The 'lexicogrammar' approach to analysing phraseology and collocation in ESP texts

Christopher Gledhill

This text was automatically generated on 10 December 2020.

The 'lexicogrammar' approach to analysing phraseology and collocation in ESP texts

Christopher Gledhill

1. Introduction

1   The aim of this paper[1] is to examine the notions of phraseology and collocation in the field of English for Specific Purposes (ESP) and to recast these terms from the point of view of Systemic Functional Linguistics (SFL). Broadly speaking, **phraseology** involves the study of formulaic sequences of words, including idiomatic phrases and proverbial expressions, which stand in contrast to other more prosaic constructions in the language in that they have a highly conventionalised form and frame of reference. For example, the rhetorical impact of the phrase *(to) cut (one's) losses* (cited in sample text T1 in the Appendix) cannot quite be captured by paraphrases such as: *accept what one has lost and move on, stop doing something in order not to make a bad situation worse,* etc. Whereas phraseology is phrase-oriented and rhetorical (involving a contrastive choice between marked phrases and their unmarked paraphrases), the notion of **collocation** is essentially word-oriented and cohesive: it refers to the extent to which the presence and meaning of a word 'coheres' or depends on the presence of another word (or words) in the same stretch of text. For example, the noun *loss* refers to 'debit, decrease in revenue' in contexts such as *to cut one's losses* and *to make a loss,* whereas *loss* refers to 'bereavement, death' when used in the context of verbs such as *mourn, regret, suffer,* etc.

2   In the first half of this paper, I contrast the traditional, lexicological approach to phraseology and collocation with the 'lexicogrammar' approach adopted by the proponents of Systemic Functional grammar (Halliday 1961, Halliday & Matthiessen 2004). The notion of lexicogrammar encompasses a much broader set of phenomena than are usually considered in mainstream lexicology. In the final sections of this paper, I demonstrate this by showing that it is possible to explore the

lexicogrammatical properties of high-frequency, closed-class grammatical items (also called 'small words'). My point is that individual grammatical signs not only enter into collocational relations, but also form relatively predictable and productive chains of expression, with one construction cascading into another. It can be shown that these extended lexical patterns are often unique to a particular register or genre. I would argue that the identification of such patterns should be a fundamental step in the systematic analysis of ESP texts. For demonstration purposes, throughout this paper I refer to examples taken from two related pieces of science writing on *Genomic Imprinting* (set out in the Appendix: T1 an extract from a popular science book by Dan Dennet, T2 an abstract from a research paper by David Haig).

2. Lexicogrammar and lexicogrammatical patterns

3   The term **lexicogrammar** refers to two distinct but related notions: (1) the typical lexical and grammatical environment of a sign as it is habitually used in naturally occurring texts or 'discourse', and (2) the core stratum of 'wording' in Michael Halliday's model of language, which serves to mediate between the lower stratum of 'sounding' (graphology/phonology) and higher 'meaning' (semantics/discourse). As this notion was first developed in the framework of Systemic Functional Linguistics (SFL) (Halliday 1961, Fries *et al.* 2002, Halliday & Matthiessen 2004), it is important to set out here some of the core features of the SFL approach.

4   One of the central tenets of SFL is that lexis (a structured system of signs which serves to organise the vocabulary of a language) and grammar (a structured system of choices which serves to organise sequences of signs into texts) are not different in nature, but rather form a unified stratum in the language: the lexicogrammar. A further central assumption of SFL, following Firth (1957), is that no aspect of lexis or grammar can be properly defined without reference to its typical context of use (or 'co-text') that is to say in actual stretches of texts or discourse. It follows from this that SFL rejects the structuralist view that the abstract system of language (*langue*) is independent from language in use or discourse (*parole*). Rather, the language system is constantly interacting with and being shaped by different types of speech event (the 'context of situation') within a community of speakers (the 'context of culture'). Another way of putting this, following Martin (2001), is to say that everything in language, from lexical items and grammatical constructions to whole texts, has evolved to express very specific discourse functions, in the form of situational 'registers' (the lexicogrammatical resources associated with a specific speech activity, such as *impersonal expressions, nominal style, taxonomies of terms,* etc.), as well as 'genres' (goal-oriented, culturally specific speech activities, such as *conversation on a scientific topic, exposition in popular science, narration in a research article,* etc.). It is this focus on the underlying communicative functions of language and the systemic choices that are made available by the language system that make SFL distinct from other models of language.

5   It follows from what has just been written that the SFL viewpoint on phraseology and collocation is very different from that of mainstream lexicology. Lexicographers and other analysts typically conceive of phraseological phenomena in terms of a continuum that ranges from 'free combinations' at one end to 'fixed phrases' at the other. Here is how Howarth (1996) puts it:

> [...] a 'scale of idiomaticity', ranging from the most freely co-occurring lexical items and transparent combinations to [...] the most cast-iron and opaque idiomatic expressions. [... It] is desirable for purposes of efficiency to eliminate from the description those combinations whose co-occurrence can be accounted for by normal grammatical and syntactic processes. (Howarth 1996: 32-47)

6 The SFL approach is diametrically opposed to this view of language. Firstly, SFL assumes that any normal construction in the language can potentially be promoted to the rhetorical status of idiom, and there is thus no need to establish a separate category of phraseological unit outside the lexicogrammar (this point is discussed in terms of lexicalisation, below). Secondly, Howarth's notion of free combination supposes that grammatical rules or structures operate independently from lexical signs or lexical relations. The lexicogrammar approach assumes instead that even the most mechanical or abstract grammatical process depends on lexical relationships and has a lexical realisation (e.g., the grammatical mechanism of 'raising' depends on cognitive, reporting verbs as in the pattern *N has been {found, shown, thought} to V*.) In this respect, it is useful to return to Firth's (1957) original conception of collocation, which states that all signs in the language are mutually dependent on and mutually defined by the other signs with which they are habitually used within actual stretches of text:

> Words must not be treated as if they had isolate meaning and occurred and could be used in free distribution. (Firth 1968b: 18)
> The collocation of a word or a 'piece' is not to be regarded as mere juxtaposition, it is an order of *mutual expectancy*. The words are mutually expectant and mutually prehended. (Firth 1957: 181)

7 The main objects of study from an SFL perspective are thus not phraseological units or grammatical constructions, but rather lexicogrammatical (LG) patterns (Stubbs 1995, Hunston & Francis 1998, Tucker 1998, Legallois & François 2006). **Lexicogrammatical patterns** have the following properties:

8 - a LG pattern is a predictable but also productive sequence of signs, which as a whole shares a stable, coherent frame of reference;

9 - a LG pattern can be composed of lexical signs, or more abstract signs, including grammatical morphemes and constructions;

10 - a LG pattern is composed of permanent 'pivotal' signs and a more productive 'paradigm', a feature which allows the pattern to be reformulated and integrated into other patterns and thus into on-going discourse;

11 - a LG pattern may extend over a long stretch of text, it may be discontinuous and it may or may not be a syntactic constituent or phrase.

12 It is possible to explore some specific examples of LG patterns that occur in the research article abstract (T2, see Appendix), such as the sequences *mount (a) response* and *gene + express*. Using a Web browser, it is possible to find over 16,000 examples of *mount a response* in texts relating to molecular biology, including:

> (1) Patients with muscular dystrophy **mount immune response to** dystophin protein prior to gene therapy.
> (2) Target cells however **mount a response to** such membrane damage...
> (3) [...] the host might **mount a response against** the cancer cells...
> (4) Pure-bed S. salar were susceptible but frequently **mounted a response to** G. salaris without eliminating the infection.

> (5) We describe an HA-A1 melanoma patient who has **mounted a** spontaneous cytolytic T cell (CTL) **response against** an antigenic peptide encoded by gene MAGE-A3 and presented by HLA-A1.

13   From the point of view of the lexis, *mount (a) response* is 'lexicalised' phrase, an extended lexical sign in which there is only a small degree of variation. For example, an on-line search for a passive sequence such as *response (is, was) mounted* reveals only four occurrences, suggesting that the pattern is relatively invariable. From a grammatical point of view, the pattern involves a Predicator[2] *mount*, which expresses a 'light' or generic Material Process (change, create), plus a Complement *response* which specifies the type of Process expressed by the verb (for Halliday & Matthiessen 2004, its semantic role is 'Process Range'). Finally, it is important to point out that the pattern is not restricted to the pivotal elements *mount (a) response*: in its wider context, it also includes a relatively stable set of Subjects (*cells, hosts, patients*) and an 'indirect' Complement (introduced by *against* or *to*), which is in effect the main (Affected) Participant of the clause.

14   The sequence *gene + express* involves a much more productive set of LG patterns, as the following examples suggest (these are taken from the 500,000 word Pharmaceutical Sciences Corpus (PSC), reported in Gledhill 1995, 1997):

> (6) Under these conditions, we did not detect PAF-R **gene expression** (Ma and Bazan, 2000).
>
> (7) However, **expression of the gene** was not confined to the hair follicle, as the transgene phenotype included not only hair abnormalities, but also vertebral defects and bladder, liver and intestinal tumors.
>
> (8) In the present study, we report our attempt to identify differentially **expressed genes** with respect to the confluence/proliferative status of MGH-U3 cells in culture.
>
> (9) [...] level was determined semiquantitatively by calculating the ratio of density metric value from specific **genes expressed** in relation to the internal standard
>
> (10) Results: the Muc2 mucin **gene was expressed** in middle ear mucosa of the control rats.

15   The signs *gene + express* occur in two basic LG patterns. The first involves a nominalisation, in which *gene* is a (pre-modifying) Classifier or (post-modifying) Qualifier of a nominalised Process (*gene expression, expression of the gene*). In these contexts, the emphasis is on the investigation or observation of a 'metaphorical' (nominal, static) process (*we did not detect, was not confined to...*). In the second pattern, *gene* is typically post-modified by an embedded passive clause, or is the Subject of a passive (examples 8-10). In these contexts, the emphasis is on explaining the physical or genetic location of a 'congruent' (verbal, dynamic) process. In both patterns, the implicit semantic role played by *gene* is not Agent but rather Medium (Halliday & Matthiessen 2004), the location or vehicle in which the self-regulating Process of *expression* takes place.

16   It is interesting to note that these examples represent two fairly typical perspectives that can be adopted in science writing. In the LG patterns typically associated with *gene + express(-ed, -ion),* there is no explicit Agent. In contrast, the LG pattern *mount (a) response* always involves an Agent: it is either the host's *cells*, the *host* or more generally the *patient*. In the contexts above (1-5) *mount (a) response* appears to be a deliberately

dramatic choice of expression, and in text T2 this fits in coherently with the other conflictual metaphors used throughout the rest of the text.

3. Phraseology and lexicalisation

17   Having set out the main principles of the lexicogrammar approach, it is now worth revisiting the well-known terms 'phraseology' and 'collocation'. One of the principal assumptions of the traditional lexicological approach is that phraseological phenomena generally correspond to lexical units. This is reflected in the terminology of phraseology studies, especially the **phraseological unit** (PU) – in contrast to 'phraseologism' and 'phraseme', which are used differently. The prototypical examples of PUs studied in the literature tend to be idioms *it's raining cats and dogs*, catchphrases *the rain in Spain stays mainly on the plain,* proverbs *it never rains but it pours,* and the like. These kinds of phrases are clearly essential to the cultural life of a language. However, examples such as these give the impression that PUs generally have an idiosyncratic structure or meaning. They also suggest that PUs correspond to fully-formed constituent phrases or clauses.

18   There have been few studies on phraseological units in ESP and science writing, at least in the traditional 'idiom-oriented' sense of the term. The exception perhaps lies in the areas of LSP, terminology and translation studies (Pavel 1993, Fiedler 2007). However, not all phraseological studies adopt this perspective, or indeed refer to phraseological units. An alternative approach has emerged in discourse analysis (Gréciano 1997, Tollis 2001, Gonzalez-Rey 2002, Gledhill & Frath 2007) and corpus-based lexicography (Moon 1994, Fernando 1996, Hunston & Francis 1998, Pecman 2005). On the basis of empirical evidence, these analysts emphasise the fact that idiomatic expressions change over time, have variable interpretations in on-going discourse, and are often reformulated or serve as the basis for new constructions. Similarly, analysts working in psycholinguistics and language acquisition (Wray 2002, Jones & Haywood 2004, Granger & Meunier 2008) refer to 'formulaic sequences', a term which can be applied to the invariable sequences encountered in children's speech (*allgone*) or in conversation (*d'you know what I mean?*).

19   Rather than concentrate on the notion of 'idiomaticity' or on specific types of phraseological phenomena, it may be more relevant to those working in the SFL perspective and areas such as ESP to refer to more general, underlying processes. An important notion to emerge recently in cognitive and comparative linguistics involves **lexicalisation**, the historical process of language change in which a sequence of signs gradually coalesces in structure and in sense to become a single sign. Brinton & Traugott (2005) claim that this process involves a continuum ranging from L1: partially productive lexicalised compounds and phrases (*airbrush, to bear witness, cutting-edge*), through L2: non-productive lexicalised composites (*auburn hair, with bated breath, to curry favour*) and finally L3: fully lexicalised items (*altogether, breakfast, causeway*). It is important to note that although lexicalisation is defined in the same terms as idioms and other phraseological phenomena, the process potentially involves a much broader set of patterns:

> Lexicalization is the change whereby in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the

constituents of the construction or the word formation pattern. Over time there
may be further loss of internal constituency and the item may become more lexical.
(Brinton & Traugott 2005: 96)

20   Various examples of lexicalisation can be seen in the sample texts T1 and T2. As might
be expected, there are few phraseological units in the traditional sense of the term in
these texts, except perhaps some stereotypes or clichés *the conflict plays out, cut her
losses* in T1 (these appear to be quite appropriate to a popular science account)*.*
However, in the same text there are also a variety of lexical frameworks (also known as
'sentence stems') *it is the embryo's best interests...that...,* locutions or lexicalised verb
phrases *given the choice, taking whatever steps are available,* lexicalised adverbial/
prepositional phrases *and so on, of course, on the one hand...on the other* and lexicalised
noun groups *by-product, tug-of-war, trying circumstances.* Similarly, text T2 (the more
'serious' research article abstract) does not contain any clear examples of
phraseological units. Instead, there are many examples of partially lexicalised noun
groups *blood glucose levels, natural selection* and lexicalised verb groups such as the
examples examined above, *genes expressed as, mount a response.*

4. Collocation and cohesion

21   Unlike phraseological units, there has been a long tradition of studies on collocation in
applied linguistics, especially in the fields of English for Academic Purposes (EAP) and
Language for Special Purposes (LSP) (Sager, Dungworth & McDonald 1980, Howarth
1996, Nesselhauf 2003, Williams 2003, Cavalla 2008) as well as in related areas such as
terminology (Gläser 1988, Béjoint & Thoiron 1992, Thomas 1993, Pearson 1998,
Grossmann & Tutin 2002, Tutin 2007). Many of these studies adopt a semantic
definition of collocation proposed by lexicologists such as Hausmann (1985) and
Mel'čuk *et al.* (1995). The key concept in this approach is the **lexical function**, a
privileged semantic relation between two lexical items in which one element retains its
core meaning as the 'base' while the other is a relatively restricted or metaphorised
'collocator.' For example, constructions such as *express a gene* and *mount a response* (to
take the examples from T2 examined above) are considered to be collocations because
they exploit a metaphorised, or in this case specialised sense of *express* ('to process
information in order to synthesise proteins or other gene products') or *response* ('a
hormonal defence mechanism'). In contrast, Predicator + Complement sequences such
as *produce hormones, provide nutrition, release hormones* (these examples are from text T1)
are considered to be simple 'combinations' (to use Howarth's 1996 term), because they
refer to one of the usual senses of a polysemous verb. This approach has been
particularly influential in LSP and terminology, as can be seen in the distinction often
made between 'LGP' and 'LSP' collocations (Sager *et al.* 1980, Benson *et al.* 1986, Howarth
1996).

22   The advent of computer-based corpus analysis has meant that many linguists use
statistical methods for identifying collocations as well as or instead of semantic
criteria. The statistical approach emphasises factors such as the frequency of co-
occurrence of lexical items (Smadja 1993, Stubbs 1995, Evert 2004), the distribution and
co-occurrence of collocations across text-types (Muller 1968, Williams 1998, Biber *et al.*
2004) and more recently the co-occurrence of lexical items and grammatical
constructions (Stefanowitsch & Gries 2003). Since corpus analysis is necessarily based

upon the observation of texts, a probabilistic approach has often been central to the lexicogrammar approach, as can be seen in this early definition from Halliday:

> Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur at *n* removes (a distance of *n* lexical items) from an item *x*, the items *a, b, c…* Any given item thus enters into a range of collocation, the items with which it is collocated being ranged from more to less probable. (Halliday 1961: 276)

23   In this light it is interesting to re-examine the collocations *express + gene* and *mount + response* mentioned in the previous section. The pair of signs *express(-ed, -es, -ing, -ion) + gene(s)* co-occur (i.e., occur together within a window of five words, right and left) 156 times in the PSC, and 263 times in the BNC. In contrast, the pair *mount (-ed, -ing, -s) + response(s)* does not occur in the PSC, and only occurs four times in the BNC. Since the BNC is twenty times the size of the PSC, there is proportionally a stronger rate of co-occurrence between *express + gene* than *mount + response.* But there are also many other ways of looking at these data. For example, using the AntConc program (Anthony 2007), it is possible to find all of the exactly repeated sequences ('clusters') that are formed within a given span of a single lexical item (related terms include 'N-Grams' and 'bundles', as reported in Biber *et al.* 2004). Thus within a span of five words, we can find 3,214 clusters for *gene(s)* in the PSC. It is interesting to note that common collocates of *gene* such as *express(-ed, -ion)* occur quite low down in the frequency list of clusters. This is because clusters, such as *expression of fibronectin gene was, expression of genes encoding biotransformation, expressed genes with respect to,* etc., mostly only occur once (although segments of the same cluster are also counted again, as parts of other clusters). This type of analysis shows that frequently co-occurring pairs of signs are not necessarily involved in strictly fixed sequences.

24   Finally, since collocation is now usually associated with the large-scale analysis of corpora, there has been less research on the role of collocation as a textual resource in individual texts. In this respect, it is appropriate to return briefly to Halliday and Hasan's (1976) view of collocation as a form of **cohesion**, that is to say a linking device that contributes to the overall coherence of a text. Halliday and Hasan (1976) originally made a distinction between grammatical forms of cohesion (reference, ellipsis, substitution, conjunction) and lexical cohesion, which involves explicit and implicit links between lexical signs, including such relations as reiteration, synonymy, complementarity, membership of ordered series or any other systematic lexical relationship, including 'collocation':

> *laugh…joke, blade…sharp, garden…dig* […]. In general, any two lexical items having similar patterns of collocation – that is, tending to appear in similar contexts – will generate a cohesive force if they occur in adjacent sentences. (Halliday & Hasan 1976: 285-6)

25   Unlike the other approaches to collocation mentioned here, Halliday and Hasan's definition is very informal and has not been generally taken up outside the SFL approach. Nevertheless, some analysts (Hoey 2005, Siepman 2005 and Gledhill 2009) have recently argued that a 'textual' approach to collocation would be a useful corrective to semantic and statistical approaches, which are essentially de-contextualised, and do not account for the role of collocation in on-going discourse. These analysts also point out that the lexical items that are usually involved in cohesive chains are necessarily embedded in lexicogrammatical patterns, whose distribution throughout a text must therefore also contribute to the development of coherence

throughout the text (Firth 1957 coined the term 'colligation' for this kind of relation). It is possible to observe this kind of development in the sample texts T1 and T2, in particular by examining the typical context of use of the key term *embryo* in T1 and the equivalent in T2 *fetus* (later reformulated as *placenta*). In the Popular Science account (T1) *embryo* is embedded in three types of cohesive chains which emphasise:

26   1. the status of the *embryo* as an Agent or (potentially conscious) Participant (*the embryo produces a hormone, [if the embryo were] given a choice, in the embryo's interests, the embryo...can be entirely oblivious of this conflict, this is just what the embryo does*);

27   2. the relationship between the *embryo* and the *mother* as co-dependent Participants (*brains of the mother and embryo, nutrition she provides her embryo, the mother bearing it [=the embryo], the genes of her embryo...*);

28   3. the *embryo* as an (Affected) Participant (*the embryo... being stillborn, its own survival, threat to the embryo's survival*).

29   These patterns contrast with the Research Article abstract (T2), in which three cohesive chains are formed around the terms *fetus* and (later in the text) *placenta*:

30   1. the *fetus* (or *placenta*) as Agent (*the attempt by the fetus... to increase its supply of nutrients, the fetus gains direct access to, the placenta is able to release hormones, Placental hormones... manipulate maternal physiology for fetal benefit.*);

31   2. the relationship between the *fetus* and *mother* as conflictual Participants (*conflict can be said to exist between maternal and fetal genes, fetal actions are opposed by maternal countermeasures, This (fetal) action ... is countered by increased maternal production of insulin, the mother is unable to ...mount an adequate response to fetal manipulation, poorly nourished fetus*);

32   3. the *fetus* as Classifier or Circumstance, a location for the activities of cells and genes (*genes expressed in fetuses, fetal genes will be selected, fetally derived cells, a similar conflict exists within fetal cells*).

33   The overall effect in Dan Dennet's text (T1) is to emphasise the *embryo*'s viewpoint or predicament and to underline that there is an equilibrium between two opposing but complementary Participants (encapsulated by metaphors such as *tug-of-war*). This kind of 'human story' is perhaps to be expected in a popular science book. In contrast, and in keeping with the conventions of science writing, the research article abstract (T2) de-humanises *the fetus* by embedding it in complex noun groups, or by transferring Agency to other Participants (such as *cells*, *genes*, the *placenta*). The text is not without drama however: the author David Haig consistently underlines the conflictual nature of the relationship between *mother* and *fetus*, and couches this in surprisingly warlike terms (*conflict, countermeasures, escalation, invasion, resistance*).

5. The lexicogrammar of grammatical signs

34   In the previous sections, I have shown that it is possible to analyse various aspects of the sample texts T1 and T2 in terms of phraseological units, lexicalised phrases, collocational pairs and cohesive chains. However, as can be seen in the above analyses, all of these phenomena can be discussed in terms of a more general unit of analysis, namely the 'lexicogrammatical pattern'. LG patterns have been studied before in the SFL and applied linguistics literature, but as Hunston and Francis (2000) point out, the

starting point has usually been that of the lexical item. In this section (following Gledhill 1995, 1997, 2000a and 2000b), I set out an alternative method, which involves the identification of LG patterns on the basis of grammatical items. More recently, the term 'small word' has been used to analyse grammatical items in LG patterns (Groom 2005). However I find it preferable to refer to the 'grammatical sign', a term which includes not only grammatical ('small') items, but also lexicalised function words (such as complex prepositions *because of, by dint of, in so far as*), grammatical morphemes (including inflections such as the plural, *-ing, -ly* and more abstract forms, such as tense) and grammatical categories (such as the active sequence N+V in *mount a response* or the passive N+V in *gene expressed,* etc.).

35   Although there exists a growing body of research on collocational frameworks and other discontinuous patterns involving grammatical items (Renouf & Sinclair 1991, Luzon Marco 2000), the idea that grammatical signs enter into collocational relations is still not generally accepted, as can be seen in the following definition:

> collocation, n. A term used in lexicology by some (especially Firthian linguists) to refer to the habitual co-occurrence of individual lexical items [...]. Some words have no specific collocational restrictions – grammatical words such as *the, of, after, in* [...]. (Crystal 2008: 86-87)

36   Yet there is ample evidence that grammatical signs are involved in collocational patterns. For example in Gledhill (2000a and 2000b), I explored the hypothesis that every text-type, and in particular every sub-section of a research article, has a particular configuration of grammatical items. The first step in this analysis is to establish the statistical distribution of grammatical words (not including morphemes and other grammatical signs) in the sub-sections of 500 research articles (the *Pharmaceutical Sciences Corpus,* PSC). For example, the first five statistically most 'salient' items in each research article sub-section are set out below in Table 1.

Table 1. The distribution of verb forms in the Pharmaceutical Sciences Articles (Gledhill 2000b: 112)

Relative Rank	Titles	Abstracts	Introductions	Methods	Results	Discussions
1	of	but	been	were	no	that
2	for	these	has	was	in	be
3	on	of	have	at	did	may
4	and	there	is	then	not	is
5	in	in	such	for	had	our

37   To some extent this kind of analysis simply confirms the findings of previous research on the research article genre (Swales 1990); for example the prevalence of *have been* in Introductions signals the perfect, the prevalence of *was, were* in the Methods signals the use of past passive forms, and so on. However, the important point is that these items are not used in isolation, but co-occur with others to form longer lexicogrammatical patterns. The following examples give some idea of how the 'salient' grammatical items in Abstracts (lines 11-15) and Discussions (16-20) co-occur in sequences which

ultimately represent some of the most typical LG patterns for each of these sub-sections of the research article (here each statistically salient item is indicated in bold):

Abstracts

> (11) the mechanism **of** action **of** {compound Y} **was** shown **to** {empirical process} (nominal expression of findings)
>
> (12) **there was** a significant increase **in** toxicity (quantitative report)
>
> (13) It is concluded **that** propagation **did** not increase (impersonal expression of quantitative report)
>
> (14) **but** subjects **who** receive active management (contrastive results expressed in embedded clause)
>
> (15) **both** normal **and** tumor cells (contrastive framework)

Discussions

> (16) data suggests **that** reactive oxygen would **be** important (projected report of biochemical process)
>
> (17) It **is** interesting to note **that** (evaluation of research process)
>
> (18) increasing data does **not** result **in** any further enhancement (metaphorical empirical report)
>
> (19) **This** evidence suggests **that** (reformulation of previous data and projection of research process)
>
> (20) **we** have found **that** (projected report of research process)

38   It is important to point out that examples such as these represent prototypical but also productive sequences. This can be shown by using a concordancer (Anthony 2007) to search for discontinuous sequences, as in * *of* * *was* *(-ed) to* (where * represents a 'wild-card', either a whole word or part of a word). This pattern, based on example (11) above, is often found in Abstracts in the PSC, usually in phrases which summarise experimental data. A search of the PSC reveals that two reporting verbs are typically involved in this pattern (*find* and *show*), and the subjects of these verbs typically have the structure: *Empirical process of Biochemical entity X*:

> (21) Another neuroprotective **activity of** cannabimimetics **was shown to** be associated with the CB1-mediated inhibition of nitric oxide release from rat microglia cells.
>
> (22) In our case the optimum **content of** acetonitrile **was found to** vary between 25 and 30%, depending on the column efficiency.
>
> (23) The **efficacy of** zidovudine **was shown to** reduce risk of transmission by 66% in the treated group.
>
> (24) The **prevalence of** restraint **was found to** be 68% (n=69).
>
> (25) As compared with the non-pregnant women, the **sensitivity** to the glucose-lowering effect **of** insulin **was found to** be reduced 45Â¯70% in the 3rd trimester

39   I have so far examined patterns on the basis of large-scale corpus analysis. But it is also worthwhile examining the specific patterns that emerge in individual texts, and asking to what extent they are related to neighbouring segments of the same text, or to examples of the same register or the general language as a whole. For example, in sample text T2 the first few lines of the text contain a number of LG patterns. In the first sentence *Pregnancy has commonly been viewed as a cooperative interaction*, the sequence * *has* * *ly been* * *(-ed, -en) as* * occurs 16 times in the BNC and 56 times in the PSC. This clearly suggests that this is a significant LG pattern in science writing. Even in the BNC, the kinds of (cognitive, communicative, reporting) verbs used in this pattern

correspond closely to the wording adopted in T2, and their context of use is typically that of 'academic exposition':

> (26) May Sinclair **has frequently been described as** shy and scholarly.
> (27) Providing support **has previously been identified as** a key aspect of the district nurses role in palliative care
> (28) This expansion of the role of the state **has variously been interpreted as** a functional response to old age incapacity,
> (29) Indeed, iconoclasm **has frequently been portrayed as** little more than mindless vandalism perpetrated by
> (30) Coeliac disease **has traditionally been regarded as** a disorder of childhood and early adult life.

40   A similar pattern emerges in the second sentence of text T2, *The effects of natural selection on genes expressed in fetuses.* The framework * *the * s of * on * s* is highly productive in the BNC and the PSC, and in both cases the noun *effects* emerges as a pivotal lexical item (20 out of 74 sequences in the BNC; in the PSC all 25 examples). However, whereas text T2 presents *effects* as a Theme/Subject, in both the BNC and PSC, the noun *effects* is usually embedded in a complex noun group as a qualifier, or as a complement of a Research-oriented process (*concentrate on, examine, investigate, questions about, work on,* here marked in italics):

> (31) Thus the scarce *research work* **on the effects of participation on** effectiveness is further limited by its inability clearly to
> (32) so too, on the other hand, do *questions about* **the effects of adrenalin on** sitters in an examination room, or family genetics on the
> (33) Based on the expectation that cellular functions would be adversely affected by such increased steroid levels, most research has *concentrated on* **the effects of glucocorticoids on** lens metabolism and ion levels.
> (34) Therefore, we undertook a series of studies that *examined* **the effects of cannabinoids on** noxious stimulation-evoked activity in nociceptive spinal and thalamic neurons
> (35) Our study is novel with respect to *investigate* **the effects of erythromycin on** LPS-induced preterm labor model in rats

41   A final, rather curious, example can be found in the third sentence in T2 *In this sense, a genetic conflict can be said to exist.* This is in fact similar to the pattern that we have already seen in research article abstracts (examples 11 and 21-25). The difference here is that the phrase used in T2 involves a modal verb rather than the past tense (*can be *(ed) to **) and the clause is introduced by a textual marker *In this sense.* From a functional point of view, the phrase *In this sense* signals that a previous discourse referent is to be reformulated by an explicit evaluation (above, marked in italics), a similar discourse function observed in the PSC examples (11, 21-25). Interestingly, if we look for this framework in the BNC, a very similar pattern emerges:

> (36) **In this sense,** Keynes's General Theory **may well be regarded as** *self-defeating* in terms of its impact on political economy.
> (37) **In this sense,** the definition of standards and routines **can be seen as** a d*efensive process*: the housewife is defending herself against the allegation that she does nothing at all.
> (38) **In this sense,** the placement in industry **will not be viewed as** *an end in itself* but as an essential ingredient in the process of change...

(39) [...] secondary-school pupils in China have only a limited opportunity to go into higher education and **in this sense,** those who do make it **can be viewed as** *a privileged elite.*
(40) **In this sense,** this strategy **may be viewed as** *a tactic* employed in the air quality management strategy to meet air quality standards.

42   The examples (36-40) all involve the same phrase *In this sense* as in T2 as well the same verb (or a with a similar meaning: *see, regard, view*). The main difference with T2 however is that the verb introduces a Complement (after *as*). Generally speaking, therefore, it can be seen here that the wording used in text T2 does not quite correspond to the 'prevailing phraseology', that is to say the typical LG patterns to be found either in the BNC or the PSC. Rather, in text T2 the author appears to have created a hybrid construction, which exploits in the first instance a pattern from the BNC (albeit in an academic register), and then reverts to a pattern that is more generally found in the PSC. This example shows that what looks like a novel creation may in fact involve the seamless joining of two regular patterns from the general discourse of academic and/or scientific writing.

6. Conclusion

43   In this paper I have argued that the lexicogrammatical pattern should be seen as a fundamental unit of analysis in text analysis, and is perhaps a more useful unit of analysis for the purposes of ESP than such notions as 'phraseological unit', 'collocational pair' and the like. I have based my observations here on Halliday's theory of 'lexicogrammar', which in this paper I conceive as a system of choices for the creation of meanings, with each choice corresponding to a cascade of lexical and grammatical features associated with a particular register or discourse function. By 'cascade' I am highlighting the fact that any choice of expression inevitably leads to a further set of choices and associated expressions, with the result that stretches of speech appear to be at the same time pre-constructed and coherent, but also highly varied and productive.

44   If the notion of lexicogrammar is such a useful concept, what are we to make of phraseology and collocation? In fact, these terms present different perspectives on the same object of enquiry. I have argued in this paper that, from an SFL perspective, it would be useful to view phraseology in terms of the diachronic process of lexicalisation. This view, as mentioned above, has the advantage of avoiding any artificial distinction between 'idiomatic' and 'normal' forms of expression. In addition, if a distinction has to be made between idiomatic expressions and other types of phrase, it is perhaps better to conceive of this in terms of rhetorical effect, a perspective that I have discussed elsewhere (Gledhill 2008). Finally, the SFL perspective on collocation is that it is a semantic concept, which refers to the dependent relationship between a sign and its habitual context of use. Although many linguists prefer to analyse collocations on the basis of large-scale corpus analysis, I have argued here (and elsewhere, Gledhill 2009) that it would also be useful, from an ESP perspective at least, to look at how collocation operates in individual texts. In this case, it is worth seeing collocation, as originally proposed by Halliday and Hasan (1976), as a form of textual cohesion.

BIBLIOGRAPHY

Anthony, Laurence. 2007. *AntConc 3.2.1w* (Windows Concordancer), Faculty of Science and Engineering, Waseda University, Japan, 10 March 2007.

Benson, Morton, Evelyn Benson & Robert Ilson. 1986. *The Lexicographic Description of English.* Amsterdam/London: John Benjamins.

Biber, Douglas, Susan Conrad, & Viviana Cortes. 2004. "If you look at...: Lexical bundles in university teaching and textbooks". *Applied Linguistics* 25/3, 371-405.

Brinton, Laurel & Elizabeth Traugott. 2005. *Lexicalization and Language Change.* Cambridge: Cambridge University Press.

Cavalla Cristelle. 2008. "Les collocations dans les écrits universitaires : un lexique spécifique pour les apprenants étrangers". In Bertrand, Olivier & Isabelle Schäffner (eds.), *Le français de spécialité.* Paris: Éditions de l'École polytechnique, 91-103.

Crystal, David. 2008 [1991]. *A Dictionary of Linguistics and Phonetics* [6th edn.] London: Blackwell.

Evert, Stefan. 2004. "The statistics of word co-occurrences: Word-pairs and collocations". Doctoral thesis, University of Stuttgart.

Fernando, Chiltra. 1996. *Idioms and Idiomaticity.* Oxford: Oxford University Press.

Fiedler, Sabine. 2007. *English Phraseology. A Coursebook.* Tübingen: Gunter Narr Verlag.

Firth, John Rupert. 1957. *Papers in Linguistics 1934-1951.* Oxford: Oxford University Press.

Firth, John Rupert. 1968. "A synopsis of linguistic theory, 1930-1955". In Palmer, F. R. (ed.). *Selected papers of J. R. Firth 1952-1959.* London: Longman*, 168-205.

Frath, Pierre & Christopher Gledhill. 2005. "Qu'est-ce qu'une unité phraséologique ?". *Cahiers de l'Institut de linguistique de Louvain* 31/2-4, 11-25.

Fries, Peter, Michael Cummings, David Lockwood & William Spruiell (eds.). 2002. *Relations and Functions within and around Language.* London: Continuum.

Gläser, Rosmary. 1988. *Phraseologie der englische Sprache.* Tübingen: Max Niemeyer.

Gledhill, Christopher. 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 Christopher. 1997. "Les collocations et la construction du savoir scientifique". *ASp* 15-18, 85-104.

Gledhill, Christopher. 2000a. "The discourse function of collocation in research article introductions". *English for Specific Purposes* 19/2, 115-135.

Gledhill, Christopher. 2000b. *Collocations in Science Writing.* Tübingen: Narr.

Gledhill, Christopher. 2008. "Le signe et le syntagme : entre phraséologie et lexico-grammaire. Une synthèse du modèle systémique fonctionnel de Michael Halliday et de la théorie sémiotique de Charles S. Peirce". Thèse d'habilitation. Université de Bretagne occidentale.

Gledhill, Christopher. 2009. "Colligation and the cohesive function of present and past tense in the scientific research article". In Banks, David (ed.), *Les Temps et les textes de spécialité.* Paris: L'Harmattan, 65-84.

Gledhill, Christopher & Pierre Frath. 2007. "Collocation, phrasème, dénomination : vers une théorie de la créativité phraséologique". *La Linguistique* 43/1, 65-90.

Gonzalez-Rey, Isabel. 2002. *La Phraséologie du français.* Toulouse: Presses Universitaires du Mirail.

Granger, Sylviane & François Meunier (eds.). 2008. *Phraseology in Foreign Language Learning and Teaching.* Amsterdam/Philadelphia: John Benjamins.

Gréciano, Gertrud. 1997. "La phraséogénèse du discours". In Martins-Baltar, M. (ed.), *La Locution, entre langue et usages.* Fontenay–Saint-Cloud: ENS Éditions, 179-200.

Groom, Nicholas. 2005. "Pattern and meaning across genres and disciplines: an exploratory study". *Journal of English for Academic Purposes* 4/3, 257-277.

Grossmann, Francis & Agnès Tutin. 2002. "Collocations régulières et irrégulières". *Revue française de linguistique appliquée* 7/1, 7-25.

Halliday, Michael. 1961. "Categories of the theory of grammar". *Word.* 17/3, 241-292.

Halliday, Michael & Hasan, R. 1976. *Cohesion in English.* London: Longman.

Halliday, Michael & Christian Matthiessen. 2004. *An Introduction to Functional Grammar* [3rd edn]. London: Arnold.

Hausmann, Franz J. 1985. "Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels". In Bergenholtz Heinrich & Joachim Mugdan (eds.), 1958. *Akten des Essener Kolloquiums zur Grammatik im Wörterbuch.* Tübingen: Max Niemeyer Verlag, 118-129.

Hoey, Michael. 2005. *Lexical Priming: A new theory of words and language.* London: Routledge.

Howarth, Peter. 1996. *Phraseology in English Academic Writing.* Tübingen: Max Niemeyer.

Hunston, Susan & Gill Francis. 1998. "Verbs observed: A corpus-driven pedagogic grammar". *Applied Linguistics* 19/1, 45-72.

Hunston, Susan & Gill Francis. 2000. *Pattern Grammar.* Amsterdam: John Benjamins.

Jones, Martha & Sandra Haywood. 2004. "Facilitating the acquisition of formulaic sequences: an exploratory study in an EAP context". In Schmitt, Norbert (ed.), *Formulaic Sequences.* Amsterdam/ Philadelphia: John Benjamins, 269-300.

Laso Martiin & Natalia Judith. 2008. "A corpus-based study of the phraseological behaviour of abstract nouns in medical English. A needs analysis of a Spanish medical community." Doctoral thesis. Universitat de Barcelona.

Legallois, Dominique & Jacques François (eds). 2006. *Autour des grammaires de constructions et de patterns. Cahier du CRISCO* 21. Caen: Presses de l'Université de Caen.

Luzon Marco, Maria José. 2000. "Collocational frameworks in medical research papers: A genre-based study". *English for Specific Purposes* 19, 63-86.

Martin, Jim. 2001. "Language, register and genre". In Burns, Anne & Caroline Coffin (eds.), *Analysing English in a Global Context.* London: Routledge, 149-166.

Mel'čuk, Igor, André Clas & Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire.* Louvain: Duculot.

Moon, Rosamund. 1994. "The analysis of fixed expressions in text". In Coulthard, Malcolm (ed.) 1994. *Advances in Written Text Analysis.* London: Routledge, 117-135.

Muller, Charles. 1968. *Initiation à la linguistique statistique.* Paris: Larousse.

Nesselhauf, Nadja. 2003. "The use of collocations by advanced learners of English and some implications for teaching". *Applied Linguistics* 42/2, 223-242.

Pavel, Sylvie. 1993. *La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques.* Montréal: Secrétariat d'État du Canada, Direction de la terminologie et des services linguistiques.

Pearson, Jennifer. 1998. *Terms in Context.* Amsterdam: John Benjamins.

Pecman, Mojca. 2005. "De la phraséologie à la traductologie proactive : essai de synthèse des fondements théoriques sous-tendant la recherche en phraséologie". *Meta : journal des traducteurs/ Meta: Translators' Journal* 50/4 <http://www.erudit.org/revue/meta/2005/v50 /n4/index.html>.

Renouf, Antoinette & John Sinclair. 1991. "Collocational frameworks in English". In Aijmer, Karin & Bengt Altenberg (eds.), *English Corpus Linguistics.* London: Longman, 128-143.

Sager, Juan, David Dungworth & Peter McDonald. 1980. *English Special Languages: Principles and Practice in Science and Technology.* Wiesbaden: Oscar Nadstetter Verlag.

Siepman, Dirk. 2005. "Collocation, colligation and encoding dictionaries, Part I: Lexicological aspects". *International Journal of Lexicography* 18, 409-443.

Smadja, Frank. 1993. "Retrieving collocations from text: Xtract". *Computational Linguistics* 19/1, 143-177.

Stefanowitsch, Anatol & Stefan Gries. 2003. "Collostructions: Investigating the interaction between words and constructions". *International Journal of Corpus Linguistics* 8/2, 209-243.

Stubbs, Michael. 1995. "Collocations and semantic profiles: On the cause of the trouble with quantitative studies". *Functions of language* 2/1: 23-55.

Swales, John. 1990. *Genre Analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Thomas, Patricia. 1993. "Choosing headwords from language-for-special-purposes (LSP) collocations for entry into a terminology data bank (term bank)". In Sonneveld, Helmi & Kurt Loenig (ed.) *Terminology. Applications in Interdisciplinary Communication.* Amsterdam/Philadelphia: John Benjamins, 43-68.

Tollis, Francis. 2001. "Bilan orienté". In Tollis, Francis (ed.), *La Locution et la périphrase : du lexique à la grammaire.* Paris: L'Harmattan, 211-268.

Tucker, Gordon. 1998. *The Lexicogrammar of Adjectives: A Systemic functional approach to lexis.* London: Cassell Academic.

Tutin, Agnès. 2007. "Autour du lexique et de la phraséologie des écrits scientifiques". *Revue française de linguistique appliquée* 12/2, 5-14.

Williams, Geoffrey. 1998. "Collocational networks: interlocking patterns of lexis in a corpus of plant biology research articles". *International Journal of Corpus Linguistics* 3/1, 151-171.

Williams, Geoffrey. 2003. "Les collocations et l'école contextualiste britannique". in Tutin, Agnès & Francis Grossman (eds.), *Les collocations : analyse et traitement.* Amsterdam: de Werelt, 33-44.

Wray, Alison. 2002. *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

APPENDIXES

In the following sample texts, the phraseological units, collocations and other lexicogrammatical patterns mentioned in this article have been emphasised in bold.

Text T1 Popular Science Extract

Dennett, Dan. 1996. *Darwin's Dangerous Idea.* London: Penguin, 255.

Homo sapiens is not exempt from the sort of **genetic conflict** David Haig postulates to explain genomic imprinting; in an important new article (1993) he analyzes a variety of **conflicts** that exist between the **genes of a pregnant woman** and the **genes of her embryo. It is in the embryo's interests, of course, that** the **mother bearing it** stay strong and healthy, for **its own survival** depends on her not only **completing her term of pregnancy** but tending for her newborn. However, if the mother, in her attempt to stay healthy under **trying circumstances** — famine, for instance, which must have been a common circumstance in most generations of human existence — should cut down on **the nutrition she provides her embryo**, at some point this becomes more of **a threat to the embryo's survival** than the alternative, a weakened mother.

If the **embryo were "given a choice"** between being spontaneously aborted early in the pregnancy or being stillborn or of low birth weight **on the one hand**, versus being born at normal weight of a weak or even dying mother on the other, what would (selfish) reason dictate? It would dictate **taking whatever steps are available** to try to ensure that the mother does not **cut her losses** (she can always try to have another child later, when the famine is over), and **this is just what the embryo does**. Both **embryo and mother can be entirely oblivious of this conflict** — as oblivious as the trees rising competitively in the forest. **The conflict plays out** in the genes and their control of hormones, not in the **brains of mother and embryo**; it is the same sort of **conflict we saw between maternal and paternal genes** in the mouse. There is a flood of hormones; **the embryo produces a hormone** that will enhance its own growth at the expense of the mother's nutritional needs, her body responds with an antagonist hormone that attempts to undo the effect of the first; **and so on**, in an escalation that can produce hormone levels many times higher than normal. This **tug-of-war** usually ends in a mutually semi-satisfactory standoff, but it produces a host of **by-products** that would be utterly baffling and senseless were they not the predictable effects of such conflict. Haig concludes with an application of the fundamental game-theoretic insight: "Maternal and fetal **genes** would both **benefit** if a given transfer of resources was achieved with a lesser production of… hormones and less maternal resistance, but such an agreement is evolutionarily unenforceable" (Haig 1993: 518).

Text T2 Research Paper Abstract

Haig, David. 1993. "Genetic conflicts in human pregnancy". *Quarterly Review of Biology* 68, 495-532.

Abstract

Pregnancy **has commonly been viewed as** a cooperative **interaction between a mother and her fetus. The effects of natural selection on genes expressed in fetuses**, however, may be opposed by **the effects of natural selection on genes expressed in mothers. In this sense, a genetic conflict can be said to exist** between **maternal and fetal genes. Fetal genes** will be **selected** to increase the **transfer of nutrients to their fetus**, and maternal **genes** will be **selected** to limit transfers in

excess of some maternal optimum. Thus a process of evolutionary escalation is predicted in which **fetal actions are opposed by maternal countermeasures**. The phenomenon of **genomic imprinting** means that a similar conflict exists within **fetal cells** between **genes** that are **expressed** when maternally derived, and **genes** that are **expressed** when paternally derived. During implantation, **fetally derived cells** (trophoblast) **invade** the maternal endometrium and remodel the endometrial spiral arteries into low-resistance vessels that are unable to constrict. This **invasion** has three consequences. First, **the fetus gains direct access** to its mother's arterial blood. Therefore, **a mother cannot reduce the nutrient content of blood** reaching the placenta without reducing the nutrient supply to her own tissues. Second, the volume of blood reaching the placenta becomes largely independent of control by the local maternal vasculature. Third, the placenta is able to **release hormones** and other substances directly into the maternal circulation. Placental hormones, including human chorionic gonadotropin (hCG) and human placental lactogen (hPL), are predicted to manipulate maternal physiology for **fetal benefit**. For example, hPL is proposed to act on maternal prolactin receptors to increase maternal resistance to insulin. If unopposed, the effect of hPL would be to maintain higher blood glucose levels for longer periods after meals. This action, however, is countered by increased **maternal production of insulin**. Gestational diabetes develops if the **mother is unable to mount an adequate response to fetal manipulation**. Similarly, **fetal genes** are predicted to **enhance** the flow of maternal blood through the placenta by increasing **maternal blood pressure**. Preeclampsia can be interpreted as an attempt by a **poorly nourished fetus** to **increase its supply of nutrients** by **increasing the resistance of its mother's peripheral circulation**.

NOTES

1. This article is based on a paper read at the GERAS workshop (*Anglais de spécialité)* at the SAES Congress, Lille 21-23 May 2010.
2. Here I follow the usual SFL practice of using capitalised initials for semantic roles and grammatical functions.

ABSTRACTS

Phraseology and collocation are central phenomena in language. However, the usual way in which these notions are understood is often highly restricted (phraseology is often associated with 'idiomatic expressions', collocation is often seen as 'a fixed sequence'). In this paper I recast these notions in terms of Halliday's notion of lexicogrammar. My particular aim is to show how collocation and phraseology, from a lexicogrammar perspective, are relevant to English for Specific Purposes (ESP). I also set out here a method of textual analysis which exploits the phraseological behaviour of grammatical signs. Far from being 'collocation free', grammatical items can be shown to be stable elements in relatively predictable but also productive cascades of

expression. I would argue that the identification of such extended lexicogrammatical patterns should be a key feature in the systematic analysis of ESP texts.

La phraséologie et la collocation sont des phénomènes linguistiques fondamentaux. Cependant, les linguistes ont souvent une conception assez limitée de ces termes (la phraséologie est associée aux « expressions idiomatiques », les collocations considérées comme des « séquences figées »). Dans cet article, je refonde ces notions du point de vue de la « lexicogrammaire » de Michael Halliday. Mon objectif est de démontrer comment ces notions, dans l'optique de la théorie systémique fonctionnelle, seraient utiles pour l'anglais de spécialité (ASP). Je propose également une méthode d'analyse textuelle qui s'appuie sur le rôle phraséologique des signes grammaticaux. Loin d'être « sans collocations », il est en effet possible de démontrer que les mots grammaticaux sont des éléments stables dans des « cascades » d'expressions relativement prévisibles mais aussi productives. L'identification de schémas lexicogrammaticaux de ce type devrait constituer une étape de base dans l'analyse systématique des textes d'ASP.

INDEX

AUTHOR

CHRISTOPHER GLEDHILL

Christopher Gledhill is *professeur des universités* in English linguistics at Lille 3 University, where he teaches linguistics, grammar, phraseology and specialised translation. He was *maître de conférences* in English at Strasbourg University and Lecturer in French at Aston and St. Andrews universities. His research interests include corpus linguistics, the linguistic analysis of artificial and invented languages, discourse analysis and phraseology. christopher.gledhill@univ-lille3.fr