How trainee translators and their teachers deal with phraseological units in the ARTES database

Mojca Pecman, Christopher Gledhill

How trainee translators and their teachers deal with phraseological units in the ARTES database

Mojca Pecman[1] and Christopher Gledhill[2]
[1]Maître de conférences, [2]Professor, [1,2]CLILLAC-ARP EA 3967 Paris Diderot University

Abstract: This paper discusses the notion of 'generic collocation' as a translational unit of relevance for achieving textual coherence, and explores how such units are analysed and processed in the ARTES database by Masters students in Specialised Translation. Generic collocations (GCs) are semi pre-constructed sequences of words, which are used widely in texts belonging to the same register or genre, and which extend beyond the narrow confines of one same domain – e.g. *in this paper/report/study we conclude/show/suggest that X, Failure to (comply/follow these instructions,* etc*.) may result in (damage/malfunction injury,* etc). GCs constitute one of the most codified 'signatures' of technical and specialised discourse, a fact that should be of interest to specialised translators. The ARTES project (*Aide à la Redaction de TExtes Scientifiques*) involves a database in which students collect and record GCs for future use by translators and writers of LSP texts. We thus explore the architecture of ARTES and its role as a template for teaching GCs. We also discuss the problems encountered by students of translation, who often have difficulty in understanding the concept of GCs and applying it to the analysis of their own data.

Key words: generic collocations, phraseological units, terminological and phraseological database, ARTES, scientific discourse, translation units

1. Introduction

This paper focuses on 'generic collocations' (GCs) as translational units and text-unifying linguistic items, and on the way teachers and students deal with GCs in the ARTES database. The ARTES database (*Aide à la Redaction de TExtes Scientifiques* / Dictionary-assisted writing tool for scientific communication)[i] is a multilingual multidomain language resource targeting various LSP users, namely students, translators and domain experts, who are all involved in specialised communication (Pecman & Kübler 2011). ARTES offers a comprehensive approach to lexical resources: terminological, phraseological, domain-specific, domain-free, semasiological and onomasiological (Pecman 2004, 2007, Kübler & Pecman 2012). Designed in 2010 by the group of researchers from Paris Diderot University, the ARTES DB is an outcome of several previous projects aiming at the creation of resources for LSPs and specialised translation (e.g. Kübler 2004, Pecman 2004, Mestivier Volanschi 2008, etc.). Apart from ARTES, there are few large-scale projects targeting phraseological phenomena such as generic collocations (GCs), with notable exceptions being *ScienText* (Tutin & Grossmann 2013) and the *Louvain English for Academic Purposes Dictionary* (LEAD) (Granger & Paquot 2015) projects.

In order to explain the way we approach GCs in the context of the ARTES project, we first focus on refining the notion of GC in relation to recent advances in the research on this specific linguistic phenomenon. We then examine how such units are identified and analysed by trainee translators in Masters studies on language industries and specialised translation,

Industrie de la Langue et Traduction Spécialisée (ILTS)[ii], at Paris Diderot University. Thus, the aim of our paper is twofold: 1) to argue generally that generic collocations are relevant units for achieving register conformity and coherence in specialised translation, as well as an important resource for terminological databases, and 2) more specifically, to improve the ability of advanced students of specialised translation to identify and analyse GCs in specialised corpora. In order to achieve these aims, we attempt to answer the following research questions: What are 'generic collocations' and what do we know about them?, How do students understand generic collocations?, and How do students analyse and categorise generic collocations?

2. What are 'generic collocations' and what do we know about them?

2.1. Past and current trends in studies on generic collocations

The term 'generic collocation' (GC) was first used by Gledhill (2000a/b), in the context of a study of collocation in scientific research articles:

> "it must be the case that examples of collocational regularity across widely different research specialisms (and across a broad range of periodicals) represent a form of coherent scientific style. The term I propose for these expressions is generic collocation." (Gledhill 2000: 238).

This admittedly very informal concept has since been explored, in various forms, by a number of linguists (Coxhead 2000, 2002; Coxhead & Hirsh 2007, Gledhill 2000a/b, Pecman 2007, 2008, 2012; Simpson-Vlach & Ellis 2010, Tutin 2007a/b). We would claim that these studies constitute a distinct area within the general field of phraseology. In these studies, the notion of GC is not addressed uniformly. Some analysts have focused on academic formulae (Simpson-Vlach & Ellis 2010), others on transdiciplinary lexicons (Tutin 2007a/b) or on specifically scientific generic collocations (Pecman 2007, 2008, 2012), and some more exclusively on the epistemological role of GCs in academic and scientific discourse (Teufel 1998, Groom 2009). All these studies have contributed to enhancing our knowledge about GCs; however putting this knowledge into practice in translation or language classes and in building dictionaries remains a challenging objective which requires further research.

Only a small number of studies have looked at how GCs are used in language classes on academic writing (Tran *et al.* 2016) and in translation classes (Gledhill & Kübler 2015). At the same time, while 'specific collocations' are increasingly being recorded in term bases (for instance, this is the case in TermiumPlus[iii] and *Grand Dictionnaire Terminologique*[iv]), generic collocations – or items that resemble them – are less recognised. Current dictionaries and databases rarely take this resource into account: rare exceptions are found in the Louvain *English for Academic Purposes Dictionary* (LEAD) (Granger & Paquot 2015) and the ARTES database (Pecman & Kübler 2011). As mentioned above, other projects involving

generic collocations exist, such as *ScienText* (Tutin & Grossmann 2013) and *Lexicoscope*[v] (Kraif & Diwersy 2012). These projects focus on providing tools for retrieving the collocational profile of words in specific discourse types through the interrogation of corpora.

Recently, we have begun to examine the extent to which our own students are able to identify generic collocations in ARTES as a way of exploring the phraseology of the specialised and technical texts which they have to translate (Gledhill & Kübler 2015, 2016). This work has led us to refine our understanding not only of what constitutes a GC, but also to redefine the way we approach the important notion of the 'phraseological pattern'.

2.2. Definition of generic collocations

There are now many terms in the research literature in phraseology and collocation studies for the key concept of a recurring sequence of words: *bundle, cluster, chunk, collocational framework, formula, multiword unit, n-gram, routine*, etc. Each of these terms defines a similar yet specific type of word association. In our view, the term 'generic collocation' is still quite useful in comparison with these terms, because it combines information about form ('collocation' as a co-selected sequence of commonly co-occurring words) and function ('generic' refers to a general rhetorical function, as opposed to any reference to a specific concept). Thus we can define **generic collocations** as: frequently co-selected sequences of words which are productive in form (allowing for some variable constituents to fit into the co-text) and predictable in function (expressing a rhetorical meaning which has evolved for a particular discourse purpose) across a wide number of domains. It is necessary to underline the final element of this definition: GCs are 'generic' in the sense that they are used across a wide number of texts belonging to the same register or 'genre', but also – and crucially – across different areas of expertise. It is also important to emphasise the fact that, unlike n-grams, formulae and other fixed phrases, the same GC can take different forms, and allow for variations (thus helping it to fit in better with the surrounding co-text). Here for example is a selection of introductory phrases in English and French:

> (1a) ***In this paper, we*** *survey the state of the art in the area of information extraction and automated analysis tools for in vivo and in vitro biomolecular imaging.*
> (2a) ***In this section we*** *provide a detailed overwiew of human rights, including: where they come from and how they can help you.*
> (3a) ***Dans cet article nous*** *allons essayer de les passer en revue, avec un accent particulier sur les fonds [...] immobiliers cotés en bourse.*
> (4a) ***Dans la première partie, nous*** *passerons en revue l'état de l'art actuel en animation d'objets déformables au chapitre 2, ce qui nous amènera à considérer aussi les travaux en modélisation par surfaces implicites au chapitre 3.*

From these sequences, we can select the following segments as potential generic collocations, that is to say units which have a generic value and belong to the general scientific discourse of English and French:

(1b) ***In this paper, we*** *survey the state of the art in the area of [domain name]*
(2b) ***In this section we*** *provide a detailed overwiew of [domain name]*
(3b) ***Dans cet article nous*** *allons essayer de les passer en revue [concept]*
(4b) ***Dans la première partie, nous*** *passerons en revue l'état de l'art actuel en [domaine], ce qui nous amènera à considérer aussi les travaux en [domaine].*

As can be seen in these examples, we would claim that GCs are not just units on the basis of syntax (i.e. whether they correspond to main clauses, verb phrases, etc.), but also because they correspond to units at the level of semantics: thus, a key component of a GC is its 'discourse function' (Gledhill 2000a/b), that is to say the specific move (cf. Biber *et al.* 2007) or the specific ideational, textual or interpersonal function which is fulfilled by a particular pattern, such as: 'stating the scope of a paper or section' (examples 1a/b and 3a/b), or 'justifying a methodological approach, establishing a transition in the argument' (examples 2a/b and 4a/b), etc. These sequences are highly relevant for achieving register unity in a source text and an important resource for compilation in terminological databases.

3. How do students understand generic collocations?

In Gledhill & Kübler (2015), we examined the problems that students encounter when they analyse generic collocations in the ARTES database. The aim of this study was to see how students conceptualise GCs, and on the basis of these observations to make some broad proposals for improving the ways in which GCs are taught and handled in the database. A further aim was to examine how we as linguists ourselves conceive of GCs. Since we believe that phraseology (in the form of GCs) is *the* key defining linguistic feature that underlies both the conformity and quality of technical and scientific genres, in our view it is important for trainee translators to be able to 1) recognise that regularities of expression such as GCs exist in both English and French specialised discourse, 2) identify GCs correctly (assuming that if trainee translators cannot identify a given linguistic feature within the particular type of text, they have not really understood the phenomenon and cannot be fully 'fluent' in the specific discourse they have been asked to work with) and 3) find equivalent GCs between their different working languages, especially when they are operating in a language or a specialised discourse in which they are not necessarily experts. In the following two sub-sections, we set out some of the key findings of Gledhill & Kübler (2015), and then we set out some of their recommendations.

3.1. Difficulties encountered by students

In Gledhill & Kübler (2015), we looked at a sample of 50 generic collocations taken from 10 student projects (as mentioned above, students are asked to find 5 GCs for each translation project, and to report on them in their projects as well as to enter them on the ARTES DB). Of the GCs analysed, 11 were found to be complex clauses or longer, with a further 28 involving predicates; there is therefore a general tendency for students to look beyond simple groups or phrases, a practice which is encouraged in our classes on phraseology. There were however a number of problems, notably with the types of GCs analysed, and these tended to coincide with shorter sequences. Thus 11 of the 50 sampled GCs failed to correspond to our usual definition of 'generic' (including examples such as: *failure to, in parallel with, key insight, known to be, let us say, related work, safety precautions, serious games for, due to...*). This is not to say that all these analyses were incorrect: in some cases (notably *failure to, let us say, known to be*), the sequence does correspond to a longer unit of phraseology, it is just that this information was not encoded correctly in ARTES. For example, the sequence *Failure to…* is a standard introductory element in the 'warning message' that typically occurs in technical manuals:

> (5) <**failure to VV** (comply, follow these instructions, read the manual,) **may VV** (result in / lead to) **NN** (damage to the product, malfunction, personal injury, etc.)>.

Thus our first finding is that students are good at analysing fixed recurrent sequences (of the type which can be identified by corpus analysis tools as 'n-grams'), but are less comfortable with looking at longer stretches of co-text for more generic patterns of expression.

A further finding, discussed below, is that students have difficulty in identifying the overall meaning of these chunks of expression. This is not a surprising result, since many linguists (including grammarians and terminologists) are themselves unfamiliar with the concept of GCs, or other extended units of expression, and thus for many of our students, this is the first time they have been exposed to the notion. As mentioned below, ARTES provides students with 80 different functions which they can use to describe the general meaning (sometimes also called 'discourse function', or 'rhetorical function') of the sequence they have chosen. It would appear however that our students have difficulty in applying this analysis. In the sample of 50 GCs analysed in Gledhill & Kübler (2015), over 25 involved problems of analysis, either by failing to assign any discourse function, by confusing the domain in question and discourse function, or sometimes by attributing various different functions to the same GC. To give just one example of the latter, the sequence *<It is worth noting (that)>* had been assigned five functions: 'Highlighting a compatibility, correlation,

analogy', 'Expressing a notion of restriction or specification', 'Expressing an addition', 'Describing, interpreting and analysing data or observed phenomena' and 'Talking about characteristics, properties, specificities.' Clearly, different functions are valid in specific contexts, but there is no evidence that these particular functions represent a useful characterisation of the typical patterns of use for this phrase.

3.2. Recommendations for teaching generic collocations

Gledhill & Kübler (2015) made three recommendations on the basis of their survey. The first finding was that we need to re-examine how GCs are analysed and presented to our students. In the first instance, this involves rethinking the borders or limits of a pattern. For example, patterns such as *In this (Noun)* belong to much longer and much richer sequences of expression, as the following aggregate example shows:

> (6) <**In this NN** (NG type Document: *paper, report, study*), **we** (VG type Discursive verb: *conclude, show, suggest*) **that NN** (NG type Drug: *dextran-coated charcoal, ICI, TBCI*) **did/did not/may VB** (VG type Empirical verb: *attenuate, block, delay, decrease, prevent*) **NN** (NG type Gene: *BRCA1, IL2, M202*)-**VD** (*associated, induced, related*) **NN** (NG type Disease-related item: *breast cancer, morphological changes, oxidation pattern...*) >

Such examples show that collocational patterns almost always extend beyond the syntactic group into the clause and beyond. The above example displays 'pivotal' elements, and as many 'paradigms' (thus with many hundreds of permissible variations). It is not possible for our students to analyse everything, of course, and to a certain extent it may not be possible to account for the full scope of any really productive expression; nevertheless it seems fairly clear that our students should be encouraged to look for the longest sequences possible. There already exist corpus-based tools which enable us to find such sequences. For example, students can be asked to start their analysis by looking at 'n-grams' which are frequently recurring sequences of words in a given corpus, or at 'tag-grams', which are the corresponding sequences of parts-of-speech (as can be seen in the previous example). Although such a procedure can produce much noise, it is important for students to learn how to distinguish between potentially productive sequences of words and sequences which have only a local value.

Inevitably, at some point our students encounter the considerable problem of variability. This then leads us to our second recommendation: it is necessary to use a system of annotation to distinguish between the elements in a generic collocation that are **pivotal** (obligatory) and **paradigmatic** (structurally productive but semantically predictable): that is to say, even when a sequence displays considerable internal variation in terms of lexical items (= productive form), it is still possible to assign an overall meaning (= predictable function) for extended units such as, < *In this NN, we VV that*... > = 'reporting', or <*Failure to VV, ...*

may result in NN > = 'warning' (as proposed in Gledhill *et al.* 2017). Again, it is possible to conduct some simple tests on the basis of corpus analysis to demonstrate to students what is a pivotal item and what is not. Thus for example, if one strips a sequence of its lexical items, one is left with sequences of grammatical items, which when searched for in the appropriate corpus (although perhaps not in any corpus) may correspond to useful patterns of expression. Students can then also be asked to formulate hypotheses about the nature of such sequences. For example, which of the following two sequences (i) <*In this *, we* > or (ii) <*Failure to * may * in* > corresponds to a regular pattern of expression in (a) academic discourse, or (b) technical manuals?

The third recommendation made by Gledhill & Kübler (2015) attempted to address the many problems involved in analysing discourse functions. To a certain extent, this issue has more to do with the way ARTES presents these items to the students, and is thus addressed in the following section. There is however a more general point, which has to do with the extent to which it is possible to harmonize semantic analysis. In Gledhill & Kübler (2015), it was suggested that instead of analysing generic collocations in terms of 80 sometimes rather specific functions, students should be given a more limited initial choice, by using for example, the three-part system based on systemic functional grammar (SFG). Such a system broadly divides functions into three: *ideational* – relating to the representation of knowledge, *interpersonal* – relating to evaluation and authorial stance, and *textual* – relating to the identification of referents and in-text periodicity. Such a system is highly symmetrical and has been applied in many other contexts – for example, it is used by Tribble (2011) to categorise 'lexical bundles' in academic discourse. Unfortunately, the categories of SFG require many further sub-divisions, and it is by no means guaranteed that such a system would be easier to apply than the current menu of functions proposed in ARTES.

There exist however further problems, which were not raised by Gledhill & Kübler (2015), but which are relevant to teaching of generic collocations and their use in ARTES. One particularly thorny problem is the extent to which students (and other users of the DB) can find the appropriate equivalent patterns in the target language. To what extent, for example, is it true to say that <*In this paper, we*> is the equivalent translational unit to <*Dans cet article, nous*>? Superficially, these look the same. But when a corpus-based analysis is conducted, it is actually possible to spot certain differences of usage. As the following sample shows, English usage (examples 7-9) is not exactly matched by typical usage in French (10-12):

(7) ***In this study, we*** *investigate redox conditions in the atmosphere and in shallow-marine environments.*

(8) ***In this study, we*** *examined the sensitivities of the viscosity structure to the decay time of the Chandler wobble.*
(9) ***In this study, we*** *analysed the effect of land use change on hydrological model parameters by calibrating the model parameters of different time periods with different land use via a linearized calibration method.*
(10) ***Dans cette étude, nous nous*** *intéressons aux assemblages minéralogiques de ces roches naturelles car ils présentent un grand potentiel pour estimer les pressions et les températures d'équilibre.*
(11) ***Dans cette étude, nous nous*** *sommes focalisés uniquement sur le taux de CD4 au diagnostic.*
(12) ***Dans cette étude, nous nous*** *intéressons à la propagation de signaux compromettants de niveaux élevés mais dont la probabilité est faible.*

There are naturally similarities: in each case, the complement of the verb in the main clause refers to the topic of the research paper (and it is interesting to note that the main noun is a 'facette' noun, expressing a quantity or a quality such as *conditions, sensitivities, effect… assemblages, taux, propagation,* etc. rather than more specific discourse referents). However, there are also several differences: in the English examples, the verb introduced in the main clause is a research-oriented epistemological verb (a mental process involving observation, analysis, sifting of data, etc.), while in the French examples the (usually reflexive) verb expresses a more general mental process (involving cognitive focus). So while these patterns are undoubtedly 'equivalent', there are also important differences of perspective which our students may or may not be able to relate in the ARTES DB.

This point leads us to mention a further issue which we have not had space to discuss here, namely the problem of ergonomics: how can we categorise patterns usefully, so that they can be easily looked for or found by other DB users? It is to such issues that we now turn in the final section of this paper.

4. How do students analyse and categorise generic collocations in the ARTES DB?

ARTES provides information on technical terms in various specific domains and on GCs found in those various domains. ARTES thus offers a template for collecting discourse phraseology, and in turn, for navigating through domain-free phraseology. There are 1 958 GCs recorded in the DB currently, and some 150-200 GCs added every year. The data is recorded in the DB by our Masters' students. Each student (that is 30 to 40 students per year) must record 5 GCs selected in a text on which they perform a translation task. Creating GC records (as well as term records and performing a translation task) is a part of the assignment they must accomplish within their Master's Dissertation.

To record a GC in this DB, the students must devise the lemmatised form of the GC they identified in their text, provide information on a discourse type where such collocation is recurrent, the type of syntactic construction of the GC, language, a context (that is, an

authentic example) to illustrate the typical usage of the GC, the source of the context, and possibly add a comment (Fig 1).

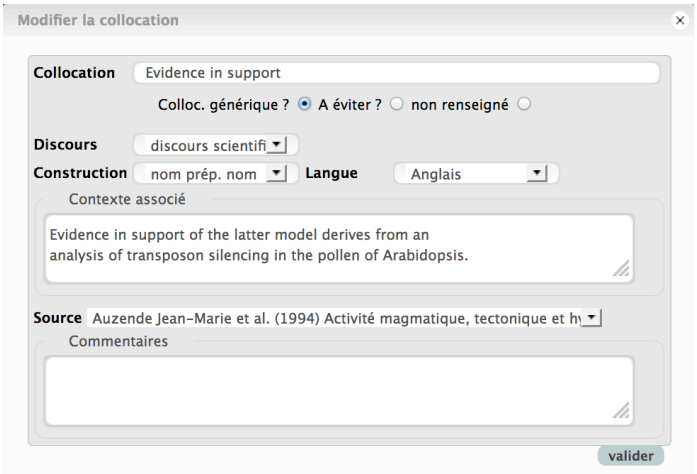

Figure 1 The form for recording generic collocations in the ARTES DB

The next step consists in assigning a discourse function to the GC and finding an equivalent pattern in a target language through target language corpus investigation (Fig 2).

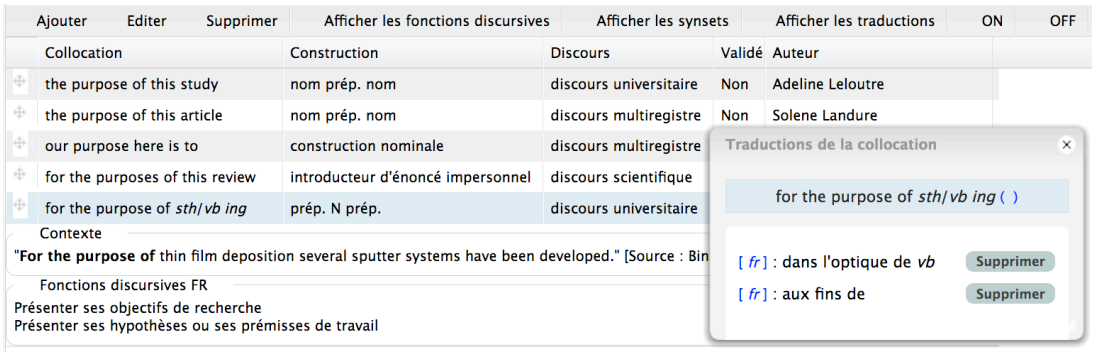

Figure 2 Examples of a pattern involving the lexical noun *purpose* in the ARTES DB

This step is generally considered to be complex by students. However, the difficulty depends on the GC they selected. Identifying the pragmatic function(s) expressed by the GC in a specific register can be more or less obvious. This identification consists in analysing the information structure of a discourse or register. Figure 2 shows the examples of a pattern with the lexical item *purpose* recorded by students and which illustrate at once:

- the attribution of a discourse function to a GC: in this case, the GC *for the purpose of sth/vb ing* was analysed as expressing the discourse function of 'presenting research objectives' and 'presenting hypothesis or work premises',
- the attribution of equivalent GC in a target language: in this case, the GC *for the purpose of sth/vb ing* was analysed as a pattern equivalent to *dans l'optique de vb* and *aux fins de* in the French language, and
- the relative homogeneity of recording strategies adopted by various students as the outcome of the identification and analysis process of this productive pattern: *the purpose of this study/article, for the purpose of sth/vb ing, for the purposes of this review, our purpose here is to*.

Students attend classes on terminology and phraseology, and on ARTES architecture to help them select the appropriate discourse function, register type and find equivalences in other languages. There are some 80 discourse functions listed in the DB devised on the basis of Pecman's study on General Scientific Language (Pecman 2004, 2007), a dozen register types (scientific, legal, political, ethical, economical, academic, multiregister, etc.), and about one hundred syntactic patterns. The discourse functions listed in ARTES are strictly based on scientific discourse, which explains one of the difficulties encountered by students in classifying GCs belonging to other discourse types. However, other major difficulties come from the functional, pragmatic and lexico-syntactic analysis that must be conducted prior to recording GCs in the DB. Such analysis should guarantee the quality of the resource collection in the DB and of handling GC during translation. Consequently, the teaching of GCs with the ARTES DB seems, despite the difficulties enumerated in the previous section, to be an excellent opportunity for guiding students through multiple-level language analysis.

5. Conclusion

Despite difficulties, teaching 'generic collocations' appears to be an innovative and constructive method for improving the abilities of trainee translators to analyse language. They learn about the regular phraseology of scientific discourse, but also about the diversity and productivity of multi-word expressions.

Using the ARTES database in the specific teaching context of a Master's in Specialised Translation has led us to re-examine our approach to GCs and to identify the next steps in developing this approach towards the more efficient treatment of GCs within the ARTES DB. As GCs are very productive sequences which play a central role in building the pragmatic structure of a discourse, we feel that the ability to analyse GCs by paying specific attention to their productive capacities and discourse functions is an important competence in the skillset for trainee translators. As mentioned in the 2017 version of the *European Master's in Translation Competence Framework,* students of translation should know how to:

> "[…] Acquire, develop and use thematic and domain-specific knowledge relevant to translation needs […] mastering systems of concepts, methods of reasoning, presentation standards, terminology and **phraseology** [...]" (European Commission, 2017: 8 [our emphasis]).

Thus an understanding of phraseology (which we defined very broadly as 'the preferred way of making meaning in a particular discourse') is argued to be a key tool for trainee translators, on a par with central language skills such as a presentation standards and terminology. It occurs to us that the analysis of generic collocations is therefore a key exercise, especially when trainee translators and writers (who are often not writing in their first language) are unsure of the conventional formulations used in a very specific domain.

As for improving the template offered by the ARTES DB for processing GCs, it is clear that we need to 1) simplify and rationalise how GCs are categorised in the DB, 2) re-examine how discourse functions are categorised in the DB in order to account for a variety of discourse types and registers, 3) rework the ARTES DB data consultation interface for GCs, and 4) explore the possibilities for exploiting GC resources for creating a tool for assisted writing of scientific abstracts and articles.

References

BIBER, (D.), CONNOR (U.) & UPTON (T. A.) 2007, *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, Amsterdam: John Benjamins Publishing.

COXHEAD (A.) 2000, "A new academic word list", in *TESOL Quarterly* 34/2, pp. 213-238.

COXHEAD (A.) 2002, "The Academic Word List: A Corpus-based Word List for Academic Purposes", in *Teaching and Language Corpora (TALC) 2000 Conference Proceedings*, Atlanta, Rodopi, pp. 73-89.

COXHEAD (A.) & HIRSH (D.) 2007, "A pilot science-specific word list", in *Revue française de linguistique appliquée* 12/2, pp. 65-78.

EUROPEAN COMMISSION 2017, *European Master's in Translation Competence Framework.* Available at: <https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf>.

GLEDHILL (C.) 2000a, *Collocations in science writing.* Language in Performance Series No. 22. Tübingen, Gunter Narr Verlag.

GLEDHILL (C.) 2000b, "The Discourse function of collocation in research article introductions", in *English for Specific Purposes* 19, pp. 115-135.

GLEDHILL (C.) & KÜBLER (N.) 2015, "How Trainee Translators Analyse Lexico-Grammatical Patterns", in *Phraseology, Phraseodidactics and Construction Grammar(s)*, M. I. González-Rey ed, Special issue of *Journal of Social Sciences* 11/3, pp. 162-178.

GLEDHILL (C.) & KÜBLER (N.) 2016, "What can linguistic approaches bring to English for Specific Purposes? ", in *Anglais de spécialité* 69, pp. 65-95.

GLEDHILL (C.), PATIN (S.) & ZIMINA (M.) 2017, « Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français », in *Corpus* 17, pp. 113-144.

GRANGER (S.) & PAQUOT (M.) 2015, "Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid", in *Lexicographica - International Annual for Lexicography / Internationales Jahrbuch für Lexikographie* 31/1, pp. 118-141.

GROOM (N.) 2009, "Phraseology and epistemology in academic book reviews: a corpus-driven analysis", In *Academic Evaluation: Review genres in university settings*, K. Hyland and G. Diani eds, Basingstoke: Palgrave Macmillan, pp. 122-139.

KRAIF (O.) & DIWERSY (S.) 2012, « Le Lexicoscope : un outil pour l'étude de profls combinatoires et l'extraction de constructions lexico-syntaxiques », in *Actes de la conférence conjointe JEP-TALN-RECITAL*, Grenoble 4-8 juin 2012, volume 2: TALN, pp. 399–406,

KÜBLER (N.) 2004, "Using Webcorp for building specialized dictionaries" In *Advances in Corpus Linguistics*, K. Aijmer K ed. Proceedings of the ICAME Conference, May 2002, Göteborg, Suède. Amsterdam: Rodopi, pp. 387-400.

KÜBLER (N.) & PECMAN (M.) 2012, "The ARTES bilingual LSP dictionary: from collocation to higher order phraseology", In *Electronic lexicography*, S. Granger and M. Paquot eds, Oxford, Oxford University Press, pp. 186-209.

MESTIVIER VOLANSCHI (A.) 2008, *Étude et modélisation des phénomènes collocationnels : Implémentation dans un système d'aide à la rédaction en anglais scientifique*, Thèse de doctorat, 5 déc. 2008, Dir. Natalie Kübler, Université Paris Diderot.

PECMAN (M.) 2004, *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique*. Thèse de doctorat. 9 déc. 2004. Dir. Henri Zinglé. Université Nice Sophia Antipolis.

PECMAN (M.) 2005, « Les apports possibles de la phraséologie à la didactique des langues étrangères », in *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)* 8/1, pp. 109-122.

PECMAN (M.) 2005, "Systemizing the notation and the annotation of collocations", in *Jezikoslovlje* 6/1, Osijek (Croatia), pp. 79-93.

PECMAN (M.) 2007, « Approche onomasiologique de la langue scientifique générale », in Lexique des écrits scientifiques, A. Tutin éd., *Revue française de linguistique appliquée* XII/2, pp. 79-96.

PECMAN (M.) 2008, "Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions", in *Phraseology in language learning and teaching,* S. Granger and F. Meunier eds, Amsterdam/Philadelphia, John Benjamins, pp. 203-222.

PECMAN (M.) 2012, « Etude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques », *Terminology, Phraseology and Translation,* M. Rogers ed, special issue of *The Journal of specialised translation (JoSTrans)* 18, pp. 113-138.

PECMAN (M.) & KÜBLER (N.) 2011, "ARTES: an online lexical database for research and teaching in specialized translation and communication", *Proceedings from International Workshop on Lexical Resources (WoLeR)* 2011 at ESSLLI. August 1-5 2011, Ljubljana, Slovenia, pp. 86-93.

SIMPSON-VLACH (R.) & ELLIS (N.) 2010, "An Academic Formulas List: New Methods in Phraseology Research", in *Applied Linguistics* 31/4, pp. 487-512.

TEUFEL, (S.) 1998, "Meta-discourse markers and problem-structuring in scientific articles", *ACL 1998 Workshop, Discourse Structure and Discourse Markers*. Montreal, Somerset/New Jersey, Manfred Stede and Leo Wanner and Eduard Hovy, pp. 43-49.

TRAN (T. T. H.), TUTIN (A.) & CAVALLA (C.) 2016, « Pour un enseignement systématique des marqueurs discursifs à l'aide de corpus en classe de FLE: l'exemple des marqueurs de reformulation », *Linguistik Online*, Bern Open Publishing, Corpus, grammaire et français langue étrangère : une concordance nécessaire, 78/4, pp.113-128.

TRIBBLE, (C.) 2011. "Revisiting apprentice texts: using lexical bundles to investigate expert and apprentice performances in academic writing", in *A Taste for Corpora. In honour of Sylviane Granger,* F. Meunier, S. De Cock, G. Gilquin and M. Paquot Eds, Amsterdam: John Benjamins, pp. 85-108.

TUTIN (A.) 2007a, « Traitement sémantique par analyse distributionnelle des noms transdisciplinaires des écrits scientifiques », in *Actes de TALN*, 5-8 juin 2007, Toulouse, pp. 283-292.

TUTIN (A.) 2007b, « Modélisation linguistique et annotation des collocations : application au lexique transdisciplinaire des écrits scientifiques », in *Formaliser les langues avec l'ordinateur*. S. Koeva, D. Maurel et M. Silberztein Eds, Besançon, Presses universitaires de Franche-Comté, pp. 189-215.

TUTIN (A.) et GROSSMANN (F.) Eds, 2013, *L'écrit scientifique : du lexique au discours. Autour de Scientext,* Rennes, Presses Universitaires de Rennes.

---

[i] ARTES (Aide à la Redaction de TExtes Scientifiques) database interface: https://artes.eila.univ-paris-diderot.fr and ARTES project webpage: http://www.eila.univ-paris-diderot.fr/recherche/artes/index. The GCs can be consulted via DB interface in the tab *Phraséologie discursive*.

[ii] Master Industrie de la langue et traduction spécialisée (ILTS) webpage: http://www.eila.univ-paris-diderot.fr/formations-pro/masterpro/ilts/index

[iii] *Termium,* La banque de données terminologiques et linguistiques du gouvernement du Canada : http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.htm

[iv] *Grand Dictionnaire Terminologique,* Dictionnaire de l'Office québécois de la langue française : http://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.htm

[v] *Grand Dictionnaire Terminologique,* Dictionnaire de l'Office québécois de la langue française : http://phraseotext.u-grenoble3.fr/lexicoscope