Review of Academic Word Lists

May 2020 – Volume 24, Number 1

Dana Therova
The Open University
<dana.therova@open.ac.uk>

Abstract

Academic vocabulary is indisputably one of the most important features of academic texts and recent decades have seen the creation of several lists of academic vocabulary intended to provide a pedagogically useful compilation of the most prevalent words used across as well as within various academic disciplines. These academic word lists have assisted not only the teaching and learning of academic vocabulary, but they have also made a contribution to academic vocabulary research. However, the pedagogical and research value of such lists is sometimes questioned. When utilising these lists, it is therefore important to understand the underlying principles and methodological decisions which had informed their creation. Thus, the purpose of this paper is to offer an up-to-date overview of existing compilations of academic vocabulary, which will hopefully assist EAP researchers and practitioners in making informed decisions regarding the choice and utility of these word lists in their research and pedagogical practices.

Keywords*: academic vocabulary, academic word lists, EAP*

Nation (2001) makes a distinction between four kinds of vocabulary: high-frequency words, technical words, low-frequency words, and academic words. High-frequency words cover a large proportion of texts and "occur in all kinds of uses of the language" (p. 13); technical words relate to the topic or subject area of texts; low-frequency words do not occur very frequently and cover small proportions of texts; and academic words are "common to a wide range of academic texts, and not so common in non-academic texts" (p. 189). These academic vocabulary items "organize scientific discourse and build the rhetoric of academic texts" (Paquot, 2010, p. 28) and are thus seen as important for both comprehension as well as production of academic texts (Coxhead & Byrd, 2007, p. 143). The knowledge of these academic vocabulary items is hence regarded as crucial in academic settings as a growing body of evidence (e.g., Gardner & Davies, 2014; Nagy & Townsend, 2012) points to a link between academic vocabulary knowledge and academic achievement.

The learning and teaching of academic vocabulary has been transformed by the creation of academic word lists containing the most frequently occurring vocabulary in academic contexts.

Such academic word lists have been extensively used for pedagogical as well as research purposes and have made invaluable contributions to the field of applied linguistics.

The creation of academic word lists is often guided by one of two main approaches to defining academic vocabulary, reflecting two different views toward the relationship between general high-frequency words and academic vocabulary. One approach regards academic vocabulary as falling outside general high-frequency words, assuming that learners are already familiar with frequently occurring words. The second approach does not assume that learners already know general high-frequency vocabulary items. Both approaches can be seen as useful ways of determining the most frequent vocabulary items found in academic texts, with the first approach enabling learners and teachers to avoid repetition of already known items, and the second approach ensuring that academic word lists avoid the potential limitations of pre-existing lists of general words (Dang, Coxhead, & Webb, 2017, p. 964).

Due to the important role of academic vocabulary in EAP contexts, this paper sets out to provide an up-to-date overview of the various academic word lists available to EAP learners and practitioners. It first provides a brief overview of the various word lists available across as well as within academic disciplines, then it focuses on compilations of academic vocabulary in common use across a range of disciplines.

Academic Word Lists

Academic word lists can be divided into two broad categories: a) word lists providing a compilation of academic vocabulary occurring across academic disciplines (i.e., general academic word lists); b) word lists containing academic vocabulary used within specific academic disciplines (i.e., discipline-specific word lists). Within these two categories exist lists of academic vocabulary occurring both in the written and spoken academic register.

One of the first attempts at a list of the most commonly used academic expressions in writing in a range of disciplines was the University Word List (UWL) created by Xue and Nation (1984), later updated and replaced by Coxhead's (2000) Academic Word List (AWL). Other compilations of academic vocabulary occurring across disciplines include Paquot's (2010) Academic Keyword List (AKL), the New Academic Word List (NAWL) created by Browne, Culligan and Phillips (2013), Gardner and Davies's (2014) New Academic Vocabulary List (AVL), and most recently the Oxford Phrasal Academic Lexicon (OPAL) (McCarthy, 2019).

Several word lists have been created in discipline-specific domains. As Nation (2016) notes, there are two reasons for the creation of such lists: to examine the nature and size of subject-specific vocabulary items and their identification; and to compile a list of subject-specific vocabulary with the aim of preparing learners for study in that domain or to test vocabulary knowledge in a specific area. Among word lists relating to specific academic disciplines are a science-specific word list (Coxhead & Hirsch, 2007), Economics Academic Word List (EAWL) (O'Flynn, 2019), Business Word Lists (BWL) (Hsu, 2011; Konstantakis, 2007), an Applied Linguistics Academic Word List (Khani & Tazik, 2013), a Chemistry Academic Word List (CAWL) (Valipouri & Nassaji, 2013), a Nursing Academic Word List (NAWL) (Yang, 2015), an Environmental Academic Word List (EAWL) (Liu & Han, 2015), a technical words list for finance (Tongpoon-Patanasorn, 2018), a Theological Word List (TWL) (Lessard-Clouston, 2010), several medical word lists (Hsu, 2013; Lei & Liu, 2016; Wang et al., 2008) and a number of engineering word lists (Hsu, 2014; Mudraya, 2006; Todd, 2017; Ward, 2009).

In terms of the academic spoken register, an English Spoken Academic Wordlist (Nesi, 2002) and more recently the Academic Spoken Word List (ASWL) (Dang et al., 2017) have been developed

as well as the Vocabulary for Academic Lecture Listening word list (VALL) (Thompson, 2015), the Hard Science Spoken Word List (HSWL) (Dang, 2018a) and the Soft Science Spoken Word List (SSWL) (Dang, 2018b). In addition to these, academic word lists containing multi-word expressions have also been developed. Among these are the Academic Collocation List (ACL) (Ackermann & Chen, 2013), Academic Formulas List (AFL) (Simpson-Vlach & Ellis, 2010), Phrasal Expression List (PHRASE List) (Martinez & Schmitt, 2012), and the Oxford Phrasal Academic Lexicon (OPAL) (McCarthy, 2019).

From the numerous compilations of academic vocabulary listed above, it becomes apparent that there has been a growing interest in the development of word lists in the last decade. Due to the vast number of various word lists available, however, the next section narrows down its focus by offering a review of word lists providing a compilation of single-word academic vocabulary items frequently occurring in written academic registers across various academic disciplines. The usefulness of such general academic word lists is relevant in EAP contexts catering to learners intending to pursue various disciplines, learners who have not yet identified their target discipline as well as learners who will be studying interdisciplinary subject areas. In addition, general academic word lists can be of great benefit to EAP teachers who do not possess sufficient background knowledge about their learners' specific disciplines (Dang et al., 2017).

General Academic Word Lists

The University Word List (UWL)

One of the first widely used academic word lists was Xue and Nation's (1984) University Word List (UWL). The UWL was created by drawing on two other word lists: Campion and Elley's (1971) list developed for the vocabulary section of the Language Achievement Test of Overseas Students, published by the New Zealand Council for Educational Research; and Praninskas's (1972) American University Word List compiled for learners of English, which excluded words appearing in West's (1953) General Service List of English Words (GSL). The GSL was created by West in 1953 on the basis of a five-million word corpus of written English resulting in a list containing 2,000 most commonly used word families in English (with a word family referring to a stem plus all related inflections and derivations containing that stem without distinguishing between word classes).

Having merged these two lists and using the GSL as the prerequisite for their combined list, Xue and Nation used two other lists to add high-frequency vocabulary items to their combined list, namely Lynn's (1973) and Ghadessy's (1979) word lists, both containing words which learners had found difficult during their reading. This resulted in a list containing 836 word families (consisting of 3,707 word types), divided into ten sub-lists "to give an indication of their relative usefulness, and to provide manageable groups of words for teaching and learning" (Xue & Nation, 1984, pp. 216–217). The 836 word families have a reported coverage of 8.5% of academic texts (Nation, 2001), specifically those contained in the Learned and Scientific sections of the LOB corpus of written British English (Johansson, 1978 in Coxhead, 2000) and the parallel Wellington corpus of written English (Bauer, 1993 in Coxhead, 2000).

The use of four pre-existing word lists for the creation of the UWL has, however, been considered its limitation due to the fact that as a combination of the four pre-existing lists, it lacked consistent selection principles. In addition, the source corpora forming the basis of the four lists were rather small and did not contain a balanced range and a wide variety of topics. Hence, the UWL had many of the weaknesses of the prior work (Coxhead, 2000). The UWL is no longer considered relevant today as it has been replaced by the more recent Academic Word List (Coxhead, 2000).

The Academic Word List (AWL)

The AWL (Coxhead, 2000) is based on a corpus of approximately 3.5 million words from 414 academic texts published by more than 400 authors mainly in New Zealand between the early 1960s and the late 1990s. The source corpus is divided into four sub-corpora representing science, arts, commerce, and law, with each discipline comprising approximately 875,000 words and further sub-divided into 7 more specific subject areas (Coxhead, 2000).

The creation of this list was guided by the following principles: *range* referring to the occurrence of a member of a word family at least 10 times in each of the four sub-corpora and in a minimum of 15 of the 28 subject areas; *frequency* based on the fact that members of a word family had to occur a minimum of 100 times in the corpus; and *specialised occurrence* referring to the exclusion of the 2,000 word families in West's (1953) General Service List of English Words (GSL).

The resulting AWL contains 570 word families organised into 10 sub-lists according to decreasing word family frequency. The AWL word families account for approximately 10% of the total words in the selected academic texts forming the source corpus. Out of the 570 word families contained in the AWL, 172 can be found in all 28 subject areas and 263 are present in at least 27 subject areas. Nonetheless, it is important to note that the AWL does not represent the four disciplines equally. As Coxhead (2000) notes, the AWL's coverage of science (9.1%), arts (9.3%) and law (9.4%) is slightly under-represented in comparison to the coverage of the commerce sub-corpus (12%).

The above characteristics of the AWL and the methodology behind its compilation have been subject to critique. Hyland and Tse (2007), for instance, argue that the items in the list are unlikely to be of equal value to students across disciplines since individual words in the list occur with different frequencies in different disciplines and behave differently across disciplines in terms of range, collocation and meaning. A further flaw has been seen in its foundation on the GSL, which has been criticised for its coverage, age, utility as well as its organising principle around word families (e.g., Brezina & Gablasova, 2015; Engels, 1968; Paquot, 2010; Richards, 1974).

The word family principle is based on the underlying assumption that "once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort" (Bauer & Nation, 1993, p. 253). The use of word families as an organising principle of the GSL has, however, been criticised due to the "semantic distance of the words that can be included under one headword in a word family," meaning that words belonging to the same word family can display different meanings, such as *please* and *unpleasantly*, *part* and *particle,* or *value* and *invaluable*, listed as belonging to the same word family in the GSL (Brezina & Gablasova, 2015, p. 4).

Despite the issue of the different meanings of words belonging to the same word family, the GSL formed the basis for the AWL, whereby the word families contained in the GSL were excluded from the AWL. The rationale behind the use of word families, however, is the argument that this practice is "supported by evidence suggesting that word families are an important unit in the mental lexicon" (Nagy et al. 1989, p. 262 in Coxhead, 2000, p. 218). This is because comprehending derived or inflected members of a word family does not require a great deal of effort on the learners' part provided that they are familiar with the base word and have a certain amount of control of basic word-building processes in English (Bauer & Nation, 1993). Against this argument are Gardner and Davies (2014), whose concern with focusing on word families is that L2 learners' knowledge of derivational word relationships is developed much later than knowledge of inflectional word relationships.

Gardner and Davies (2014) thus propose that lemmas, defined as "words with a common stem, related by inflection only, and coming from the same part of speech" (p. 308), are preferable to word families when creating pedagogical word lists. This would also eliminate the problem of meaning caused by the fact that grammatical parts of speech are not considered word families. Gardner and Davies (2014) provide the example of the verb *proceeds* (meaning *continues*) and the noun *proceeds* (meaning *profits*) which belong to the same word family in the AWL despite the difference in meaning. Another example provided by Gardner and Davies (2014) is the word family *react* with the following members: *react* (headword), *reaction*, *reactive*, *reacted*, *reactions*, *reactivate*, *reacts*, *reactionaries*, *reactivation*, *reacting*, *reactionary*, *reactor*, and *reactors*, which demonstrate the differences in primary meanings between *react* (respond), *reactionary* (strongly opposed to social or political change), *reactivation* (to make something happen again), and *reactor* (a device or apparatus). This is further supported by Durrant (2016, p. 51), who also points out that "word families often bring together forms with very diverse meanings" and illustrates this with the AWL item *constitute* which subsumes *constituting*, *constituent*, and *unconstitutional*. These examples highlight the issue with the meaning stemming from the fact that a word family does not consider grammatical word classes.

Despite these potential limitations, the AWL has served as the basis for identification and exploration of academic vocabulary usage in various contexts (e.g., Choo et al., 2017; Cons, 2012; Hyland & Tse, 2007, 2009; Knoch et al., 2015, 2014; Li & Qian, 2010; Martínez et al., 2009; Masrai & Milton, 2018; Nadarajan, 2011; Neufeld et al., 2011; Olinghouse & Wilson, 2013; Storch & Tapper, 2009; Valipouri & Nassaji, 2013).

The Academic Keyword List (AKL)

Another compilation of academic vocabulary is Paquot's (2010) Academic Keyword List (AKL). The AKL was intended to result in an extraction of potential academic words by employing Rayson's (2008) data-driven approach drawing on corpus-based and corpus-driven approaches in corpus linguistics, following five main steps: 1) corpus design and compilation; 2) corpus annotation; 3) retrieval of relevant features using quantitative and qualitative analyses; 4) devising a research question or model; 5) interpretation of results (Paquot, 2010). As for the source corpus design and compilation, Paquot opted for professional as well as student writing as the basis for the creation of the AKL. The rationale behind this was that students produce texts which are likely to be used for assessment purposes in the context of their study rather than for publication. The total size of the source corpus comprised over three million words from four different corpora: the *Miro-Concord Corpus Collection B* and the *Baby BNC Academic Corpus* representing a variety of professional academic writing covering a number of genres including book sections, scientific journals, and texts written by British scholars and American researchers, and the *Louvain Corpus of Native Speaker Essays* (LOCNESS) and the *British Academic Written Corpus* (BAWE) *Pilot Corpus* forming the *Student Writing Corpus* representing university student writing (Paquot, 2010).

The AKL is primarily based on *keyness* to extract distinctive words, followed by applying the criteria of *range* and *evenness of distribution* measured using Juilland's D value to narrow down the potential academic words (Paquot, 2010). This resulted in a "list of potential academic words [which] has been named the *Academic Keyword List* (AKL) to emphasize the fact that it is the output of a data-driven set of criteria, the first of which is keyness, and not a list of academic vocabulary in its functional sense" (Paquot, 2010, p. 55).

The AKL contains 930 individual words, out of which 38.17% belong to the noun category, which is found to be particularly common in the academic register (Biber & Gray, 2016). The second largest category of the words in the AKL are verbs (accounting for 25% of the words in the AKL),

following Hinkel's (2004) classification of verbs into activity verbs, reporting verbs, mental verbs, linking verbs, and logico-semantic relationship verbs. The next category is adjectives representing 19.35% of the potential academic words in the AKL, highlighting the function that adjectives play in academic discourse in terms of modifying nouns and noun phrases. Another important function in academic discourse is represented by adverbs (accounting for 9.35% of the words in the AKL), with mainly linking and evaluative function. The remaining 8.06% of the potential academic words include prepositions, conjunctions, pronouns, articles, determiners and ordinal numbers (Paquot, 2010).

Nevertheless, it needs to be acknowledged that the AKL still needs validation as it is not a final product which would guarantee pedagogical relevance (Paquot, 2010). Thus, each AKL vocabulary item ought to be "subject to a careful corpus-based analysis to confirm its status as an academic word and establish how it is used in academic prose in terms of meaning, phraseology, and sentence position" (Paquot, 2010, p. 63). Other limitations of the AKL could be seen in the fact that the Student Writing Corpus was skewed towards humanities and social sciences; hence, the AKL might not fully represent academic vocabulary used in the 'hard sciences' (Paquot, 2010). Next, 57% of the words in the AKL are found in the GSL, suggesting that they are not unique to academic contexts. Paquot (2010) argues, however, that the rather high occurrence of the AKL vocabulary items in the GSL highlights the important role that general vocabulary plays in academic registers. Other potential weaknesses of the AKL, as acknowledged by Paquot (2010), relate to the relatively small size of the source corpora and a limited number of texts used for the compilation of the AKL. A further limitation, inherent in the keyness approach, concerns the reference corpus used for comparative purposes with the corpus of interest. This is because the selected reference corpus may have contained a number of potential academic words, resulting in an omission of these academic words from the AKL (Paquot, 2010).

Despite these potential limitations, the AKL vocabulary items have been shown to belong to semantic categories relating to academic discourse and can hence be seen as a valuable contribution to academic vocabulary research and pedagogy.

The New Academic Word List (NAWL)

As a response to Coxhead's AWL, Browne et al. (2013a) developed the New Academic Word List (NAWL) following the same principle as Coxhead (2000) in the creation of the AWL; that is, excluding general high-frequency vocabulary. To exclude general vocabulary, the creation of the NAWL was preceded by the development of the New General Service List (NGSL) (Browne et al., 2013b), which is based on a 273 million-word sub-section of the 2 billion-word Cambridge English Corpus (CEC) and contains approximately 2,800 high-frequency words. The NGSL is based on a modified lexeme approach. Unlike the traditional definition of lexeme, which groups related forms with the same meaning and part of speech, the modified lexeme approach counts the headword in all its various parts of speech and includes all inflected forms (Browne, 2014).

The NAWL is based on an academic corpus containing 288 million words comprising the following three components: CEC Academic Corpus containing 248 million words forming 86.3% of the source corpus; Textbooks totalling 36 million words forming 12.6% of the source corpus; and Oral corpus containing 3 million words and forming 1.1% of the source corpus. The following information is available about the source corpus composition: the CEC Academic corpus contained academic journals, non-fiction, student essays and academic discourse; the Textbooks corpus was based on hundreds of top-selling academic textbooks; the Oral corpus was formed by the Michigan Corpus of Academic Spoken English (MICASE) and the British Academic Spoken Corpus (BASE).

The NAWL contains 960 academic words, which together with the NGSL cover 92% of the 283-million words Academic Corpus used for its compilation, whereby the NGSL is reported to cover 86% of the Academic Corpus. In other words, the NAWL adds 6% more words to the words covered by the NGSL. For convenience, the NAWL is available in a number of formats, such as in alphabetical order, with inflected forms, with standard frequency indexes and with definitions in English and Japanese.

Some of the advantages of this list could be seen in the size of the source corpus forming the basis of the NAWL, the fact that it is more recent than the AWL, and the various formats of the lists. However, no information regarding the methodological procedures involved in the creation of this word list or information regarding the organising principle of the NAWL is available. Nevertheless, considering that this list was developed to work in conjunction with the NGLS, it is reasonable to assume that the NAWL is based on the same unit of counting as the NGSL; that is, a modified lexeme principle. The basis of the NAWL on this organising principle could, however, be seen as a potential weakness of the list due to the modified lexeme approach collapsing various grammatical classes, resulting in grouping words with different meanings. This can be illustrated by the following NAWL entries: *articulate* (having a different meaning as an adjective and verb) or *integral* (displaying a different meaning as a noun and adjective), for instance.

The New Academic Vocabulary List (AVL)

A further response to Coxhead's (2000) AWL, particularly the issue of using word families as an organising principle of the AWL (discussed above), was Gardner and Davies's (2014) new Academic Vocabulary List (AVL), which was created using lemmas. The aim of using lemmas for the creation of the AVL was to achieve a more accurate assessment of word forms, functions as well as meanings (Gardner & Davies, 2014). The lemmas contained in the AVL were identified on the basis of a corpus comprising 120 million words of academic texts (that is, a corpus nearly 35 times larger than the corpus used for the compilation of the AWL) taken from the 425-million-word Corpus of Contemporary American English (COCA) composed of the following nine academic disciplines: education; humanities; history; social sciences; philosophy, religion, psychology; law and political science; science and technology; medicine and health; business and finance (Gardner & Davies, 2014). Hence, in terms of size, range and representativeness, the AVL source corpus can be seen as an advance on the AWL.

There were five key considerations behind the compilation of the AVL: the use of lemmas instead of word families; the use of a representative corpus of academic texts covering various academic disciplines; a statistical derivation of the new list from a large and balanced corpus of academic and non-academic materials; the use of contemporary academic and non-academic materials for comparison purposes; and the validity and reliability of the new list being tested against both academic and non-academic corpora or corpus-derived lists (Gardner & Davies, 2014).

The creation of the AVL was then guided by these four principles: *ratio*, *range*, *dispersion*, and *discipline measure*. The ratio criterion was used to eliminate general high-frequency words from the list by specifying that the frequency of the lemma must be at least 50% higher in the 120-million-word academic corpus than in the 305-million-word non-academic section of COCA. According to the range principle, the lemma had to appear in a minimum of seven of the nine academic disciplines with at least 20% of the expected frequency. In terms of the dispersion criterion (using the Juilland's D measure), the lemma had to have a dispersion of at least 0.80 (with a dispersion value ranging from 0.01 indicating that the occurrence of a word is in an extremely small part of the corpus to 1.00 representing perfectly even dispersion in all parts of the corpus). The discipline measure was intended to exclude discipline-specific and technical words, whereby

words occurring more than three times the expected frequency in any of the nine disciplines were not considered. In sum, the ratio criterion had the function of eliminating general high-frequency words, while the remaining three criteria were applied to exclude discipline-specific and technical words (Gardner & Davies, 2014).

Despite these carefully selected criteria, it cannot be argued that the AVL is without limitations. Durrant (2016), for instance, points out that about half of the words in the list are used very little. However, he also found that 427 out of the total of 3015 lemmas forming the list are frequent across 90% of disciplines, supporting the notion of generic academic vocabulary, though at a considerably smaller scale than the entire AVL. The strengths of the AVL can be seen in the size of the source corpus, the use of contemporary materials, the use of lemmas eliminating the issue of meanings caused by derivational relationships, and the thorough methodological procedures employed in its creation.

Oxford Phrasal Academic Lexicon (OPAL) Written Words

The Oxford Phrasal Academic Lexicon (OPAL) is, to my knowledge, the most recent compilation of academic expressions available to EAP teachers and learners. It comprises a collection of four different word lists: written words, spoken words, written phrases, and spoken phrases. These four lists aim to "provide an essential guide to the most important words to know in the field of English for Academic Purposes (EAP)" ('Oxford Learner's Dictionaries', 2019). The written words list is based on the 71-million-word Oxford Corpus of Academic English (OCAE) containing academic texts published by Oxford University Press in four disciplines: arts and humanities; social sciences; life sciences; and physical sciences. The creation of the OPAL is a result of a keyword analysis, which was intended to identify the most frequent words in an academic setting by extracting words that are typical of the OCAE corpus. This is claimed to make the lists "more than just a checklist of words that learners need to know" but it also ought to assist students in using the most appropriate language in their academic language production ('Oxford Learner's Dictionaries,' 2019). Similar to the AKL and AVL, the OPAL is not based on a pre-existing list of general academic words. However, no other information regarding the methodological procedures involved in the compilation of this word list is available, such as the selection of the reference corpus typically involved in a keyword analysis approach. Nor does the OPAL report on the coverage of the various academic disciplines represented by the source corpus.

The OPAL written word list contains 1,200 most frequent words in academic writing divided into twelve sub-lists according to decreasing frequency with the most frequent words in sub-list 1, the next most frequent in sub-list 2 and so on. Each sub-list contains 100 individual words (as opposed to word families or lemmas) and includes information about the words' part of speech.

The advantages of the OPAL word list could be seen in its division into sub-lists, which means that learners and teachers can choose the sub-list that is most appropriate for the learners' level. Specifically, beginners should start with sub-list 1 to start building their vocabulary. More advanced learners, who have already mastered the most frequent academic words, can utilise sub-lists 11 and 12 to extend their academic vocabulary knowledge, for example. The size of the sub-lists can be regarded a further advantage of the list as it makes it manageable to work with the list over a shorter period of time, reflecting the needs of individual English language courses. The OPAL is also conveniently linked to The Oxford Learner's Dictionary of Academic English, which is a learner's dictionary focusing exclusively on academic English, giving both teachers and students access to a powerful dictionary database. This enables the learners to select a word from the OPAL list, which will take them to its dictionary entry providing information about the selected lexical item's meaning, use and example sentences. This makes the OPAL a convenient list for students to also

use independently with the dictionary enabling them to further explore the words' meanings and typical use in context.

One of the potential weaknesses of the OPAL written word list relates to its collapsing of individual words with different parts of speech, resulting in individual entries carrying different meanings. This become apparent from the following entries: *current adj., .n.* (displaying a different meaning as adjective and as a noun), *relative adj., n.* (displaying a different meaning as an adjective and as a noun), *potential adj., n.* (displaying a different meaning as an adjective and as a noun), or *objective n., adj.* (displaying a different meaning as a noun and as an adjective). This makes it necessary for learners to use the word list in conjunction with the Dictionary to become aware of how the meanings of some of the words in the list change depending on their word class.

General Academic Word Lists: Summary

From the above review of general academic word lists, despite their common aim of providing a compilation of general academic vocabulary, significant differences between them become apparent including their age, size, source corpus and organising principles. The various methodological decisions reflecting one of the two approaches to academic vocabulary also become apparent; that is: a) regarding academic vocabulary as being excluded from general high-frequency words based on the assumption that learners have already become familiar with these lexical items, adopted for the creation of the UWL (Xue & Nation, 1984), the AWL (Coxhead, 2000) and the NAWL (Browne et al., 2013a); and b) not assuming that learners are already familiar with general high-frequency vocabulary, reflected in the AKL (Paquot, 2010), the AVL (Gardner & Davies, 2014) and the OPAL (McCarthy, 2019). A summary of the main features of the discussed lists is provided in Table 1.

Table 1. Overview of General Academic Word Lists

Characteristic	UWL	AWL	NAWL
Date of creation	1984	2000	2013
Basis of creation	4 existing word lists; GSL	GSL	NGSL
Corpus size	N/A	3,513,330 words	288,000,000 words
Source texts	Existing word lists published in the 1970s: Campion and Elley's word list (1971) Praninskas's American University Word List (1972) Lynn's word list (1973) Ghadessy's word list (1979)	414 texts published between the 1960s-1990s 64% of the texts sourced in New Zealand 20% sourced in Britain 13% sourced in the USA 2% sourced in Canada 1% sourced in Australia	Academic Corpus of the Cambridge English Corpus: 248 million words (86.3% of the source corpus) Textbooks: 36 million words (12.6% of the source corpus) Oral corpus: MICASE and BASE: 3 million words (1.1% of the source corpus)
Disciplines	N/A	4 disciplines: science; arts; commerce; law; each sub-divided into 7 subject areas	Not specified
Methodology	Combination of existing word lists Exclusion of GSL items	Specialised occurrence (i.e., exclusion of GSL items); Range; Frequency	Not specified
Organisation	Word families, 10 sub-lists	Word families, 10 sub-lists	In alphabetical order, with inflected forms, with standard frequency indexes, with

			definitions in English and Japanese
Size	836 word families (3,707 word types)	570 word families (3,110 word types)	960 words
Reported coverage	8.5% of the Learned and Scientific sections of the LOB corpus of written British English and the parallel Wellington corpus of written English; 9.8% of the AWL source corpus	10% of academic texts in the source corpus 51% of UWL items (i.e., 435 overlapping word families)	NGSL and NAWL combined cover 92% of the source corpus (86% covered by NGSL, 6% covered by NAWL) compared to 87% coverage of the source corpus by GSL and AWL combined

Characteristic	AVL	AKL	OPAL single written words
Date of creation	2014	2010	2019
Basis of creation	–	–	–
Corpus size	>120,000,000 words	3,040,004 words	71,000,000 words
Source texts	13,000 contemporary academic texts from the 425-million-word COCA with texts as recent as 2011 published in the USA 85 million words from academic journals 31.5 million words from academically oriented magazines 7.5 million words from the finance sections of newspapers	Professional writing corpora: Micro-Concord Corpus Collection B & Baby BNC Academic Corpus (both containing mainly British English) Student writing corpora: Louvain Corpus of Native Speaker Essays (LOCNESS) containing American and British English & British Academic Written Corpus (BAWE) Pilot Corpus containing British English	Academic texts published by Oxford University Press
Disciplines	9 disciplines: education; humanities; history; social sciences; philosophy, religion, psychology; law and political science; science and technology; medicine and health; business and finance	Professional writing corpus of soft science: arts; belief and religion; social sciences; humanities; politics, education and law Professional writing corpus of hard science: science,applied science, technology and engineering Student writing corpus (BAWE): language studies; social sciences; psychology; history	4 disciplines: arts and humanities; social sciences; life sciences; physical sciences
Methodology	Ratio; Range; Dispersion; Discipline measure	Keyness; Range; Evenness of distribution	Keyword analysis
Organisation	Lemmas	Individual words	12 sub-lists
Size	3,015 lemmas (~2,000 word families)	930 words	1,200 words
Reported coverage	13.8% of academic texts in the source corpus	57% of GSL items 40% of AWL items	Not reported

Conclusion

This paper set out to provide an up-to-date overview of single-word general academic word lists used in the written academic register. It has described the word lists currently available and highlighted the differences between them in terms of their age, size, source corpora, organising principles and methodological principles which guided their compilation. The awareness of the different academic word lists available as well as their characteristics is intended to inform EAP practitioners' decisions with regards to the choice and utility of these lists in their practices.

It ought to be noted, however, that despite the usefulness of general academic word lists in applied linguistics research and pedagogy, the generic label of academic vocabulary is a concern expressed by some. Hyland and Tse (2007, p. 251), for instance, note that even though it "may be a convenient shorthand for describing a general variety, it conceals a wealth of discursive variability which can misinterpret academic literacy as a uniform practice and mislead learners into believing that there is a single collection of words which they can learn and transfer across fields." It is, therefore, important to emphasise that "[w]ordlists are based on the premise that some words are likely to be more useful to learners than others" (Durrant, 2016, p. 50). Hence, one of the main criticisms of general academic word lists remains the difficulty of meeting the needs of students from various disciplines as considerable differences between the vocabulary used in different academic contexts and disciplines mean that no single generic word list is likely to be adequate (Durrant, 2016).

A further limitation points to the reliance of these word lists on single words, which fails to reflect the fact that words tend to co-occur with other items in preferred combinations which can sometimes cause them to take on additional meanings. These combinations are important in academic contexts as they contribute to meaning-making. Thus "[b]y breaking into single words items which may be better learnt as wholes, vocabulary lists simultaneously misrepresent discipline-specific meanings and mislead students" (Hyland & Tse, 2007, p. 247). Nonetheless, despite these potential limitations, academic word lists have made an invaluable contribution to vocabulary research as well as pedagogy in academic contexts.

About the Author

Dana Therova is an EAP practitioner currently pursuing a PhD at The Open University, where she is investigating the deployment and development of academic vocabulary in multilingual foundation-level students' assessed academic writing in the context of UK Higher Education.

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. https://doi.org/10.1016/j.jeap.2013.08.002

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Browne, C., Culligan, B., & Phillips, J. (2013a). New academic word list. Retrieved from http://www.newgeneralservicelist.org/nawl-new-academic-word-list

Browne, C., Culligan, B., & Phillips, J. (2013b). New general service list. Retrieved from http://www.newgeneralservicelist.org/

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

Choo, L. B., Lin, D. T. A., Singh, M. K. M., & Ganapathy, M. (2017). The significance of the Academic Word List among ESL tertiary students in a Malaysian public university. *3L: Language, Linguistics, Literature*, *23*(4), 56–65. https://doi.org/10.17576/3L-2017-2304-05

Cons, A. M. (2012). The use and misuse of academic words in writing: Analyzing the writing of secondary English learners and redesignated learners. *TESOL Journal*, *3*(4), 610–638. https://doi.org/10.1002/tesj.36

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

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, *16*(3), 129–147. https://doi.org/10.1016/j.jslw.2007.07.002

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *XII*, 65–78.

Dang, T. N. Y. (2018a). A hard science spoken word list. *International Journal of Applied Linguistics*, *169*(1), 44–71.

Dang, T. N. Y. (2018b). The nature of vocabulary in academic speech of hard and soft-sciences. *English for Specific Purposes*, *51*, 69–83. https://doi.org/10.1016/j.esp.2018.03.004

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997. https://doi.org/10.1111/lang.12253

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, *43*, 49–61. https://doi.org/10.1016/j.esp.2016.01.004

Engels, L. K. (1968). The fallacy of word-counts. *International Review of Applied Linguistics in Language Teaching*, *6*(3), 213–231.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Ghadessy, M. (1979). Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum*, *17*, 24–27.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Oxon: Routledge.

Hsu, W. (2011). A business word list for prospective EFL business postgraduates. *The Asian ESP Journal*, *7*(4), 63–99.

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454–484. https://doi.org/10.1177/1362168813494121

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, *33*, 54–65. https://doi.org/10.1016/j.esp.2013.07.001

Hyland, K., & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL Quarterly*, *41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies*, *9*(2), 111–129.

Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC Journal*, *44*(2), 209–232. https://doi.org/10.1177/0033688213488432

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, *28*, 39–52. https://doi.org/10.1016/j.jslw.2015.02.005

Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English-medium university? *Assessing Writing*, *21*, 1–17. https://doi.org/10.1016/j.asw.2014.01.001

Konstantakis, N. (2007). Creating a business word list for teaching business English. *Estudios de Linguistica Inglesa Aplicada (ELIA)*, *7*, 79–102.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42–53. https://doi.org/10.1016/j.jeap.2016.01.008

Lessard-Clouston, M. (2010). Theology lectures as lexical environments: A case study of technical vocabulary use. *Journal of English for Academic Purposes*, *9*(4), 308–321. https://doi.org/10.1016/j.jeap.2010.09.001

Li, Y., & Qian, D. D. (2010). Profiling the academic word list (AWL) in a financial corpus. *System*, *38*, 402–411. https://doi.org/10.1016/j.system.2010.06.015

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, *39*, 1–11. https://doi.org/10.1016/j.esp.2015.03.001

Lynn, R. W. (1973). Preparing word lists: a suggested method. *RELC Journal*, *4*(1), 25–28.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, *28*, 183–198. https://doi.org/10.1016/j.esp.2009.04.003

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320. https://doi.org/10.1093/applin/ams010

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, *31*, 44–57. https://doi.org/10.1016/j.jeap.2017.12.006

McCarthy, M. (2019). *Why do we need EAP word lists and what can we do with them? BALEAP Conference: Innovation, Exploration and Transformation, University of Leeds 12th – 14th April 2019*. Leeds.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*, 235–256. https://doi.org/10.1016/j.esp.2005.05.002

Nadarajan, S. (2011). The challenges of getting L2 learners to use academic words in their writings. *Electronic Journal of Foreign Language Teaching*, *8*(2), 184–200.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*(1), 91–108. https://doi.org/10.1002/RRQ.011

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

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

Nesi, H. (2002). An English spoken academic wordlist. In *Proceedings of the tenth EURALEX International Congress* (p. 351–357, Vol.1). Copenhagen, Denmark.

Neufeld, S., Hancioğlu, N., & Eldridge, J. (2011). Beware the range in RANGE, and the academic in AWL. *System*, *39*(4), 533–538. https://doi.org/10.1016/j.system.2011.10.010

O'Flynn, J. (2019). An economics academic word list (EAWL): Using online resources to develop a subject-specific word list and associated teaching-learning materials. *Journal of Academic Language and Learning*, *13*(1), 28–87.

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, *26*, 45–65. https://doi.org/10.1007/s11145-012-9392-5

Oxford Learner's Dictionaries. (2019). Retrieved from https://www.oxfordlearnersdictionaries.com/about/wordlists/opal

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London: Continuum International Publishing Group.

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

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519–549.

Richards, J. C. (1974). Word Lists: Problems and prospects. *RELC*, *2*, 69–84.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, *8*, 207–223. https://doi.org/10.1016/j.jeap.2009.03.001

Thompson, P. (2015). Changing the bases for academic word lists. In P. Thompson & G. Diani (Eds.), *English for academic purposes: Approaches and implications* (pp. 317–342). Newcastle upon Tyne: Cambridge Scholars Publishing.

Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31–39. https://doi.org/10.1016/j.esp.2016.08.003

Tongpoon-Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes*, *51*, 45–54. https://doi.org/10.1016/j.esp.2018.03.002

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, *12*(4), 248–263. https://doi.org/10.1016/j.jeap.2013.07.001

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*, 442–458. https://doi.org/10.1016/j.esp.2008.05.003

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*, 170–182. https://doi.org/10.1016/j.esp.2009.04.001

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

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, *3*(2), 215–229.

Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, *37*, 27–38. https://doi.org/10.1016/j.esp.2014.05.003