Verifying the General Academic Status of Academic Verbs: An Analysis of co-occurrence and Recurrence in Business, Linguistics and Medical Research Articles

Natassia Schutz

Centre for English Corpus Linguistics, Université catholique de Louvain
E-mail: natassia.schutz@uclouvain.be

Abstract

General academic vocabulary lists have been the subject of much debate. Because they focus on single words, they have been criticized for not considering "the importance of contextual environments which reflect different disciplinary practices" (Hyland & Tse 2007: 251). This study aims to provide insight into the reliability of such vocabulary lists by analyzing cross-disciplinary phraseological variation. To do so, I analyze the collocations and lexical bundles used with c. 30 academic verbs found in a 3-million-word corpus containing research articles in business, linguistics and medicine. The results seem to suggest that there are sufficient commonalities, both in terms of use and meaning, to justify the creation and use of general academic vocabulary lists. In addition to their discipline-specific uses, many of the verbs under focus also have general academic uses that relate to the core business of research, irrespective of the academic discipline (e.g. *provide + information/ insight* and *as can be seen in*). The results of this study also demonstrate the benefit derived from adopting a bottom-up approach to phraseology, as it identified a considerable number of verb-based patterns that are not found in existing corpus-driven academic phraseology lists.

Keywords: academic vocabulary, verbs, collocations, lexical bundles

1   Introduction

While research in English for Academic Purposes (EAP) discovered quite early on that academic disciplines can differ in the way they use language to construct knowledge, the increase in the number of general EAP courses across the world acted as a catalyst for the search of a teachable common-core (de Chazal 2013). This led to a growing interest in the linguistic devices found to be common to various academic disciplines, and thus useful for mixed groups of EAP learners. The linguistic device that has undoubtedly attracted the most attention is academic vocabulary, i.e. the vocabulary that is "neither highly technical and specific to a certain field of knowledge, nor obviously general in the sense of being everyday words which are not used in a distinctive way in specialized texts" (Baker 1988: 91). One of the reasons for this is that academic vocabulary is said to be the most difficult type of vocabulary for EAP learners, as it is "not central to the topics of the texts in which they occur" (Coxhead 2000: 214) and "tend[s] to pass unnoticed" (Granger 2017: 9) – as opposed to technical vocabulary.

To meet this need, quite a number of general academic vocabulary lists have been created. Before the advent of corpus linguistics, academic vocabulary lists were based on manual frequency analyses of small corpora (Campion & Elley 1971; Praninskas 1972), the annotations found in students' textbooks (Ghadessy 1979; Lynn 1973), or a combination of both (the University Word List, Xue & Nation 1984). Twenty years later, the need for a more representative and up-to-date vocabulary list was felt, and EAP scholars set out to propose academic vocabulary lists based on the analysis of EAP corpora. The very first corpus-based academic vocabulary list, viz. the 'Academic Word List'

(AWL, Coxhead 2000), quickly met with great success. The AWL is based on a 3.5-million-word corpus of academic texts in various disciplines and contains 570 word families (i.e. a headword and its inflectional and derivational affixed forms, e.g. *authority*, *authorities* and *authoritative*) sorted in decreasing order of word family frequency. Notwithstanding its popularity, this list also met with some criticism. For example, it excludes high-frequency words on the grounds that learners should already master these vocabulary items. Research has however demonstrated that such vocabulary items can have academic uses as well, e.g. *gain weight* vs. *gain insight* (Martínez et al. 2009; Paquot 2007; Schutz 2013). Another issue is its organization according to word family. It has been shown, for instance, that some members of word families are not always very frequent in academic English (e.g. *establish* vs. *disestablish*) and do not always share the same core meaning as their headword (e.g. *react* vs. *reactivation*) (Gardner & Davies 2013). EAP scholars have recently attempted to address these weaknesses by using more empirical vocabulary extraction methods based on statistical analyses and by analyzing word lemmas only. This resulted in the creation of the 'Academic Keyword List' (AKL, Paquot 2010) and the 'Academic Vocabulary List' (AVL, Gardner & Davies 2013). The former is based on a two-million-word corpus of academic texts and contains 930 words. The latter is based on a 120-million-word corpus and contains over 3,000 words.

While such lists have attracted considerable attention and have been extensively used to write EAP teaching materials, the possible existence of general academic vocabulary has however also been questioned. Hyland and Tse (2007: 238), for example, question "the assumption that a single inventory can represent the vocabulary of academic discourse and be valuable to all students irrespective of their field of study". They support their argument by showing that a number of vocabulary items found in the AWL (1) are not evenly distributed, and (2) show semantic variation across academic disciplines (e.g. the noun *volume* is mostly used to refer to a book in applied linguistics and sociology whereas, in the hard sciences, it refers to a type of quantity; ibid.: 246). Those scholars defending the idea of general academic vocabulary, on the other hand, argue that potential cross-disciplinary variation is not a reason to "throw out generalized word lists altogether" (Gardner & Davies 2013: 6). Gardner and Davies (ibid.: 2), for instance, underline the importance of such lists in helping EAP practitioners establish learning goals and design learning materials and tools. Granger and Paquot (2009) and Ming-Tzu and Nation (2004) further argue that it is possible to teach many vocabulary items by focusing on the central concept found behind the variety of uses: e.g. the verb *measure* should be described as referring to the 'action of determining the size, amount, level, etc. of something' despite the fact that disciplines use different methods, data and criteria (Granger & Paquot 2009). When analyzing the weight of general academic verbs compared to that of discipline-specific verbs, Schutz (2013) demonstrated the importance of general academic verbs as they represent over half of the verb tokens occurring in a corpus of research articles in business, linguistics and medicine; the verbs that were considered as discipline-specific only represented around 5% of the verb tokens occurring in each discipline. These results provide additional evidence that general academic vocabulary lists should not be discarded, but rather further investigated to best help EAP practitioners and learners.

The aim of the present paper is to provide insight into the reliability of general academic vocabulary lists by analyzing the collocations and lexical bundles used with general academic verbs in three strongly contrasting academic disciplines: business, linguistics and medicine. More specifically, the objective is to determine the extent to which general academic verbs have shared academic meanings and phraseological patterns. While the phraseology of academic English has been the object of quite a number of studies (see for example, Ackermann and Chen (2013) and Durrant (2009) for an analysis of collocations, and Simpson-Vlach and Ellis (2010) for an analysis of lexical bundles), none of these studies fully answer the criticism leveled against general academic vocabulary lists, as they adopted a textual rather than a lexical approach to phraseology. To best inform what can be

called the specificity debate (i.e. the debate centered around the question as to whether general EAP teaching is worthwhile, and thus the degree of specificity that such courses should adopt) this paper takes general academic verbs as a starting point to better identify their general academic uses vs. their discipline-specific uses.

2   Methodology

2.1   The Data

This study makes use of a sub-corpus of the *Louvain Corpus of Research Articles*[1] which totals 3,035,510 words and contains 421 research articles from peer-reviewed top-rated journals in three different academic disciplines: business, linguistics and medicine (hereafter LOCRA, BUS, LING and MED) (cf. Table 1).

Table 1: The LOCRA sub-corpus.

Disciplines	Number of texts	Number of words
Business	116	1,053,479
Linguistics	109	1,004,829
Medicine	196	977,202
TOTAL	421	3,035,510

In order to identify the verbs occurring in LOCRA, the three sub-corpora were lemmatized and POS-tagged with WMatrix (Rayson 2009) using the Constituent Likelihood Automatic Word-tagging System (CLAWS7) (Gardside & Smith 1997). As the tagset includes different tags for each verb form (e.g. VV0 for the base form or VVD for the past tense), a Perl program was applied to the CLAWS output so as to simplify the verb tags and conflate them into a single VV tag (cf. Granger & Paquot 2009).

2.2   Academic Verb Selection

Rather than selecting verbs that are found in existing general academic vocabulary lists, this paper zooms in on the verbs that stand out as being typical of the academic corpus under study. To do so, I join the forces of two different vocabulary extraction methods which have hitherto never been combined: the keyness analysis (e.g. the AKL, Paquot 2010) and the analysis of traditional frequencies (e.g. the AWL, Coxhead 2000). The novelty of this combined selection procedure is that it takes into account different types of verbs that have never been considered so far in the context of academic English (e.g. *find*, *make* and *see*) (cf. Schutz 2013; 2017). To be considered as general academic verbs, the verbs occurring in LOCRA had to be identified as either key (Scott 2001) or highly frequent across BUS, LING and MED. Key verbs are those which "occur with unusual frequency in a given text" (Scott 2001: 236) when compared to a "strongly contrasting reference corpus" (Tribble 2001: 396). In this study, WordSmith Tools 5 (Scott 2008) and a corpus of fiction writing, viz. the one-million-word fiction sub-corpus of the Baby British National Corpus, were used to extract the key verbs occurring in LOCRA. The verbs that were considered as highly frequent are those that appear among the top verbs covering up to 80% of the total number of verb tokens in each discipline (cf. Coniam 1999).

---

1   https://uclouvain.be/en/research-institutes/ilc/cecl/locra.html

The comparison of the academic verbs occurring in BUS, LING and MED revealed that 177 academic verbs were used across all three disciplines. These cover 62%, 54% and 69% of the total number of verb tokens occurring in BUS, LING and MED, respectively (cf. Schutz 2013). To reduce this list to one that is more manageable for the purpose of this study, this paper focuses on the top 15 general academic verbs occurring in each discipline. When merging the top 15 academic verbs used in each discipline, the final verb list totals 31 verbs (see Table 2), which represent a sizeable 26% of the verb tokens occurring in LOCRA.

Table 2: Top 31 academic verbs occurring across BUS, LING and MED.

appear, associate, base, consider, compare, describe, determine, develop, examine, express, find, follow, give, include, increase, indicate, influence, involve, make, observe, occur, perform, provide, receive, relate, report, see, show, suggest, take, use

2.3   The Analysis of Phraseology

For the analysis of phraseology, this study makes use of automated tools to analyze co-occurrence and recurrence. While the former seeks to identify the words that co-occur with a specific node more often than by chance (viz. significant collocates; Sinclair et al. 1970: 150) the latter focuses on "recurrent expressions [of three or more words], regardless of their idiomaticity, and regardless of their structural status" (viz. lexical bundles; Biber et al. 1999: 990). The reason for choosing these two types of analyses is that they describe two different aspects of multi-word units (MWUs) that have rarely been studied concurrently. It is therefore hoped to provide a more complete description of cross-disciplinary phraseology, and thus better inform the specificity debate.

The collocates that are used with the 31 verbs under focus were extracted thanks to the Word Sketch option of the Sketch Engine (SkE; Kilgarriff, Rychly, Smrz & Tugwell 2004). This tool automatically extracts the collocates of a specific node (using the logDice measure; Rychlý 2008) and categorizes them according to their grammatical function. As illustrated in Figure 1, Word Sketch identifies, for example, the words *support*, *evidence* and *difference* as the object collocates of *find* in BUS. To ensure the pedagogical relevance of the collocates we focus on, an additional frequency threshold of a minimum of five occurrences with the node was set.

Figure 1: Word Sketch of *find* in BUS.

To compare the collocates used in BUS, LING and MED, a simple three-step procedure was adopted. First, the collocate lists obtained for each verb in each discipline were extracted from Word Sketch. The different collocate lists extracted for each verb were then automatically compared so as to identify the collocates that are used across BUS, LING and MED vs. those that are used in one discipline only, i.e. the potential cross-disciplinary and discipline-specific collocates.

As regards the analysis of recurrence, a Perl script was used to extract the three-to-10-word lexical bundles occurring at least five times with the academic verbs under focus in BUS, LING and MED. After having identified the bundles used with each verb, the script then automatically generated the list of shared and discipline-specific bundles. While this considerably accelerated the extraction procedure, the bundle list then needed to be cleaned up as the output also contained, for example, bundles that did not include the actual verbs under focus (e.g. *increasing number of* or *the following variables*). Because of the extremely large number of bundles that were extracted for some verbs, this last step was restricted to the bundles that were found across BUS, LING and MED.

3   Results

3.1   The Analysis of Co-occurrence

The significant collocates extracted from LOCRA were categorized into 10 different grammatical categories: subject, object, modifier, prepositional complement, *wh-* complement, infinitive complement, particle, object complement, adjective complement and *–ing* complement. In the following sections, we give particular attention to the subject and object collocates used in LOCRA, as these are the collocates for which we find the most significant and conclusive findings.

3.1.1 The Shared and Discipline-specific Collocates

Out of the 31 verbs under focus, 22 verbs were found to share significant subject collocates across all three disciplines: *appear*, *associate*, *consider*, *compare*, *describe*, *examine*, *find*, *follow*, *include*, *indicate*, *influence*, *involve*, *make*, *observe*, *provide*, *receive*, *report*, *see*, *show*, *suggest*, *take* and *use*. As regards object collocates, 14 verbs were found to show cross-disciplinary similarities: *base*, *compare*, *examine*, *find*, *give*, *include*, *increase*, *make*, *perform*, *provide*, *report*, *show*, *take* and *use*. Tables 3 and 4 list the shared subject and object collocates that were identified in LOCRA and indicate which academic verbs they were generally found to co-occur with. The qualitative analysis of these collocates revealed that most of them could be grouped under various research-related semantic categories. The majority of the shared subjects were grouped under the categories RESEARCHER(S), RESEARCH, FRAMEWORK and INFORMATION (cf. Table 3). Most of the shared objects were grouped under the categories FRAMEWORK/METHOD, INFORMATION, RELATIONSHIP, RESEARCH and PHENOMENON (cf. Table 4). The remaining shared collocates were grouped under the category OTHER.

As can be seen from below, the shared collocates identified in LOCRA clearly relate to the core business of research, irrespective of the discipline. They are all (except for the subjects *it*, *they* and *we* and the objects *it*, *place*, *time* and *detail*) listed in recent general academic vocabulary lists (the AKL and/or the AVL). However, only a couple of the shared collocational pairs identified in LOCRA also appear in existing general academic collocation lists: only 10 verb-object patterns were found to overlap with those listed in the Academic Collocation List (ACL, Ackermann & Chen 2013) (e.g. *use* + *approach*/*method*/*strategy*/etc. and *provide* + *opportunity*/*data*/*support*/etc.). The reason for this small overlap is that Ackermann and Chen adopted a textual approach to phraseology whereas we adopt a lexical approach. In other words, Ackermann and Chen concentrated on the most frequent collocational pairs appearing in academic English, whereas this study focuses on those used with a particular node.

Table 3: Shared subject collocates occurring in LOCRA.

Semantic category	Shared subjects	+	Shared academic verbs
RESEARCHER(S)	we, they, study (metonymic use)		*compare, consider, describe, examine, find, follow, include, make, observe, provide, report, see, show, suggest, take, use*
RESEARCH	study, analysis		*appear, include, indicate, involve, show, suggest*
FRAMEWORK	model	+	*suggest*
INFORMATION	score, evidence, data, factor, result, finding		*indicate, influence, show, suggest*
OTHER	it, they		*appear, associate, provide, receive, show, suggest, use*

Table 4: Shared object collocates occurring in LOCRA.

Shared academic verbs	+	Semantic category	Shared objects
use		FRAMEWORK/ METHOD	*approach, criterion, instrument, measure, method, model, procedure, strategy, system, technique, test*
compare, find, give, increase, provide, report, show, use	+	INFORMATION	*data, detail, estimate, evidence, explanation, information, insight, level, number, rate, result, sample, score*
compare, examine, find, show		RELATIONSHIP	*correlation, difference, effect, relationship*
base, include, perform, use		RESEARCH	*analysis, study*
show		PHENOMENON	*pattern*
give, make, provide, take		OTHER	*advantage, assumption, contribution, decision, it, opportunity, place, rise, time, support*

The quantitative analysis of what these collocates cover compared to those that are discipline-specific revealed a rather complex picture of verb patterning: academic verbs appear to show different phraseological preferences according to the discipline they are used in. For example, our results indicate that the verbs under focus show varying degrees of formulaicity according to the discipline they occur in (e.g. the verb *receive* seems to have preferred discipline-specific collocates in medicine, such as *infant/patient/mouse/women + receive + care/therapy/placebo*, but not so much in linguistics). Similarly, some verbs appear to have shared collocates showing different coverage values in different disciplines (e.g. the shared subject collocates of the verb *find*, viz. *we* and *study*, cover 86% of *find's* total number of subject co-occurrents in medicine while they only cover 36% in linguistics). Despite this finding, our results suggest that, generally speaking, both commonalities and discipline-specificities are important. For quite a number of verbs, whether they show a preference for shared and/or discipline-specific collocates, we found that both types of collocates are often sufficiently frequent to be considered in the context of EAP teaching. Typical examples of such verbs can be found in Table 5.

A closer look at each grammatical category also reveals interesting trends as regards the frequency of shared subject and object collocates. As can be seen from Figure 3, the shared subject collocates often tend to cover more than those that are discipline-specific. Figure 4, on the other hand, shows that the shared object collocates mostly cover as much as or less than the discipline-specific object collocates. Bearing in mind the complexity of cross-disciplinary verb patterning and also the fact that

Table 5: Examples of frequent shared and discipline-specific collocates.

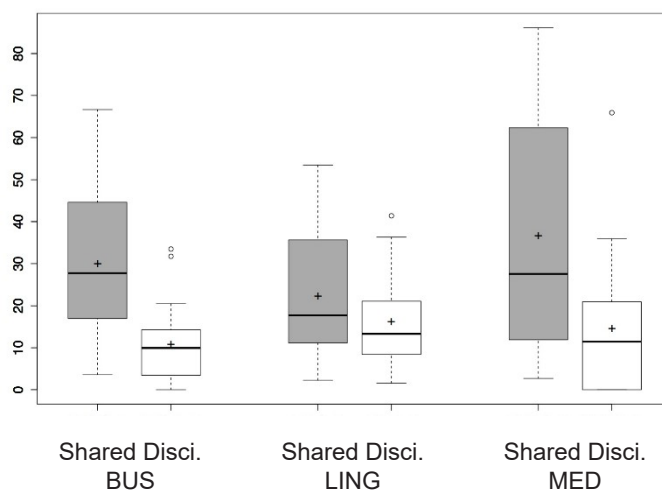| Verbs	Shared subject collocates (coverage/total number of subject co-occurrents)	Discipline-specific subject collocates (coverage/ total number of subject co-occurrents)
show	study, analysis, we, data, etc. (MED = 35%)	biopsy, cell, microscopy, etc. (MED = 25%)
report	study, we, they (LING = 25%)	he, learner, student, etc. (LING = 28%)
Verbs	**Shared object collocates (coverage/total number of object co-occurrents)**	**Discipline-specific object collocates (coverage/ total number of object co-occurrents)**
provide	evidence, insight, information, etc. (BUS = 29%)	access, firm, value, etc. (BUS = 22%)
show	evidence, correlation, result, etc. (LING = 19%)	interaction, meaning, variation, etc. (LING = 15%)

Figure 2: Coverage (%) of the shared and discipline-specific subject collocates used with the 22 verbs sharing subject collocates.
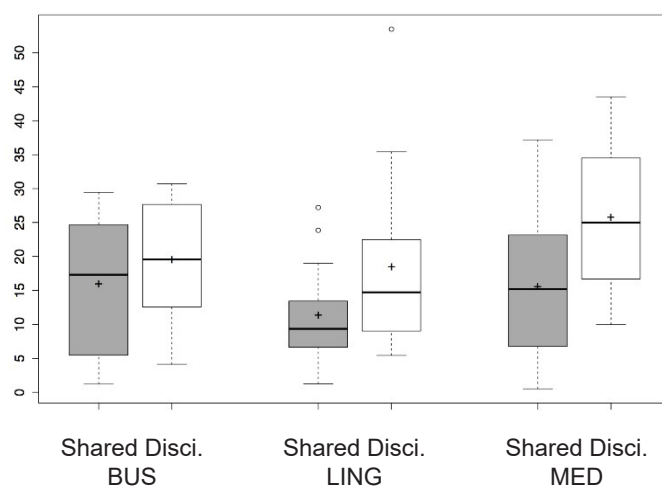

Figure 3: Coverage (%) of the shared and discipline-specific object collocates used with the 14 verbs sharing object collocates.

box plots can oversimplify results, this finding is still particularly interesting given that existing EAP collocation lists contain very few subject-verb collocational pairs: the ACL does not list any subject-verb patterns and Durrant's (2009) collocation list only lists three (*figure + shows*, *we + assume* and *these + suggest*). Our results, however, suggest that, given the high coverage this type of collocate can have for some verbs (e.g. *study/we + find* and *data/result/study/*etc. *+ suggest* in BUS, LING and MED), shared subject-verb collocational pairs should also have a place in general academic collocation lists.

3.1.2 Core Academic Meaning and Semantic Variation

While the results presented above revealed that the verbs under focus can indeed show variation in collocational patterning, the contextual analysis of the shared and discipline-specific collocation pairs shows that, despite this, many of the academic verbs under investigation can also be used in similar meanings across BUS, LING and MED to describe/report on results and information, and report on research activities, for example. Only a handful of verbs were found to have discipline-specific uses and meanings.

The collocational patterns used to describe/report on results and information in LOCRA contain the verbs *appear*, *associate*, *give*, *find*, *indicate*, *observe*, *provide*, *relate*, *report*, *show* and *suggest*. For most of these verbs, our study identified both shared and discipline-specific significant collocates. For example, the verbs *find* and *observe* were most often used in the meaning 'discover/notice something after a careful examination of data/results/etc.'[2] no matter the context they were used in (see examples 1-6). To take another example, the verb *report* means 'provide information about something' whether it is a researcher, a study, a respondent, a learner or a patient that reports something. The only verbs for which we found no real discipline-specific collocates were the verbs *indicate* and *suggest*. When performing this rhetorical function, these verbs were used in the following collocational patterns: RESEARCH(ER)/ INFORMATION / FRAMEWORK + *suggest* and RESEARCH(ER) / INFORMATION + *indicate*.

(1) […] we ***found*** the *entrepreneurs* […] to be heterogeneous with respect to their initial growth intentions. (BUS)

(2) As a result, hardly any evaluative *words* were ***found*** in this move of the abstract. (LING)

(3) Iron *deficiency* was ***found*** in 46.6% of case patients and 69.4% of controls. (MED)

(4) *Spreitzer and Quinn (1996)*, for example, ***observed*** this social support *effect* in Ford's transformational change programme. (BUS)

(5) In our data, *we* ***observe*** that Mandarin 'meiyou' acts as a signal for self-repair or other-repair, particularly in repair types 1 and 2. (LING)

(6) *We* also ***observed*** *that* greater increases in levels of the 3 biomarkers were associated with significantly higher all-cause and cardiovascular disease mortality […]. (MED)

The verbs used to report on cognitive and physical research activities are the verbs *base*, *compare*, *determine*, *examine*, *include*, *perform* and *use*. Most of these verbs also occurred with both shared and discipline-specific collocates in LOCRA. However, in this case, cross-disciplinary variation was primarily to be found among the object collocates. This suggests that, while researchers in business, linguistics and medicine can undertake similar research activities (e.g. *compare*, *examine*, *use*), they also show differences in their object of study (e.g. *effectiveness*, *verbs* and *cells*) and methodology (*use + index*, *image* and *primer*). Despite the variation found among their object collocates, all the

---

2   Note that all the definitions given in this section are (adapted) from either the *Macmillan Dictionary* or *Longman Dictionary*.

verbs listed above were, in most instances, found to have the same core meaning across BUS, LING and MED. For example, the prepositional verb *base on* was predominantly used in the passive voice in the sense 'to use something as the thing from which something else is developed' no matter the context it occurred in, as illustrated in examples 7-9. Similarly, the verbs *determine*, *include* and *perform* were used across BUS, LING and MED in the sense 'discover something through the examination of evidence/data/etc.', 'take something into consideration for the reported study' and 'carry out/complete an action or an activity', respectively.

(7) Such *strategies* are ***based on*** two different dimensions […]. (BUS)

(8) The *task* is ***based on*** the alphabetic principle: words that have more sounds need more letters to represent those sounds. (LING)

(9) Efficacy and safety *analyses* were ***based on*** the intention-to-treat population (that is, all persons who underwent randomization and received at least 1 dose of medication in the double-blind phase). (MED)

Among the 31 verbs we analyzed in this study, only seven were found to have discipline-specific uses in LOCRA: *develop* (MED), *express* (LING and MED), *give* (MED), *make* (BUS, LING and MED), *perform* (BUS), *receive* (MED) and *take up* (MED). In addition to its shared meaning 'create a new product/method' (e.g. with the objects *vaccine*, *approach*, *product* or *model*), the verb *develop* was found to mean 'begin to be affected by a medical condition' when co-occurring with subjects such as *cell*, *mouse* and *woman*, and objects such as *cancer*, *diabetes* and *lesion* in MED. When used, for example, with the subjects *clause*, *speaker* and *they* and the objects *proposition* and *meaning* in LING, we found that *express* means 'utter or state'. In MED, on the other hand, *express* means to 'produce something' when used, for instance, with the subjects *cell* and *mice*, objects *gene* and *receptor*, and modifiers *constitutively* and *differentially*. As for *make*, we found that this high-frequency verb is very often used in delexical constructions typical of either BUS, LING or MED. In BUS, *make* co-occurred, for instance, with the nouns *payment*, *investment* and *purchase*. In LING, it co-occurred with *judgment*, *recording* and *suggestion*. Finally, in MED, *make* was used with *diagnosis*, *measurement* and *visit*. Thanks to its discipline-specific modifier collocates *effectively*/*well*/*etc.*, we found that the intransitive use of *perform* also has a meaning which seems to be more frequent in BUS: in the sense 'do something with a particular amount of success' when discussing the performance of firms and employees, for instance. Finally, the verbs *receive*, *give* and *take up* appeared to be used as technical terms in MED. When used with objects such as *treatment*, *placebo* and *injection*, *receive* and *give* were used in the sense 'to have/give a particular treatment'. *Take up* was used in the sense 'to absorb or incorporate into itself'.

3.2   The Analysis of Recurrence

While the results presented above already provide considerable insight into the specificity debate, what the analysis of recurrence offers in this study is additional evidence supporting a general approach to academic English. As can be seen from Table 6, 20 of the 31 verbs under focus were found to share frequent lexical bundles across BUS, LING and MED (minimum frequency of 20 occurrences per million words). A closer look at their coverage values reveals that 11 of these verbs are used in shared lexical bundles which cover, on average, at least 15% of the total number of verb tokens occurring in LOCRA. These verbs are *appear*, *associate*, *base*, *consider*, *determine*, *find*, *indicate*, *involve*, *relate*, *show*, *suggest* and *use* (see coverage values in Table 6). In other words, about every six occurrence (at the minimum) of these verbs is used in a shared lexical bundle in LOCRA. As illustrated in Table 7, a closer look at these bundles further shows that many of them represent phrases that would have not been found in an analysis of lexical co-occurrence and that could prove useful to

Table 6: Frequent shared verb-based lexical bundles in LOCRA: ≥ 20 occurrences per million words.

appear	*appears to be, appear to be, it appears that* (mean coverage = 37%)
associate	*associated with the, is associated with, associated with a, are associated with, be associated with, were associated with, associated with an, not associated with* (mean coverage = 32%)
base	*based on the, is based on, based on a, are based on, was based on, is based on the* (mean coverage = 39%)
compare	*as compared with, compared with the*
consider	*considered to be*
describe	*as described in*
determine	*to determine whether, to determine the, determined by the* (mean coverage = 34%)
examine	*to examine the*
find	*we found that, found to be, found that the, found in the, was found to, be found in, were found to, can be found, and found that, been found to, can be found in* (mean coverage = 22%)
follow	*followed by a*
include	*included in the, were included in*
indicate	*indicate that the, results indicate that, indicates that the* (mean coverage = 17%)
involve	*involved in the* (mean coverage = 16%)
provide	*to provide a*
relate	*related to the, is related to, be related to, are related to* (mean coverage = 23%)
report	*reported in the*
see	*can be seen, seen in the, see figure #*
show	*as shown in, been shown to, have shown that, shown in figure, has been shown, shown in table, table # shows, are shown in, shown to be, show that the, has been shown to, as shown in figure, has shown that, have been shown, showed that the, have been shown to, studies have shown, results show that, we show that, # shows the* (mean coverage = 26%)
suggest	*suggest that the, suggests that the, findings suggest that, results suggest that, this suggests that, we suggest that, suggesting that the, suggested that the, to suggest that, these results suggest, these findings suggest, these results suggest that* (mean coverage = 27%)
use	*used in the, was used to, can be used, used as a, be used to, by using the, were used to, are used to, used in this, is used to, was used as, was used for, by using a* (mean coverage = 17%)

EAP students no matter what their academic discipline: e.g. *appears to be*, *can be found in*, *results indicate that*, *as shown in figure* and *by using the*.

When comparing our shared bundles with those found in an existing list of shared academic lexical bundles, viz. the Academic Formulae List (AFL, Simpson-Vlach & Ellis 2010), we found that our results identified frequent shared lexical bundles for verbs that are not described at all in the AFL: for example, those used with the verbs *compare* (e.g. *as compared with*), *indicate* (e.g. *results indicate that*), *describe* (e.g. *as described in*) and *suggest* (e.g. *these findings/results suggest that*). One of the reasons why these bundles were not found in the AFL is because this list only contains the top 200 bundles used in academic English. Given that academic prose has been shown to prefer NP- and PP-based bundles rather than VP-based bundles (Biber et al. 2004; Biber et al. 1999), it is normal that very few of the top bundles contain verb phrases. It thus appears that our approach to lexical bundles can prove beneficial when analyzing individual words, as it enables the identification of frequent verb-based bundles that would not especially appear at the top of corpus-driven academic bundle lists.

Table 7: Examples of frequent shared lexical bundles in LOCRA.

appears to be	Trust *appears to be* the central component that enhances perceived quality […]. (BUS) However, the trend in modern corpus construction *appears to be* toward bigger and broader. (LING) Elevated pulmonary capillary wedge pressure *appears to be* common in those with emphysema and may be an important determinant of pulmonary artery pressure in these patients. (MED)
can be found in	Other examples of counter-intuitive results *can be found in* the empowerment literature and many suggestions have been made as to why empowerment interventions do not succeed. (BUS) Sample items from the language test and the meta-language test *can be found in* the Appendix. (LING) Different DC subsets *can be found in* the lung, each with a functional specialization. (MED)
results indicate that	The *results indicate that* 135 firms […] went public with founder CEOs. (BUS) The *results indicate that* […] English-writing skills show up as early as the second grade. (LING) These *results indicate that* the assessment of immunogenicity after immunization with DNA alone is not a reliable measure of the priming ability of DNA candidate vaccines. (MED)
as shown in figure	*As shown in Figure* 1, stock prices rose after the unexpected deaths of such CEOs. (BUS) This consisted of 12 drawings accompanied by a verb and an NP, *as shown in Figure* 2, where the target response would have been "El vaso se rompi". (LING) *As shown in Figure* 4A, apocynin treatment increased survival of ALS mice […]. (MED)
by using the	It is hoped that *by using the* organizing framework of competence [...] some clarity is offered regarding the issues which are receiving empirical attention and existing gaps. (BUS) The notion of invariance was investigated *by using the* moving word task. (LING) All events were coded *by using the* Medical Dictionary for Regulatory Activities conventions. (MED)

4   Conclusion

The purpose of this paper was to take a closer look at the phraseological patterning of academic verbs with the aim of verifying the reliability of general academic vocabulary lists. To do so, this paper drew on an exploratory cross-disciplinary comparison of the collocational patterns and lexical bundles used with a set of general academic verbs. While our study is limited, for example, by the number of academic disciplines and verbs it focuses on, it nevertheless provides valuable insight into an ongoing debate for which, as pointed out by Nhã (2015: 43), EAP scholars currently lack empirical evidence to support their arguments; most scholars have indeed based their arguments (either in favor of or against general EAP teaching) on but a handful of illustrative examples (e.g. de Escorcia 1985; Bruce 2011; Granger & Paquot 2009; Ming-Tzu & Nation 2004). In this study, we provide empirical evidence that, in addition to showing cross-disciplinary differences, general academic verbs also seem to have general academic meanings, collocates (e.g. *study/results + suggest*) and lexical bundles (e.g. *it appears that*).

More specifically, the analysis of co-occurrence showed that academic verbs have both shared and discipline-specific significant collocates. However, in many cases, variation in collocational patterning does not appear to affect the core meaning of the academic verbs. While our results revealed a rather complex picture of cross-disciplinary variation, many of the verbs under investigation were used in

the same core academic meaning no matter the context; only seven verbs had real discipline-specific uses in LOCRA (in addition, some had cross-disciplinary uses). Our results also showed that cross-disciplinary similarities were often to be found among subject collocates rather than object collocates. This finding is particularly interesting given the fact that subject collocates receive considerably less attention in EAP teaching materials. This type of collocate would thus deserve further investigation in future research.

While the analysis of recurrence did not provide evidence as compelling as that provided by the analysis of co-occurrence, it did however provide additional support for our main finding: academic verbs are also used in frequent shared lexical bundles. In addition to completing our description of cross-disciplinary phraseology, the results yielded by this analysis are just as important given the salience and systematic functionality of lexical bundles (e.g. Biber & Barbieri 2007), the challenge MWUs represent for learners (e.g. Nation 2001) and the importance of lexical bundles for academic proficiency (e.g. Hyland 2008). It is therefore important to include such patterns in general EAP teaching material to best help EAP learners.

More generally, adopting a lexical approach rather than a textual one to phraseology proved particularly valuable in the context of the present study. While a large majority of frequency-driven studies of EAP phraseology adopt a corpus-driven approach, we decided to concentrate on the phraseological patterning of particular nodes. Not only did this enable us to provide a comprehensive picture of cross-disciplinary language variation, it also helped identify phraseological patterns that are not found in existing lists of general academic MWUs. To establish whether there is such a thing as general academic vocabulary, our study also suggests taking a lexical approach to better determine how vocabulary items are used across academic disciplines.

I acknowledge however that this paper was based on a very small corpus. It would thus be necessary to replicate this study using a corpus of academic English containing a wider variety of disciplines and larger number of texts to better identify, for example, the patterns that are cross-disciplinary and those that are discipline-specific. Nevertheless, it is hoped that this paper has shown how, conducted on a larger scale, cross-disciplinary comparisons along the lines presented here can highlight typical phraseological patterns which EAP teachers could use to raise their students' awareness as to how general academic vocabulary behaves across disciplines and in their own field of study.

References

Ackermann, K., Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. In *Journal of English for Academic Purposes*, 12(4), pp. 235-247.

Baker, M. (1988). Sub-technical Vocabulary and the ESP Teacher: An analysis of Some Rhetorical Items in Medical Journal Articles. In *Reading in a Foreign Language*, 4(2), pp. 91-105.

Biber, D., Barbieri, F. (2007). Lexical bundles in university spoken and written registers. In *English for Specific Purposes*, 26(3), pp. 263-286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical Bundles in University Teaching and Textbooks. In *Applied Linguistics*, 25(3), pp. 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, England: Longman.

Bruce, I. (2011). *Theory and Concepts of English for Academic Purposes*. New York: Palgrave Macmillan.

Campion, M., Elley, W. (1971). *An academic vocabulary list*. Wellington: New Zealand Council for Educational Research.

Coniam, D. (1999). Second language proficiency and word frequency in English. In *Asian Journal of English Language Teaching*, 9, pp. 59-74.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. In *English for Specific Purposes*, 23(4), pp. 397-423.

Coxhead, A. (2000). A New Academic Word List. In *TESOL Quarterly*, 34(2), pp. 213-238.

de Chazal, E. (2013). The general–specific debate in EAP: which case is the most convincing for most contexts? In *Journal of Second Language Teaching and Research*, 2(1), pp. 135-148.

de Escorcia, B. (1985). ESP and beyond: a quest for relevance. In R. Quirk, H. G. Widdowson (eds.), English in the World: Teaching and learning the language and literatures. Cambridge: Cambridge University Press, pp. 228-237.

Durrant, P. (2009) Investigating the viability of a collocation list for students of English for academic purposes. In *English for Specific Purposes*, 28(3), pp. 157-169.

Gardner, D., Davies, M. (2013). A New Academic Vocabulary List. In *Applied Linguistics*, 34(3), pp. 305-327.

Garside, R., Smith, N. (1997). 'A hybrid grammatical tagger: CLAWS4'. In R. Garside, G. Leech & A. McEnery (eds.) Corpus annotation: linguistic information from computer text corpora. New York: Addison Wesley Longman, pp. 102-121.

Ghadessy, P. (1979). Frequency counts, word lists, and material preparation: a new approach. In *English Teaching Forum*, 17(1), pp. 24-27.

Granger, S. (2017). Academic phraseology: A key ingredient in successful L2 academic literacy. In *Oslo Studies in Language*, 9(3), pp. 9–27.

Granger, S., Paquot, M. (2009). In search of a General Academic Vocabulary: A Corpus-driven Study. In K. Katsampoxaki-Hodgetts (ed.) Options and Practices of LSP Practitioners. Crete: University of Crete Publications, pp. 94-108.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge: MIT Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. In *English for Specific Purposes*, 27(1), pp. 4-21.

Hyland, K., Tse, P. (2007). Is there an "Academic vocabulary"? In *Tesol Quarterly*, 41, pp. 235-253.

Kilgarriff, A., Rychly, P., Smrz, P. & Tugwell, D. (2004). The Sketch Engine. In G. Williams, S. Vessier (eds.) Proceedings of the Eleventh EURALEX International Congress. Bretagne: Université de Bretagne Sud, pp. 105-116.

Longman Dictionary of Contemporary English Online. Accessed at: https://www.ldoceonline.com [31/03/2018].

Lynn, R. W. (1973). Preparing word lists: a suggested method. In *RELC Journal*, 4(1), pp. 25-32.

Macmillan English Dictionary Online. Accessed at: http://www.macmillandictionary.com [31/03/2018].

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. In *English for Specific Purposes*, 28(3), pp.183-198.

Ming-Tzu, K., Nation, P. (2004). Word meaning in academic English: Homography in the academic word list. In *Applied Linguistics*, 25(3), pp. 291-314.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nhã, V. T. T. (2015). Should an ESP Course be Specific or General? A Literature Review of the Specificity Debate. In *VNU Journal of Science: Foreign Studies*, 31(4), pp. 37-45.

Paquot, M. (2007). Towards a productively-oriented academic word list. In J. Walinski, K. Kredens, & S. Gozdz-Roszkowski (eds.) Corpora and ICT in Language Studies. Frankfurt am Main: Peter Lang, pp. 127-140.

Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London: Continuum.

Praninskas, J. (1972). *American university word list*. London: Longman.

Rayson, P. (2009). *Wmatrix: a web-based corpus processing environment*, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/.

Rychlý, P. (2008). A Lexicographer-Friendly Association Score. In P. Sojka, A. Horák (eds.) Proceedings of Recent Advances in Slavonic Natural Language Processing. Brno: Masaryk University, pp. 6-9.

Schutz, N. (2013). How specific is English for Academic Purposes? A look at verbs in business, linguistics and medical research articles. In G. Andersen, K. Bech (eds.) English Corpus Linguistics: Variation in Time, Space and Genre. Amsterdam: Rodopi, pp. 237-257.

Schutz, N. (2017). Verbs in English for Academic Purposes: a cross-disciplinary corpus driven study. PhD thesis. Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Scott, M. (2001). Comparing corpora and identifying key words, collocations and frequency distributions through the WordSmith Tools suite of computer programs. In P. Ghadessy, A. Henry & R. Roseberry (eds.) Small corpus studies and ELT. Amsterdam: John Benjamins, pp. 47-67.

Scott M. (2008). *WordSmith Tools 5*. Oxford: Oxford University Press.

Simpson-Vlach, R., Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. In *Applied Linguistics*, 31(4), pp. 487-512.

Sinclair, J., Jones, S. & Daley, R. (1970). *English lexical studies*. Department of English, University of Birmingham.

Tribble, C. (2001). Small corpora and teaching writing: towards a corpus-informed pedagogy of writing. In M. Ghadessy, A. Henry & R. Roseberry (eds.) Small corpus studies and ELT: theory and practice. Amsterdam: Benjamins, pp. 381-408.

Xue, G., Nation, I. (1984). A University Word List. In *Language Learning and Communication*, 3(2), pp. 215-229.