid to analyse patterns which do not belong to her specialist domain:

> 'The following generic collocations [are] from both the corpus and my source text. I was somewhat puzzled by this task, as the generic collocations in my text, such as *Dans les années récentes* [In recent years] that serve to punctuate the narrative more than to provide information are not specific to the discourse of my field.' (Student 3).

What conclusions can be drawn from comments such as these? In the light of these comments and of our general findings set out above, we clearly need to re-evaluate not only our teaching methods, but also perhaps the way in which we present and conceptualise such notions as 'generic collocation' and 'discourse function'. Although these points are essentially pedagogical, the issues involved are also clearly related to more theoretical issues about dealing with different types of phraseological phenomena, especially regarding the notion of collocation in Languages for Specific Purposes (LSP).

As discussed in the first half of this study, in order to identify a lexico-grammatical pattern in terms of structure and discourse function, it is necessary to examine a representative selection of examples on the basis of corpus analysis. However, this is a time-consuming activity and it requires research-oriented skills. In the M2 ILTS course, we attempt to provide both time and training in corpus-informed analysis. However, it occurs to us that the use of concordance software provides such a wealth of data that students are sometimes overwhelmed. This may then lead them to focus on patterns of expression which emerge purely on the basis of frequency from software such as AntConc. So even though the segments may be statistically valid, the sheer quantity of data may lead students to neglect the much more detailed analysis of co-text and context, which is necessary in order to identify useful lexico-grammatical patterns.

Furthermore, our students are sometimes confused about what constitutes a specific instance (an example) and what qualifies as more general, underlying pattern. An anonymous reviewer of this study very appositely characterised these responses as a 'downwards' and an 'upwards' problem of analysis. We would suggest that when these phenomena are presented to students, a clearer distinction should be made between the 'specific instance' (which serves as evidence of an overall pattern) and the 'generic pattern' itself (i.e. a pattern which, although abstract, should nevertheless involve a named grammatical item or structure as well as a paradigm of related lexical items). In other words, we are arguing that currently fashionable 'bundles', 'clusters', 'n-grams' etc. are good tools for throwing up statistically significant fragments of phrases, but this corpus-based methodology needs to be supplemented by the systematic recognition and analysis of what we (and other corpus-informed linguists) have called 'lexico-grammatical patterns'.

The final point involves the fact that our students are often unable to analyse 'discourse functions' in a systematic way. As we believe that this notion is crucial – indeed a discourse function is the defining feature of each lexico-grammatical pattern – then we may need to re-think our system of analysis. One such way of doing this, as we have suggested in the final section of this study, may be to use the very systematic, although also rather elaborate system of analysis inspired by a model such as Systemic Functional Grammar (SFG). We do not currently teach this model of analysis to our students. After all, our students are not specialists in linguistics and they have chosen to work in the world of professional translation (or other areas of language industry). It may not be appropriate for them to learn a new, sometimes very complex model of analysis. But SFG has its origins in language teaching and applied linguistics and it has been used successfully in a variety of highly practical contexts. What is more, in our

opinion, there is no reason why language professionals should not also acquire a range of transferable research skills. Such key competencies might include: (a) the intellectual skills necessary to design a research project, (b) the computational and technical ability to conduct corpus-informed analysis and (c) the terminology and metalanguage necessary to examine and interpret corpus-informed data in relation to a systematic model of language.

Acknowledgment

The authors would like to thank two anonymous reviewers and the editors and for their detailed remarks on an earlier version of this study. Any errors remaining are the sole responsibility of the authors.

Author's Contributions

Both Christopher Gledhill and Natalie Kübler contributed fully to the data collection, data analysis and the writing of this study.

Ethics

Permission has been obtained from all the participants in this study to provide short anonymous citations of their work for research purposes. The numbered examples cited in this study are all available online, or in the corpora and databases cited in the text.

References

Anthony, L., 2002. A Machine learning system for the automatic identification of text structure and application to research article abstracts in computer science. PhD Thesis, University of Birmingham, Birmingham.

Banks, D., 1994. Writ in Water: Aspects of the Scientific Journal Article. Université de Bretagne occidentale: Equipe de recherche en linguistique appliquée.

Biber, D., S. Conrad and V. Cortes, 2004. If you look at …: Lexical Bundles in University teaching and textbooks. Applied Linguist., 25: 371-405. DOI: 10.1093/applin/25.3.371

Brinton, L. and E.C. Traugott, 2005. Lexicalization and Language Change. 1st Edn., Cambridge University Press, Cambridge, ISBN-10: 0521540631, pp: 220.

Cheng, W., C. Greaves, J. M.H. Sinclair and M. Warren, 2008. Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. Applied Linguist., 30: 236-252. DOI: 10.1093/applin/amn039

Firth, J.R., 1957. Papers in Linguistics, 1934-1951. 1st Edn., Oxford University Press, Oxford, pp: 233.

Frath, P. and C. Gledhill, 2005. Qu'est-ce Qu'une Unité Phraséologique? In: La Phraséologie dans tous ses états, Bolly, C., J.R. Klein and B. Lamirov (Éds.), Louvain-La Neuve, pp: 11-25.

Gledhill, C., 1995. Collocation and genre analysis: the phraseology of grammatical items in cancer research abstracts and articles. Zeitschrift fiir Anglistik und Amerikanistik, 43: 11-36.

Gledhill, C.J., 2000. Collocations in Science Writing. 1st Edn., Gunter Narr Verlag, Tübingen Narr, ISBN-10: 3823349457, pp: 268.

Gledhill, C., 2011. The 'lexicogrammar' approach to analysing phraseology and collocation in ESP texts. Anglais de Spécialité, 59: 5-23. DOI: 10.4000/asp.2169

Gledhill, C., 2012. The discourse function of collocation in research article introductions. In: Benchmarks in Language and Linguistics, Biber, D. and R. Reppen (Eds.), Sage Publications, London, pp: 23-45.

Goldberg, A.E., 1995. Constructions: A Construction Grammar Approach to Argument Structure. 1st Edn., University of Chicago Press, Chigago, ISBN-10: 0226300862, pp: 271.

Groom, N., 2005. Pattern and meaning across genres and disciplines: An exploratory study. J. English Acad. Purposes, 4: 257-277. DOI: 10.1016/j.jeap.2005.03.002

Halliday, M.A.K. and C.M.I.M. Matthiessen, 2014. Introduction to Functional Grammar. 4th Edn., Routledge, London, ISBN-10: 1444146602, pp: 808.

Hasselgren, A., 2002. Learner Corpora and Language Testing: Small Words as Markers of Learner Fluency. In: Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching, Granger, S., J. Hung and S. Petch-Tyson (Eds.), John Benjamins Publishing, Amsterdam, pp: 143-173.

Humbley, J., 2001. Quelques enjeux de la dénomination en terminologie/Some issues in term formation. Cahiers de Praxématique, 31: 93-115.

Hunston, S. and Francis, Gilll 2000. Pattern Grammar. 1st Edn., John Benjamins, Amsterdam.

Hunston, S., 2008. Starting with the small words: Patterns, lexis and semantic sequences. Int. J. Corpus Linguist., 13: 271-295. DOI: 10.1075/ijcl.13.3.03hun

Hyland, K., 2008. As can be seen: Lexical bundles and disciplinary variation. English Specific Purposes, 27: 4-21. DOI: 10.1016/j.esp.2007.06.001

Kübler, N. and M. Pecman, 2012. The ARTES Bilingual LSP Dictionary: From Collocation to Higher Order Phraseology. In: Electronic Lexicography, Granger, S. and M. Paquot (Eds.), Oxford University Press, Oxford, pp: 186-208.

Kübler, N. and A. Volanschi, 2012. Semantic prosody and specialised translation, or how a lexico-grammatical theory of language can help with specialised translation. In: Corpus-Informed Research and Learning in ESP: Issues and applications, Boulton, A., S. Carter-Thomas and E. Rowley-Jolivet (Eds.), Studies in Corpus Linguistics, pp: 103-134.

Lee, D.Y.W. and C. Xiao, 2008. Small words, big deal: Teaching the use of function words and other key items in research writing. Proceedings of the 8th Teaching and Language Corpora Conference, (LCC' 08), ISLA, Lisbon, pp: 198-206.

Legallois, D. and J. François, 2006. Autour des grammaires de constructions et de patterns. Presses de l'Université de Caen, Caen.

Luzón-Marco, M.J., 1999. The phraseology and meanings of the pattern be+adjective + to-infinitive. La Linguistique, 35: 47-60.

Martin, J. and D. Rose. 2003. Working with Discourse. Meaning Beyond the Clause. 1st Edn., Continuum, London.

Mhedbhi, M., 2014. Lexical Bundles and the construction of an academic voice in business writing. Adv. Lang. Literary Stud., 5: 1-9.

Pavel, S., 1993. La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques. Terminol. Nouvelles, 10: 23-35.

Pearson, J., 1998. Terms in Context. 1st Edn., John Benjamins, Amsterdam.

Pecman, M., 2012. Etude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques. J. Specialised Translat., 18: 113-138.

Renouf, A. and J. Sinclair, 1991. Collocational Frameworks in English. In: English Corpus Linguistics, Aijmer, K. and B. Altenberg (Eds.), Longman, London, pp: 128-143.

González Rey, I., 2005. L'espace Réservé à la Phraséologie Dans la Didactique du FLE. In: Espace et Texte Dans la Culture Française, Ramos, A.S. (Éds.), Université d'Alicante, Alicante, pp: 1421-1439.

González Rey, I., 2008. Le rôle de la phraséologie dans la mise en discours de la langue juridique. In: Aspectos Formales y Discursivos de Las Expresiones Fijas, Tarrío, G.C. (Eds.), Frankfurt am Main, Berlin, Bern, Bruxelles, New York, pp: 121-140.

Scott, M. and C. Tribble. 2006. Textual Patterns: Keyword and Corpus Analysis in Language Education. 1st Edn., Benjamins, Amsterdam.

Stefanowitsch, A. and S. Gries, 2003. Collostructions: Investigating the interaction of words and constructions. Int. J. Corpus Linguist., 8: 209-243. DOI: 10.1075/ijcl.8.2.03ste

Williams, G.C., 1998. Collocational networks: Interlocking patterns of lexis in a corpusof plant biology research articles. Int. J. Corpus Linguist., 3: 151-71. DOI: 10.1075/ijcl.3.1.07wil

Williams, G. 2003. Les Collocations et L'école Contextualiste Britannique, In: Les Collocations: Analyse et Traitement. Travaux et Recherches en Linguistique Appliquée, Francis, G. and A. Tutin (Eds.), Amsterdam DeWerelt, pp : 33-44.

Appendix 1

Table 1. 50 Lexico-grammatical patterns from 10 Student Reports

No.	Pattern[8]	Lexico-grammatical Structure[9]	Discourse function (according to students)	Discourse system (according to SFG)
#1	Failure to + VG (comply, follow, do so, heed, observe) +NG + will result in +NG (Not supplied)	Effective clause (Subject includes embedded expansion clause)	Ideational: material, logical ('so')	Interpersonal: engagement: source (projection 'failure to', modality 'will') Textual: identification ('so')
#2	We reserve the right to VG	Effective clause (Complement includes embedded expansion clause)	(Not supplied)	Ideational: Relational Interpersonal: Engagement: Source Textual: identification ('we')
#3	Previous works focus on...	Effective clause (Subject Predicate active)	'Talking about the subject of the present study/'Talking about one's position or the theoretical context to which the study belongs'	Ideational: Relational Interpersonal: Engagement: Source Textual: identification
#4	This paper describes...	Effective clause (Subject Predicate active)	'Talking about the subject of the study'	Ideational: mental/communicative Interpersonal: Engagement: Source Textual: Identification
#5	Figure X depicts...	Effective clause (Subject Predicate active)	'Referring to non-textual elements (tables, graphs,	Ideational: mental / communicative Interpersonal: engagement: source

No.	Pattern[8]	Lexico-grammatical Structure[9]	Discourse function (according to students)	Discourse system (according to SFG)
			figures...'	Textual: identification
#6	Finally, we consider...	Effective clause with Adjunct in Marked Theme position	'Referential introductory phrase'	Ideational: Mental Interpersonal: Engagement: Source Textual: Periodicity
#7	Next on our agenda is...	Effective clause with Adjectival phrase in Marked Theme position	'Impersonal introductory phrase'	Ideational: Relational Interpersonal: Engagement: Source Textual: Periodicity
#8	In this paper we outline...	Effective clause with Prepositional phrase in Marked Theme position	'Referential introductory phrase'	Ideational: Mental, communicative Interpersonal: Engagement: Source Textual: Periodicity and identification
#9	This paper was supported in part by...	Receptive clause (Subject Predicate passive)	'Partial phrasal pattern'	Ideational: Material Interpersonal: Engagement: Source Textual: identification
#10	The analysis is based on...	Receptive clause (Subject Predicate passive)	'Presenting methods, tools, approach, technique'.	Ideational: relational Interpersonal: N/A Textual: Periodicity (prospective) and identification
#11	Proceed as follows ...	Imperative clause (with Predicate as Theme)	(Not supplied)	Ideational: Material Interpersonal: N/A Textual: Identification (*as follows)* and periodicity (imperative)
#12	On estime que... 'it is thought that'	Projecting clause	Verbal construction/ 'Scientific discourse'	Ideational: mental Interpersonal: Engagement: Source/ projection Textual: N/A
#13	On observe que... 'it is observed that'	Projecting clause	'Impersonal introductory phrase'	Ideational: Mental Interpersonal: Engagement: Source/ projection Textual: N/A
#14	X shows that...	Projecting clause	'Demonstrative discourse'	Ideational: mental Interpersonal: Engagement: Source/ projection Textual: Identification
#15	Les travaux ont démontré que... 'Studies have shown that'	Projecting clause	'Expressing anteriority/ Making an anonymous reference'	Ideational: Mental Interpersonal: Engagement: Source/ projection Textual: Identification
#16	Let us say ...	Imperative Projecting clause (Predicate as Theme)	'Verbal construction: Introducing a hypothesis/ Making estimations, calculations and interpretations'	Ideational: Mental/communicative Interpersonal: Engagement: Source/ projection Textual: Identification
#17	Il s'ensuit que... 'It follows that...'	Projecting clause (empty Subject)	(Not supplied)	Ideational: Logical connection Interpersonal: Engagement: Source/ projection Textual: Periodicity
#18	Supposer que... 'To suppose that'	Projecting clause (non-finite)	'Partial phrasal pattern'	Ideational: mental Interpersonal: engagement: source/ projection Textual: N/A
#19	Force est de constater... 'It must be acknowledged (that)'	Projecting clause with complement clause (with Nominal group as Marked Theme)	'Verbal construction/ Academic discourse'	Ideational: Mental Interpersonal: Engagement: Source/ projection Textual: N/A
#20	The intent of this article is to provide...	Projecting clause with Complement clause	'Presensting one's research goals'	Ideational: Relational Interpersonal: Engagement: Source/ projection Textual: periodicity & identification
#21	It is worth noting (that)...	Projecting clause with postponed clause	'Highlighting a compatibility, correlation, analogy' 'Expressing a notion of restriction or specification.' 'Expressing an addition.' 'Describing, interpreting and analysing data or observed phenomena.' 'Talking about characteristics, properties, specificities.'	Ideational: Relational + mental Interpersonal: Attitude: Evaluation Textual: N/A

No.	Pattern[8]	Lexico-grammatical Structure[9]	Discourse function (according to students)	Discourse system (according to SFG)
#22	Il est intéressant de constater ... 'it is worth noting (that)'	Projecting clause with postponed clause	Verbal construction/ 'Multi-register discourse'	Ideational: Relational + mental Interpersonal: Attitude: Evaluatio Textual: N/A
#23	Reste à comprendre... 'It remains to be seen (whether)	Projecting clause with postponed clause (as Subject)	(Not supplied)	Ideational: Relational + mental Interpersonal: Engagement: Projection Textual: N/A
#24	It is widely known that...	Receptive (passive) Projecting clause with postponed clause	'Referential introductory phrase'	Ideational: Relational + mental Interpersonal: Engagement: Source/ projection Textual: N/A
#25	It can be seen that...	Receptive (passive) Projecting clause with postponed clause	'Impersonal introductory phrase'	Ideational: mental Interpersonal: engagement: source/ projection Textual: N/A
#26	It should be noted that..	Receptive (passive) Projecting clause with postponed clause	'Impersonal introductory phrase' 'Talking about characteristics, properties, specificities' 'Expressing a necessity'	Ideational: mental Interpersonal: Engagement: Source/ projection Textual: N/A
#27	Have been shown to + V...	Partial (Subject-less) projecting clause (receptive / passive)	'Verbal construction/ Scientific discourse'	Ideational: Mental, Interpersonal: Engagement: Source Textual: Periodicity and identification
#28	Known to be...	Partial (Subject-less) projecting clause (receptive/passive)	'Verbal construction/ Scientific discourse'	Ideational: Mental, Interpersonal: Engagement: Source/ projection Textual: N/A
#29	Comme le montre la figure x... 'As shown by figure X'	Expansion clause	'Adverbial construction/ Making reference to non-textual elements (tables, graphs, figures...)'	Ideational: Mental/communicative Interpersonal: Engagement: Source Textual: Identification
#30	When considering...	Reduced Expansion clause	'Referential introductory phrase'	Ideational: Mental, logical connection Interpersonal: N/A Textual: N/A
#31	As stated in...	Reduced Expansion clause	'Partial phrasal pattern'	Ideational: Mental/communication Interpersonal: Engagement: Source Textual: identification
#32	In the same way that...	Prepositional phrase (introducing an expansion clause)	'Prep + NG + conj' 'Emphasising compatibility, orrelation, analogy, similarity'	Ideational: logical connection/ comparison Interpersonal: N/A Textual: Identificatio
#33	A critical issue in handling x...	Noun group (modified by embedded expansion clause)	'Discussing difficulties, problems or limits encountered'	Ideational: N/Al Interpersonal: Attitude: Evaluation Textual: N/A
#34	Studies undertaken to date...	Noun group (modified by embedded expansion clause)	'Talking about one's position or the theoretical context to which the study belongs''	Ideational: N/A Interpersonal: Engagement: Source Textual: Identification
#35	Les études menées jusqu'ici... 'Studies undertaken to date'	Noun group (modified by embedded expansion clause)	'Talking about one's position or the theoretical context to which the study belongs''	Ideational: N/A Interpersonal: Engagement: Source Textual: Identification
#36	The assumption underpinning the concept...	Noun group (modified by embedded expansion clause)	'Talking about one's position or the theoretical context to which the study belongs''	Ideational: N/A Interpersonal: Engagement: Source Textual: Identification
#37	To remain an unsolved challenge for...	Verbal phrase (Predicate plus Complement)	'Partial phrasal pattern'	Ideational: Relational Interpersonal: Attitude: Evaluation Textual: identification
#38	Trouver dans la littérature... 'to find in the literature'	Verbal phrase (Predicate plus Complement)	(Not supplied)	Ideational: mental Interpersonal: engagement: source Textual: identification
#39	Recommended by...	Verbal phrase (Predicate without Complement)	'Adjective + preposition' 'Technical discourse'	Ideational: Mental Interpersonal: Engagement: Source Textual: idenTification
#40	To this end...	Prepositional phrase (functioning as Adjunct)	'Presenting one's research goals'	Ideational: N/Al Interpersonal: N/A Textual: Identification, periodicity
#41	Pour une synthèse... 'For a summary' (see...)	Prepositional phrase (functioning as Adjunct)	'Verbal construction' 'Academic discourse'	Ideational: N/A Interpersonal: Engagement: Source Textual: Identification
#42	Due to the presence of...	Partial prepositional phrase (missing Nominal group)	'Adjective + preposition' 'Technical discourse'	Ideational: Logical connection Interpersonal: Textual: N/A

No.	Pattern[8]	Lexico-grammatical Structure[9]	Discourse function (according to students)	Discourse system (according to SFG
#43	In accordance with...	Partial prepositional phrase (missing Nominal group)	'Presenting methods, tools, approach, technique'.	Ideational: Logical connection Interpersonal: Engagement: Source Textual: N/A
#44	Key insight...	Nominal group	'Talking about the specifics of the present study' 'Making empirical observations'	Ideational: I Interpersonal: Graduation: Force Textual: Identification, Periodicity
#45	Safety precautions...	Nominal group	(Not supplied)	Ideational: N/A Interpersonal: N/A Textual: N/A
#46	Related work...	Nominal group	'Nominal construction' 'Talking about the subject of the study'	Ideational: Interpersonal: Engagement: source Textual: Identification
#47	Serious games for...	Nominal group postmodified by preposition	(Not supplied: see discussion in main text)	Ideational: N/A Interpersonal: N/A Textual: N/A
#48	N-driven game...	Nominal group pre-modified by reduced embedded clause	(Not supplied: see discussion in main text)	Ideational: N/A Interpersonal: N/A Textual: N/A
#49	Due to...	Complex prepositional group	'Adjective + preposition' 'Technical discourse'	Ideational: Logical connection Interpersonal: N/A Textual: N/A
#50	In parallel with...	Complex prepositional group	'Explaining the conditions in which the analysis takes place'	Ideational: Logical connection Interpersonal: N/A Textual: N/A

[8]The analysis in columns 1 and 3 is that of our students. Note also that in the ARTES data base, discourse functions are described in French. To save space here, we have presented this information in English.
[9]The analysis in columns 2 and 4 follows the conventions of Systemic Functional Grammar (c.f. Halliday and Matthiessen 2014 and Martin and Rose 2003). See the main text for details.