Book Section:

Vocabulary in university tutorials and laboratories: Corpora and word lists

Averil Coxhead and Thi Ngoc Yen Dang

INTRODUCTION

The focus of this chapter is vocabulary in university tutorials and laboratories and the extent to which existing word lists of academic English, both single and multi-word units, can help prepare learners for the vocabulary they will encounter in these interactive small group academic contexts. University laboratories (usually in hard sciences, e.g., Biology, Engineering, and Chemistry) and tutorials (usually in soft sciences, e.g., History, Education, and Political Science) are important because they provide learners with opportunities to develop their understanding of the content of their disciplinary subjects and express complex ideas (Basturkmen, 2016). Lecturers from a university in Aotearoa/New Zealand, interviewed as part of a study by Coxhead, Dang and Mukai (2017),emphasised that students needed to verbalisetheir thinking about the content of their studies in small group interactions. Being unable to do so possibly signals a lack of understanding of disciplinary knowledge. Students who were speakers of English as a second or foreign language in the same study suggested that their capacity for small group interaction was hampered a lack of spoken fluency, struggling to keep up with the speed of native English speakers in class, and the highly interactive nature of these speaking events (Coxhead et al., 2017; see also Hunter & Coxhead, 2007). Coxhead, Hunter, Pierard, and Cooke (2008) note that tutorials can contain local cultural references and slang that can be difficult for English for Academic Purposes students who do not have experience or knowledge of such cultural and linguistic information.

Vocabulary in university laboratories and tutorials

The linguistic features of academic spoken English vary according to speech events (Biber, 2006; Dang & Webb, 2014; Schmitt, Cobb, Horst, & Schmitt, 2015), but very little research has differentiated between them(Dang , Coxhead, & Webb, 2017; Simpson-Vlach & Ellis, 2010).The corpus-based study of university laboratories and tutorials by Coxhead et al. (2017) is the only study, to the best of our knowledge, which focuses on lexis in these academic events. Coxhead et al.(2017) analysed the vocabulary of a tutorial corpus (380,078 running words)and a laboratory corpus (137,399 running words) and found a large amount of high frequency vocabulary (i.e., the words that occur very often in a wide range of spoken and written discourse types such as see, know, think, good). They also analysed a range of EAP and ESP textbooks, looking for evidence of focus on vocabulary in tutorials and/or laboratories, and found little on tutorials and nothing on laboratories. That said, Coxhead et al. (2017)found that the textbooks did contain 176 functional phrases or lexical patterns that were recommended for use in spoken interactions at university, but very few of these multi-word units appeared in the tutorial or laboratory corpus. This led Coxhead et al.(2017)to identify and categorise multi-word units which did occur in the corpora (see below).

This chapter follows on from Coxhead et al. (2017) by analysing the laboratory and tutorial corpora from that study using five existing word lists of academic English: three lists of single words and two lists of multi-wordunits (made up of two words or more).These lists were all developed to support learners and teachers in EAP.Of the single word lists, Coxhead's (2000) Academic Word List (AWL) and Gardner and Davies's (2014) Academic Vocabulary List (AVL)were based on written academic corpora while Dang, Coxhead, and Webb's (2017) Academic Spoken Word List (ASWL) was based on spoken academic corpora. The current research looks at the percentage of the corpora which is covered by the word lists (coverage). This is important because research by Dang and Webb (2014) showed that to reach 95% coverage of lectures and seminars, 4,000 word families plus proper nouns

and marginal words (such as um and ah) are needed; while it takes takes 8,000 word families plus proper nouns and marginal words to reach 98%. The higher the coverage of the word lists over the corpora, the more potential the word lists offer learners and teachers.The two multi-word unit lists are Simpson-Vlach and Ellis's (2010) spoken Academic Formulas List (AFL)and Biber, Conrad, and Cortes's (2004) list of lexical bundles from university classroomdiscourse. The focus on multiword units is on whether they support each other by including the same items, and the major functions these items perform in the discourse.

RESEARCH QUESTIONS

RQ1. How useful are existing word lists to students in tutorials?

RQ2. How useful are existing word lists to students in laboratory sessions?

RQ3. Do the existing wordlists support each other by including the same items?

METHODOLOGY

This study draws on two spoken corpora: one for university laboratory sessions and one from university tutorials (see Coxhead et al., 2017 for more). The laboratory corpus (137,399 running words) comprised texts from 10 academic subject areas, sourced from the Michigan Corpus of Academic Spoken English, Limerick-Belfast Corpus of Academic Spoken English, and Newcastle Corpus of Academic Spoken English. The tutorials corpus contains 380,078 running words from nine subject areas, and is made up of texts from the Limerick-Belfast Corpus of Academic Spoken English and the Hong Kong Corpus of Spoken English.

The coverage in the labs and tutorial corpora of the AWL, AVL and ASWL word lists was carried out using the Range programme (Heatley, Nation & Coxhead, 2002). The multi-word units in tutorials and laboratories identified by Coxhead et al. (2017) were compared with

Biber et al.'s (2004) lexical bundles and Simpson-Vlach and Ellis's (2010) spoken AFL formulasusing Antconc (Anthony, n.d.).

Single-word academic lists

The three single-word lists were selected because they are widely known in the case of the AWL and AVL, and based on spoken academic English in the case of the ASWL. Table 1 outlines the key features of these lists, including their unit of counting words, which is important for word list studies. The AVL uses lemmas, the smallest unit of counting. Lemmas contain a stem (apply) and its inflections (applies, applied, applying). The AWL and the ASWL use word families, a larger unit of counting. They include a stem (apply), its inflections (applies, applied, applying) and related derivations (e.g., application,reapply, reapplication).Lemmas distinguish between word classes but word families do not. For example, form (verb) and form (noun) are counted as two lemmas but as only one word family (see Bauer & Nation (1993) for more on word families).

Table 1 shows that the coverage of the three word lists over the original corpora varies from 10% for the AWL up to over 90% for the ASWL. The AWL looked outside the first 2,000 words, represented by West's (1953) General Service List, for academic vocabulary that met Coxhead's selection criteria. However, academic vocabulary can also be found in the high frequency words of English (Nation, 2016), and decisions made in the development of the GSL have an impact on the AWL. Also, EAP learners may not know high frequency vocabulary in English. The AVL and ASWL did not take this approach, and as a consequence, the AVL contains high frequency words which contribute a great deal to the coverage of these lists.TheAVL also contains some proper nouns, including Africa, Anglo, Asia, Darwinian, Dominican, Europe, Tanzania and Greece. The ASWL is divided into proficiency levels, unlike other twoword lists.

Table 1. Key features of the AWL, AVL and ASWL

Word list	Number of items	Corpora	Coverage (%)
Academic Word List	570 word families	Written; 3.5 million words; university reading texts, textbooks, articles, technical reports; 28 subjects areas fomrfour disciplinary groups (Arts, Commerce, Science, and Law)	10%
Academic Vocabulary List	3,000 lemmas; 1,991 word families	Written; 120 million words from COCA (Davies, 2008); journal articles, newspapers, and magazines; nine academic disciplines	Just under 14%
Academic Spoken Word List	1,741 word families	Spoken; 13 million words; 24 subjects areas from hard-pure, soft-pure, hard-applied and soft-applied	90.13%

Lists of academic multi-word units

The three multiword lists used for comparison in this study, Biber et al. (2004), Simpson-Vlach & Ellis (2010), and Coxhead et al. (2017) were chosen because they are made up of four-word sequences. This means that the comparison is between items that are made up of

the same amount of multiword units, rather than trying to compare units made up of two, three or more words.Table 2 outlines key features of these three lists.

Table 2. Key features of the Biber, Conrad, and Cortes (2004), Simpson-Vlach & Ellis (2010) and Coxhead, Dang & Mukai (2017) multiword unit lists

Word list	Number of items	Corpora	Selection criteria
Biber, Conrad, and Cortes (2004)	84 items	Spoken; 1.2 million words; university classroom teaching; Business, Education, Engineering, Humanities, Natural Science, and Social Science	Frequency: 40 times per million words
Simpson-Vlach and Ellis's (2010) spoken Academic Formulas List (AFL)	200 four-word formulas	Spoken; 2.1 million word; lectures, seminars, labs, tutorials, office hours, study groups; Humanities and Arts, Social Sciences, Natural Sciences/Medicine, and Technology and Engineering	Corpus-comparison and expert consultation
Coxhead, Dang & Mukai (2017) Laboratory Multiword Unit	183	Spoken; 137,399 running words	25 occurrences per million words

List			
Coxhead, Dang & Mukai (2017) Tutorial Multiword Unit List	125	Spoken; 380,078 running words	25 occurrences per million words

An important feature of the multiword unit word lists is categorising the items by their function. This means that teachers and learners can identify the purpose of the bundles in discourse and work on which ones to use in their own writing or speaking. Biberet al. (2004) categorised the bundles into stance expressions (e.g. it is important to), discourse organisers (e.g. on the other hand), and referential expressions (e.g. on the basis of).Hyland (2008) analyzed the bundles from Biber et al. (2004) usinga corpus of published academic writing, student dissertation, and thesis writing to explore disciplinary differences. He found limited evidence of shared bundles across Biology, Electrical Engineering, Applied Linguistics and Business Studies. Hyland identified research, text, and participant-oriented bundles in his corpora, and found that Science and Engineering writers used more research-oriented bundles, Applied Linguistics and Business Studies had higher instances or text-oriented bundles, and participant-oriented bundles were more commonly in the academic writing of the Social Science writers. Byrd and Coxhead (2010) investigated lexical bundles from Biber et al. (2004) and Hyland's (2008) study in a corpus of academic written English developed for Coxhead's (2000) AWL study and found 35 lexical bundles in common, including on the basis of, on the other hand, as a result of, the end of the, at the end of, at the same time, the nature of the, in the form of, and in terms of the. Note that the boundaries of four-word lexical bundles can be blurry (at the end of/the end of the).

An issue in multiword units is overlaps between word sequences of three or four words, for example (Byrd & Coxhead, 2010).Following Wood and Appel (2014), Coxhead et al. (2017) broke each 4-word cluster (e.g., the rest of it) into two constituent 3-word clusters (e.g., the rest of, rest of it) to deal with overlap among 4-word clusters. The more frequent cluster was considered as the root structure and the 4th word were classified as a word that commonly occurred with that structure (e.g., the rest of (it)). It should be noted that most of Coxhead et al.'s (2017) multi-word units were made up of words from Nation's (2012) first 1,000 high frequency word list.

RESULTS AND DISCUSSION

The results below for the single word lists focuses on the percentage of coverage over the tutorial and laboratory corpora, with the aim of finding out which list has the highest coverage and would therefore be potentially the most useful for learners who are preparing to take part in these small academic speaking events at university.The single word list analysis of the laboratory corpus (Table 3) revealed that the ASWL has the highest coverage at 90.58%, followed by the AVL (19.65%),and the AWL (2.52%). One reason for the high coverage of the ASWL is the large proportion of high frequency items in this list.Table 3 shows the coverage of each list across the first 1,000-3,000 and beyond 3,000 frequency levels of Nation's BNC (2012) lists. This analysis shows how prevalent high frequency vocabulary (the first 1000-3000 words) is in the laboratory corpus.

Table 3. Coverage of different academic word lists in the laboratory corpus

Relevant BNC/COCA levels	Coverage (%)		
	AWL	AVL	ASWL
1st 1000	0.33	15.51	85.05
2nd 1,000	0.89	2.18	3.57

3rd 1,000	1.22	1.52	1.59
Beyond 3,000	0.08	0.44	0.37
Total	2.52	19.65	90.58

It is important to show the number of word families from each list that occurred in the laboratory corpus, because this shows us how much of the word lists actually occur in the corpora. If word lists contain a large number of items which do not occur in the corpora, then learners might spend precious time and effort on lexical items which do not appear in tutorials and laboratories.Table 4 shows that the ASWL has the largest percentage of word families appearing in the corpus (84.38%). The AWL is next at 61.23%. The AVL has the smallest percentage of word families appearing in the corpus at 41%.

Table 4. Number of AWL, AVL and ASWL word families in the laboratory corpus

BNC/COCA levels	Number of word families		
	AWL	AVL	ASWL
1st 1000	20	197	782
2nd 1,000	108	221	378
3rd 1,000	192	270	256
Beyond 3,000	29	134	53
Total	349	822	1,469
% of word families appearing in the corpus	61%	41%	84%

The analysis of the tutorial corpus again showed much higher coverage by the ASWL (92.35%), followed by the AVL (22.73%), and then the AWL (3.56%) (Table 5). All three lists have higher coverage over the tutorial corpus than over the laboratory corpus.

Table 5. Coverage of different academic word lists in the tutorial corpus

BNC/COCA levels	Coverage (%)		
	AWL	AVL	ASWL
1st 1000	0.4	17.22	86.16
2nd 1,000	1.31	2.82	3.84
3rd 1,000	1.76	2.26	2.23
Beyond 3,000	0.09	0.43	0.12
Total	3.56	22.73	92.35

Like in the laboratory corpus, the ASWL has the largest percentage of words appearing in the tutorial corpus (97.42%)(Table 6). The AWL ranks second (91.23%), followed by the AVL (66.21%).

Table 6. Number of AWL, AVL and ASWL word families in the tutorial corpus

BNC/COCA levels	Number of word families		
	AWL	AVL	ASWL
1st 1000	20	207	826
2nd 1,000	134	265	450
3rd 1,000	297	437	365
Beyond 3,000	69	404	55
Total	520	1,313	1,696
% of word families appearing in the corpus	91.23	66.21	97.42

All in all, the ASWL provided much higher coverage and has a larger percentage of items appearing in the laboratory corpus and tutorial corpus than the two written word lists. These results are to be expected because of the provenance of the ASWL. It was developed from aspoken corpus and includes large amounts of high-frequency words, ascan be seen clearly

from the coverage provided by the most frequent 3,000 BNC/COCA words in Tables 3 and 5. The AVL also contains a reasonably large number of high frequency words, reflected in its fairly high coverage over the two corpora. These findings highlight the value of the ASWL to help learners comprehend labs and tutorials (Coxhead et al., 2017).

The multiword unit word lists analysis (Table 7) shows the overlapping items in the laboratory corpus accounted for 20.77% (38 out of 183 items) (Biber et al.'s list) and 23.50% (43 out of 184 items) (Simpson-Vlach and Ellis's list) of the total multi-words in Coxhead et al.'s list.Thetutorial analysis showedtheBiber list overlapped by 28.8% (36 out of 125 items) and Simpson-Vlach and Ellis's (2010) list overlapped by 24.8% (31 out of 125 items) with the Coxhead list.The appendixes contain the full list of items (including those which occurred in one or two lists). These results suggest that there is a number of core items which are useful for different kinds of academic speech acts, but there are a substantial number from Coxhead et al's (2017) lists of laboratory and tutorials multiword units which do not appear in the other two lists. If learners and EAP teachers are preparing for academic speaking at university, they may find the overlapping list to be a useful starting point. Beyond that group of items, the tutorial and laboratory lists from Coxhead et al. (2017) would perhaps be more useful.

Table 7 shows that among Coxhead et al.'s multi-words that appearing in Biber et al.'s and Simpson-Vlach and Ellis's list, attitudinal/ modality stances such as (that/so/what) you need to (know/be/have) and (to/should) be able to (do)) is the largest group of functions. They occur in both corpora. The next biggest group is topic introduction focus,for example,(if) you look at (the) and I think it (s/was). It was followed by epistemic stance such as (do) you know what (i/you). It should be noted that these groups all belong to stance expression and discourse organizers functions. The dense use of stance expression function reflects the spoken nature of the corpora while the dense use of discourse organizers function reflects

instructors' attempts in organizing and structuring discourse that can facilitate listeners' comprehension under the real time proceeding circumstances (Biber et al., 2004).

Table 7. Number and frequencyof items in Coxhead, Dang, and Mukai's (2017) lists appearing in Biber, Conrad, and Cortes's (2004) and Simpson-Vlach and Ellis's (2010) lists.

Function of the multi-words	Labs		Tutorials	
	Biber et al.'s list	Simpson-Vlach & Ellis's list	Biber et al's list	Simpson-Vlach & Ellis's list
I. Stance expression				
Epistemic stance	3	4	6	1
Attitudinal/ modality stances	17	14	8	11
II. Discourse organizers				
Topic introduction focus	5	10	4	5
Topic elaboration	1	3	2	6
Textual reference	0	1	0	0
III. Referential expression				
Identification focus	5	2	4	1
Imprecision	1	0	3	0
Specification of attributes	3	2	4	2
Time/Place/Text reference	1	1	3	2
Vagueness markers	0	1	0	0
IV. Special conversational functions	2	5	2	3
Total	38	43	36	31

There are three possible reasons for the small overlap between Coxhead et al.'s list and the other two lists. The first reason is the difference in the kind of speech events that three lists represent. Coxhead et al.'s lists focus on either labs or tutorials. In contrast, Biber et al.'s list focuses on classroom teaching while Simpson-Vlach and Ellis's list represent multi-words from a wide range of academic speech events. Second, the corpora used to develop Coxhead et al.'s lists (137,399 words; 380,078 words)were smaller than those used to develop Biber et al.'s list (1.2 million words) and Simpson-Vlach and Ellis's list (2.1 million words). Third, the variation in the selection criteria used in the three studies may be another reason for the modest overlap between Coxhead et al.'s and other two lists of multi-words.

IMPLICATIONS FOR PEDAGOGY

The findings in this study have implications for EAP pedagogy, materials and course design. The comparison of single word lists in this study illustrates how academic spoken language is different from academic written language (Dang et al., 2017). It is not surprising that the AWL had the lowest coverage of the three single word lists because of the way this word list was developed. The analysis of single word lists illustrates the benefits of Dang et al.'s (2017) Academic Spoken Word List for preparing learners for the vocabulary they will encounter in university laboratories and tutorials, and for that they will use in their speaking in these contexts too. The results of this study show that high frequency vocabulary plays a large role in demonstrating content knowledge in academic speaking, so it is important that EAP learners have a strong knowledge of these lexical items. Learners need to be able to recognise them in speaking and be able to use them fluently in speaking in order to keep up with highly interactive and fluency-challenging small group environments. EAP learners need practice in participating in small group discussions in their classrooms because they are an important, targeted and deliberate part of a course that focuses on the vocabulary that is needed for university study.

Our analysis shows quite a large variation in coverage between the single academic word lists, and a number of formulas and bundles that overlap between the multi-word unit lists. It seems clear from our analysis that lists of multi-word units which have been developed from different spoken corpora have a fairly small overlap because of their different origins. The items which overlap could form a useful core of multiword units for speaking in EAP courses. The remaining items in the tutorial and laboratory lists (see Appendixes 1 and 2) would be useful for preparing for speaking in those events, depending on the avenue of future studies for students. The appendixes note the frequency of these items per million in the two corpora, as well as their functions. High frequency items should be focused on first. The categorisation of the four word sequences can help learners and teachers with the purpose of the multiword units (to organise the discourse, to indicate a stance, and so on) both when learners are speaking or listening.

Textbooks and materials designers could draw on the results of this study first of all by considering whether and how they take vocabulary in small group academic speaking into account. They could examine any multi-word units or phrases that they present in existing textbooks and materials, along with functional analysis, and consider using the multi-word unit analysis presented in Table 5 and Appendixes 1 and 2 to decide what units they might focus on and why. The presentation of laboratory data in Appendix 1 and tutorial data in Appendix 2 could be used to inform the development of materials and textbooks for EAP learners who are heading for either the hard or soft sciences.

One of the main limitations of this study is the small tutorial and laboratory corpora. Clearly a larger scale study with millions of words from these two academic speech events is needed to help confirm these findings and enable broader generalisations. Another limitation is the word lists which we selected for our analysis. There are, of course, other word lists which have been developed for specific academic purposes, such as Ackermann and Chen's (2013)

Academic Collocations List and Liu's (2012) list of multiple types of multiword units in academic written texts. We limited our analysis to studies which included spoken corpora in the case of the multiword unit analysis and to widely known and newly developed word lists in the single word analysis.This study is a reminder that teachers and learners and users of word lists need to know how lists were made as well as what they might have to offer.

FUTURE RESEARCH

One avenue for future research is replicationwith larger corpora for tutorials and laboratory. Another avenue could be using other multiword units to the analysis, such as Ackermann and Chen (2013) and Liu (2012). Investigating the integration of findings into EAP courses and materials would also be useful.

CONCLUSION

Participating in and being able to follow small group discussions is a key part of university study. This study highlights the importance of high frequency vocabulary in both tutorials and laboratories, and high coverage of the ASWL over spoken academic corpora, compared to the AVLand the AWL. This study also illustrates the importance of high frequency vocabulary in multiword units in tutorials and laboratories, and the functional analysis shows that these high frequency words are used in particular patterns for particular reasons. This study contributes to our overall understanding lexis in the university space, including commonly used tools such as word lists.

REFERENCES

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

Anthony, L. (n.d.)Antconc [Online]. Available at:

   http://www.laurenceanthony.net/software/antconc/ (Accessed 15 March, 2018)

Basturkmen, H. (2016)'Dialogic interaction' in Hyland,K. & Shaw,P. (eds.)The Routledge

   handbook of English for Academic Purposes. London: Routledge, pp. 152–164.

Bauer, L., & Nation, I. S. P. (1993)'Word families',International Journal of Lexicography, 6,

   253–279.

Biber, D. (2006)University language: A corpus-based study of spoken and written registers.

   Amsterdam: John Benjamins.

Biber, D., Conrad, S. & Cortes, V. (2004)'If you look at…: Lexical bundles in University

   teaching and textbooks',Applied Linguistics, 25(3), 371–405.

Byrd, P. & Coxhead, A. (2010). 'On the other hand: Lexical bundles in academic writing and

   in the teaching of EAP',University of Sydney Papers in TESOL, 5, 31-64.

Coxhead, A. (2000)'A new academic word list',TESOL Quarterly, 34(2), 213−238.

Coxhead, A. Dang, T.N.Y., & Mukai, S. (2017)'University tutorials and laboratories:

   Corpora, textbooks and vocabulary',English for Academic Purposes, 30, 66–78

Coxhead, A., Hunter, J., Pierard, J. and Cooke, D. (2008)'Cultural context and academic

   context. What difference do they make in EAP students?' in Burns,A. & J. Burton

   (eds.), Language teacher research in Australia and New Zealand. Alexandria,

   Virginia: TESOL, pp. 25-41.

Dang, T. N. Y., Coxhead, A. & S. Webb. (2017)'The Academic Spoken Word

   List',Language Learning, 67(3), 959–997.

Dang, T. N. Y. & Webb, S. (2014)'The lexical profile of academic spoken English',English

   for Specific Purposes33, 66–76.

Davies, M. (2008)The Corpus of Contemporary American English: 520 million words, 1990-present. [Online]. Available at: http://corpus.byu.edu/coca/ (Accessed: 15 March, 2018).

Gardner, D. & M. Davies. (2014)'A new academic vocabulary list', Applied Linguistics 35(3), 305–327.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002)Range: A program for the analysis of vocabulary in texts. [Online]. Available at: http://www.vuw.ac.nz/lals/staff/paul-nation/ nation.aspx(Accessed: 15 March, 2018).

Liu, D. (2012)'The most frequently-used multi-word constructions in academic written English: A multi-corpus study',*English for Specific Purposes*, 31, 25–35.

Hunter, J. & Coxhead, A. (2007)'New technologies in university lectures and tutorials: Opportunities and challenges for EAP Programmes',TESOLANZ, 15, 30-41.

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

Nation, I. S. P. (2012)The BNC/COCA word family lists. [Online]. Available at: http://www.victoria.ac.nz/lals/about/staff/paul-nation (Accessed: 15 March, 2018).

Nation, I. S. P. (2016)Making and using word lists for language learning and testing. Amsterdam: John Benjamins.

Schmitt, N., Cobb, T., Horst, M. & Schmitt, D. (2015)'How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007)',Language Teaching, 50(2), 212-226.

Simpson-Vlach, R. & Ellis, N. C. (2010)'An Academic Formulas List: New methods in phraseology research',Applied Linguistics, 31(4), 487–512.

West, M. (1953)A general service list of English words. London: Longman, Green.

Wood, D. C. & Appel, R. (2014)'Multiword constructions in first year business and engineering university textbooks and EAP textbooks',Journal of English for Academic Purposes, 15, 1–13.

APPENDIX 1. Coxhead, Dang, and Mukai's (2017) multiwords in labs appearing in Biber, Conrad, and Cortes's (2004) and Simpson-Vlach and Ellis's (2010) lists

Function	Coxhead et al's list of labs MWU	Freq per million	In Biber et al.'s list	In Simpson-Vlach & Ellis's list
STANCE EXPRESSIONS				
Epistemic stance	(do) you know what (I/the)	189.23	1	1
	know what I (mean)	36.39	1	1
	(we/you) don t know (how /I/if/is/it/what/why)	655.03	1	0
	(I) know what the (energy/velocity)	116.45	0	1
	(as/and) you can see (the/it)	167.40	0	1
Attitudinal/ modality stances	(so/what/because/is/and/now/that/all/do) i want to (change/have/select/find/reuse/use/get/count/put/do)	1033.49	1	1
	(do/what) you want to (have/use/do)	291.12	1	1
	(that/so/what) you need to (know/be/have)	232.90	1	1
	(I) want you to (do)	189.23	1	1
	(and/so) what I want (you)	109.17	1	1
	you might want to	58.22	1	1
	(to/should/ 'll) be able to (do)	225.62	1	1

(and/do) you have to (go/be/decide)	167.40	1	0
(what/so) you have to (do)	116.45	1	0
don t have (to)	131.01	1	0
don t want (to)	80.06	1	0
want to do is	116.45	1	0
re going to (build/do/get/have/use)	240.18	1	0
(we/you/they) re going to	473.07	1	0
(is/are) going to be (the)	232.90	1	0
going to have (to)	29.11	1	0
It's going to	43.67	1	0
do we have to	29.11	0	1
have to do (is/it/that)	101.89	0	1
(do/so/that) we need to (see/do)	174.67	0	1
(if) I wanted to	50.95	0	1
(to) make sure that (you)	109.17	0	1
you (re /were) trying to	65.50	0	1
we are trying to	36.39	0	1
DISCOURSE ORGANIZERS			
Topic introduction focus	I think it (s/was)	276.57	1	1
	(if) you look at (the)	101.89	1	1
	(going) to look at (it/the)	116.45	1	1
	what do you (think)	58.22	1	0

	(you) look at it (and)	72.78	1	0
	You're looking at	36.39	0	1
	we (were/are) talking about	58.22	0	1
	you're talking about	29.11	0	1
	are you talking about	36.39	0	1
	I was talking about	29.11	0	1
	We've talked about	43.67	0	1
	(okay) so if you (re)	80.06	0	1
Topic elaboration	to do with (the)	36.39	1	1
	(what) you can do (is/it)	160.12	0	1
	I'm doing is	36.39	0	1
Textual reference	(to) go back and	29.11	0	0
	going to go back	29.11	0	0
	and then there s	36.39	0	0
	and then we'll	29.11	0	0
	and then you can	29.11	0	1
REFRENTIAL EXPRESSIONS				
Identification focus	(so/and/if) this is the	116.45	1	1
	(okay) and this is (what)	58.22	1	1
	it's one of	29.11	1	0
	(is) one of the (things)	80.06	1	0

	(for) those of you	29.11	1	0
Imprecision	or something like that	65.50	1	0
	(it) a matter of	36.39	0	0
Specification of attributes	(and) a lot of (the/this)	101.89	1	0
	a bit of (a)	43.67	1	0
	a little bit (and)	29.11	1	1
	(it/that) s kind of	109.17	0	1
Time/Place/Text reference	the end of (this)	36.39	1	1
Vagueness markers	and so on (so)	43.67	0	1
SPECIAL CONVERSATIONAL FUNCTIONS				
	(if/do) you have a	72.78	1	1
	doesn't have (a)	50.95	1	0
	it doesn't (work/matter/have)	174.67	0	1
	how do you (get)	29.11	0	1
	it looks like a	58.22	0	1
	It's looking up	29.11	0	1

Note: 1= appear; 0= does not appear

APPENDIX 2. Coxhead, Dang, and Mukai's (2017) multiwords in tutorials appearing in Biber, Conrad, and Cortes's (2004) and Simpson-Vlach and Ellis's (2010) lists

Function	Coxhead et al's list of labs MWU	Freq per million	In Biber et al.'s list	In Simpson-Vlach & Ellis's list
STANCE EXPRESSIONS				
Epistemic stance	(do) you know what (I/you)	163.12	1	1
	(I/we/you) don t know (how/what/I/if)	526.21	1	0
	you know if you	55.25	1	0
	you know I (m/mean)	65.78	1	0
	you know when you	36.83	1	0
	you know one of	26.31	1	0
Attitudinal/ modality stances	(what) I want to (do)	65.78	1	1
	(if/do/I) you want to (do/be)	292.05	1	1
	(that/so) you need to (be/do)	155.23	1	1
	(to/should) be able to (do)	221.01	1	1
	(and/that/do/so) you have to (be/do/have)	318.36	1	0
	don t want (to)	76.30	1	0
	(we/you/it) re going to	276.26	1	0

	are (we/you) going to	71.04	1	0
	we have to do	28.94	0	1
	have to do it	26.31	0	1
	we need to (do)	36.83	0	1
	i just want to	26.31	0	1
	don t need (to)	36.83	0	1
	(to) make sure that (you)	92.09	0	1
	(you/we/I) re trying to	184.17	0	1
DISCOURSE ORGANIZERS				
Topic introduction focus	(if) you look at (the)	157.86	1	1
	you re looking at	26.31	1	1
	(what) do you think (that/it/you)	223.64	1	0
	have a look at	71.04	1	0
	(you/we) re talking about	118.40	0	1
	(we/you) were talking about	65.78	0	1
	(okay) so if you	52.62	0	1
Topic elaboration	(nothing/has) to do with (the)	113.13	1	1
	I mean you (know)	55.25	1	0
	you can do (it)	42.10	0	1

	know what I mean	63.14	0	1
	I mean if you	26.31	0	1
	what you (re/were) saying	71.04	0	1
	what I'm (saying)	63.14	0	1
REFRENTIAL EXPRESSION				
Identificatio n focus	(so) this is the	26.31	1	1
	one of the (things)	160.49	1	0
	the things that (we/you)	86.82	1	0
	(some) of the things (you/that)	152.60	1	0
Imprecision	(or) something like that	47.36	1	0
	and things like that	47.36	1	0
	that kind of (stuff)	26.31	0	0
	and stuff like that	28.94	1	0
Specification of attributes	a little bit (of/about/more)	102.61	1	1
	(quite/have) a lot of (the/people/them/it)	226.27	1	0
	a bit of (a)	99.98	1	0
	(all) the rest of (it/the)	102.61	1	0
	(it) s kind of	34.20	0	1
Time/Place/ Text	(at) the end of (the)	186.80	1	1

reference				
	end of the (day)	26.31	1	1
	at the same time	44.73	1	0
SPECIAL CONVERSATIONAL FUNCTIONS				
	if you have a	71.04	1	1
	(so/and/if) you've got (the/a)	207.85	1	0
	thank you very much	128.92	0	1
	it doesn't (matter)	31.57	0	1