Providing ESP Learners with the Vocabulary They Need: Corpora and the Creation of Specialized Word Lists

Simon FRASER
Institute for Foreign Language Research and Education
Hiroshima University

Introduction

EAP/ESP (English for Academic/Specific Purposes) has enjoyed a rising profile in recent years. Here in Japan, this has resulted in an increasing awareness among EFL practitioners of the necessity to provide graduate-level learners with the English they will need for their educational studies and in professional contexts. Providing learners with the language they require for specific purposes to a large extent means providing them with the *words* they need, although this of course does not guarantee performance in the language. Two obvious questions which come to mind are:

1. Exactly what vocabulary does the ESP learner need?
2. Once identified, how should such vocabulary be taught and learned?

In this paper, I would like to focus on the first of these questions, and demonstrate how word lists created using frequency and range criteria can be helpful in answering it. I will also attempt to show how the findings can throw some light on the way in which the second question might best be addressed.

Frequency lists in EFL

Frequency information is important to ensure that the words studied by learners will in fact be the ones they encounter most often. One of the first frequency lists was the General Service List (West, 1953). Often referred to as the "classic" frequency list, the GSL is still considered to be useful despite its age, the errors it contains, and its focus on written language. The 2,000 most frequent word families in this list are considered to be an essential basis for all language use. The Brown Corpus (Francis and Kucera, 1982) features lemmatized word lists ranked according to frequency (1,000 up to 6,000 words). More recent word lists, which have attempted to redress the written/spoken balance, include the JACET 8000 (2003) list and Nation's (2006) fourteen 1,000 word family lists, derived from the *British National Corpus* (http://info.ov.ac.uk/bnc/).

In EAP, the University Word List (UWL) (Xue and Nation, 1984) was developed to provide learners with a general academic vocabulary. This list of just over 800 words has now been superseded by the 570 word family Academic Word List (Coxhead, 2000). The AWL is based on a 3,500,000 token corpus of academic English which is divided into four

groupings of Arts, Science, Law, and Commerce. The purpose of the UWL or AWL is to provide learners with words that are not in the first 2,000 words of the GSL, but are frequent across a wide range of academic disciplines.

For learners with more specific requirements, attempts have been made to create lists of the specialized vocabulary of particular disciplines. Baker (1988), comparing a specialized corpus with a general corpus in her study of medical journal articles, was one of the first researchers to produce a list of specialized lexis using frequency and distribution criteria. Other studies have produced word lists for the English of electronics (Farrell, 1990), medicine (Salager, 1983), economics (Sutarsyah, Nation and Kennedy, 1994), and engineering (Ward, 1999; Mudraya, 2006). Chujo and Utiyama (2006) used statistical measures to produce rank-ordered lists of business English words targeted to different proficiency levels.

Cobb and Horst (2001) have suggested that if the most frequent words and the academic words are extracted from a corpus of domain texts, a residue of terms will be left which characterize the domain. This was the methodology followed by Konstantanis (2006) in his attempt to create a business word list. Konstantanis wanted a list which, in conjunction with the GSL and AWL, would take coverage of a business English corpus up to at least 95% of the total number of words occurring in the corpus. It has been suggested (Laufer, 1989a, 1992; Hirsh and Nation, 1992) that a lexical coverage of 95% in a text is the threshold for learners to achieve adequate reading comprehension; this equates to a knowledge of around 3,000 word families. For pleasurable reading, it may be that learners need to know as many as 98% of the running words (Hu and Nation, 2000).

How easy or difficult might it be to reach this kind of coverage with a list that goes beyond the 2,000 high-frequency words and the academic words? According to Nation and Waring (1997), the GSL gives an average 82% coverage in written texts; in academic texts, the coverage by the most frequent words can be expected to be slightly less, with the AWL providing an additional coverage of 10% or so. Coxhead (2000) found that the GSL and the AWL gave the following coverage of her academic corpus (Table 1):

Table 1. Coverage of Coxhead's academic corpus by the first three word lists

Word list	Coverage of corpus
Most frequent 1,000 words	71.4%
2nd 1,000 most frequent words	4.7%
Academic Word List	10.0%
TOTAL	86.1%

Similar results were obtained by Nation (2004), who found that the GSL and AWL gave an average of just under 90% coverage on four different corpora. It would seem, then, that any specialized word list needs to give coverage of 5–10% if the 95% mark is to be reached.

Konstantanis' study used the *Published Material Corpus* (Nelson, 2000). The PMC

consists of about 600,000 words, texts coming from 33 business English course books. He found that a list of business English words would need to give additional coverage of around 5% (Table 2).

Table 2. Coverage of the Published Material Corpus by the first three word lists

Word list	Coverage of corpus
Most frequent 1,000 words	80.26%
2nd 1,000 most frequent words	5.46%
Academic Word List	4.66%
TOTAL	90.38%

Using the criteria of range (appearing in more than five books) and frequency (appearing more than 10 times), Konstantanis arrived at a list consisting of 561 word families. Although this list, the Business Word List (BWL), gave only an additional 2.79%, it increased total coverage of the PMC to 93.17%, pushing the total fairly close to the magic 95% (see Table 3).

Table 3. Coverage of the business PMC by the first three lists and the BWL

Word list	Coverage of corpus
Most frequent 1,000 words	80.26%
2nd 1,000 most frequent words	5.46%
Academic Word List	4.66%
Business Word List	2.79%
TOTAL	93.17%

For business English, at least, this would seem to be a fairly promising way of identifying the words that learners will need. I was interested to see whether it would be possible to use similar methodology to create a word list for a medical discipline, pharmacology. Medicine differs from business or economics, for example, in that a large proportion of its terms are totally incomprehensible to the layperson, whereas a great many business or economics terms, if not their precise meanings, are familiar to most people. This is evident from Konstantanis' finding that the first 2,000 words provide coverage of almost 86% of a business corpus; for an economics text, Sutarsyah, Nation and Kennedy (1994) found a corresponding figure of 82.5%.

Creating a pharmacology word list

The present study was carried out using a pharmacology corpus currently under construction as part of wider research into the characteristics of specialized language. Pharmacology was chosen for my investigation as I have studied it at the tertiary level, and felt that my specialist knowledge of the subject would be useful. At present, the 185,000-word

corpus consists of 51 articles taken from a wide international selection of pharmacology journals. The journals use both British and American English, and areas represented include cardiovascular pharmacology, autonomic pharmacology, biochemical pharmacology, clinical pharmacology, alimentary pharmacology, and toxicology. The journals are all relatively recent (1997-2006), with most being published in 2006.

The corpus was run against the RANGE computer program (available at http://www.wuw.ac.nz/lals/). In addition to counting the total number of word types and tokens, the program can be used to compare one or more texts against the GSL and AWL lists. At present, the program uses the GSL rather than any more up-to-date list, but Hwang and Nation (1995) and Nation (2004) have shown that newer lists give only slightly better coverage. Coverage provided by the GSL and AWL is expressed as a percentage, and the results can be presented either in order of frequency (total number of occurrences) or range (the number of different texts in which a word is found).

The following table shows the proportion of words found in the most frequent word lists and the AWL.

Table 4. Coverage of the Pharmacology Corpus by the first three word lists

Word list	Coverage of corpus
Most frequent 1,000 words	56.51%
2nd 1,000 most frequent words	4.46%
Academic Word List	9.47%
TOTAL	70.44%

When we compare this with the Academic Corpus and the Published Material Corpus, some striking differences can be seen. The biggest of these is the coverage that the most frequent 1,000 words give of the different corpora: 71% for the Academic Corpus, 80% for the PMC, and a mere 56% for the Pharmacology Corpus. We might expect the figure for pharmacology to be substantially lower than that for business, but it is surprising that it is also so much lower than the first 1,000 word coverage of the Academic Corpus. Both the PMC and the Pharmacology Corpus diverge considerably from the Academic Corpus, although in opposite ways: in business English, compared to "typical" academic English, many more terms are found in the most frequent lists; in pharmacology, on the other hand, the reverse appears to be true.

It is interesting to note that the AWL coverage is very similar for both the Academic Corpus and the Pharmacology Corpus (10.0% and 9.5%, respectively). The figure for the PMC, however, is a much lower 4.7%. This might be interpreted that knowledge of the AWL is indeed useful for pharmacology students, but perhaps less so for business students; it could, of course, be a reflection of differences between textbook and journal article genres.

Work was started on creating a pharmacology word list with the knowledge that the list

would have to increase coverage of the pharmacology corpus by 25% if 95% coverage was to be attained — no mean feat, bearing in mind that the Business Word List achieved less than 3% coverage. In order to produce the list, the words not found in the GSL or AWL were sorted according to their range across the 51 articles of the corpus. The criterion used was that words appearing in fewer than six articles were ineligible (an admittedly arbitrary cut-off point, but similar criteria were useful in the compilation of the Academic Word List and Business Word List). From the resulting word list, words considered to be either known or with a very low learning burden (e.g., proper names, nationalities, and numbers) were taken out, a procedure followed for the GSL, the AWL, and the BWL. Abbreviations and acronyms (e.g., O_2, ADH) were considered to fall into the same category, and were likewise removed.

On examination of the resulting word list it became apparent that there were a number of words which had been eliminated on grounds of insufficient range, but which either occurred with a noticeably high frequency or were very obviously part of the fundamental vocabulary of pharmacology. Probably these had not been picked up due to the relatively small size of the corpus; *parasympathetic*, for example, was found in only three articles, but it is certainly a word that all students of pharmacology should know. In order to ensure that such words were included in the list, it was decided to expand the list by including a frequency criterion. Words which appeared ten times or more, and appeared in at least two articles (to eliminate extremely specialized words) were selected. These words were then checked against the range list for cases of overlapping. The words that were not in the range list were added and, where appropriate, expanded into their families up to and including level 6 as described in Bauer and Nation (1993). This was a fairly straightforward procedure, but it was found in a few cases that a word was being added to the list even though it might be considered to belong to a word family in the GSL or AWL; the families in those lists had not been expanded sufficiently to include them. Examples include *interestingly* (*interesting* is in the GSL) and *cessation* (*cease* is in the AWL).

The words from the frequency list were added to the range list, making a total of 601 word families. The resulting Pharmacology Word List (PWL) was added to RANGE as a fourth list (see Appendix for a complete list of the headwords of this list). The following result was obtained (Table 5):

Table 5. Coverage of the Pharmacology Corpus by the first three word lists and the PWL

Word list	Coverage of corpus
Most frequent 1,000 words	56.51%
2nd 1,000 most frequent words	4.46%
Academic Word List	9.47%
Pharmacology Word List	12.91%
TOTAL	83.35%

The PWL is of similar size to both the AWL and the BWL, and thus fairly manageable, so a figure of around 13% coverage is certainly respectable (c.f. coverage of under 3% for the BWL and 5% coverage for Sutarsyah, Nation and Kennedy's economics word list). However, we still have some way to go to achieve even 90% coverage, let alone 95%.

One thing that became clear from an examination of the words not in the GSL, AWL, or PWL was that the decision not to consider abbreviations or acronyms meant that a large number of terms of both high frequency and wide range were being excluded. It was therefore decided to compile an additional list of these using the same range and frequency criteria, and to see to what extent total coverage would be increased. It was found that the resulting list of 140 items, the Pharmacology Abbreviations List (PAL), provided a not insubstantial coverage of 4.31% (see Table 6). When the PWL and PAL are combined, total coverage reaches 17.22%. Clearly, abbreviations and acronyms have an important part to play in the lexis of pharmacology journal articles.

Table 6. Coverage of the Pharmacology Corpus by the first three word lists, the PWL, and the PAL

Word list	Coverage of corpus
Most frequent 1,000 words	56.51%
2nd 1,000 most frequent words	4.46%
Academic Word List	9.47%
Pharmacology Word List	12.91%
Pharmacology Abbreviations	4.31%
TOTAL	87.66%

Table 6 shows mixed results — the combined PWL and list of abbreviations provides coverage of over 17%, which is almost twice that of the AWL; coverage has increased quite dramatically from around only 70% with the GSL and AWL to almost 88%. We are still, though, a long way from 95% coverage, and even the 93% attained by Konstantinis remains elusive.

Trying out the PWL on a different corpus

It would, of course, be advantageous to see whether the kind of coverage provided by the Pharmacology Word List can be replicated using a different pharmacology corpus. In order to get a better idea of the overall usefulness and validity of the PWL, the RANGE program was run on a corpus created from a pharmacology textbook (*Medical Pharmacology at a Glance,* Neal, 2003). Although the corpus at present consists of only one textbook (58,413 words in total), as an introduction to the subject this particular text covers all main areas of pharmacology. The findings are shown in Table 7.

Table 7. Coverage of the two pharmacology corpora by the first three word lists, the PWL, and the PAL

Word list	Coverage of corpus (journal articles)	Coverage of corpus (textbook)
Most frequent 1,000 words	56.51%	57.53%
2nd 1,000 most frequent words	4.46%	4.66%
Academic Word List	9.47%	6.58%
Pharmacology Word List	12.91%	14.76%
Pharmacology Abbreviations	4.31%	0.58%
TOTAL	87.66%	84.11%

We can see very similar coverage of both corpora by the two most frequent lists and the PWL; in fact the PWL figure is even better for the textbook corpus than for the article corpus (although abbreviations clearly have less of a role to play). This would seem to hint at the general applicability of a list such as the one created in this study, and its potential usefulness for learners who might need English either for reading and writing research articles, or for studying their specialist discipline at university using an English-language textbook. The AWL, on the other hand, would appear to be less useful for those studying with textbooks — a finding borne out by Konstantinis, who found that the AWL gave less than 5% coverage of the business corpus. Konstantanis (personal communication, January 2007) is currently constructing a corpus of business articles; it will be interesting to see if this corpus gives similar results to the business textbook corpus.

Discussion

The Pharmacology Word List was created using frequency and range criteria (with just a pinch of "expert" intuition); the fact that it provides coverage of around 13-14% on two different pharmacology corpora suggests its general usefulness and validity. The high coverage means that it has the potential to provide the ESP learner with an extremely useful set of words, and the return for effort invested in ensuring that these words are known would be considerable. Importantly, it is, like Coxhead's similarly-sized Academic Word List, small enough to be a realistic and attainable learning objective.

Tempering the result, however, is that despite a combined coverage of as much as 17%, the PWL and PAL together were unable to bring up total coverage to more than 88%, which is a long way off the ideal of 95%. Of course, we are assuming that a criterion of 95% — or, indeed, 98% — is applicable in a subject like medicine or pharmacology, and it is by no means certain that this is the case. We need to keep in mind that these figures were produced in studies investigating the reading of general texts for pleasure. Nevertheless, it cannot be argued that the 12% residue of words not found in any list comprises a sizeable proportion of unknown words.

Having produced our list of "core" pharmacology words, the next step will be to

determine how these words can best be taught and learned. To do this, it will first be necessary to look in more detail at the characteristics of the words. Fraser (2005) offers some suggestions for categorizing the different kinds of words found in a specialized text which will be useful in this regard. There are other types of specialized lexis in addition to "strictly" technical words such as *norepinephrine* and *vasodilator*; of particular interest are polysemous "cryptotechnical" words (*sympathetic* or *channel*, for example), which could be said to have hidden technical meanings in addition to their commonly-known meanings.

One problem with the study is that a large number of these cryptotechnical words are found in the most frequent 2,000 word lists and the AWL. Fraser (*ibid.*), investigating the lexis of a pharmacology textbook, found that as many as 25% of cryptotechnical words come from the two high-frequency word lists, and over 12% are found in the AWL. Even if the learners "know" all the words in the GSL and AWL, they may not necessarily be aware that many of them have an additional meaning in the field of pharmacology. Focusing only on the Pharmacology Word List, then, would not in itself be sufficient for learners: time would also need to be spent on those words in the other lists which have a particular, specialized meaning in the discipline. One way of getting around the problem might be to create a list based on frequency and range which does not distinguish between general purpose, academic, and specialized vocabulary; Ward (1999) found that a 2,000 word family vocabulary was sufficient to provide over 95% coverage of engineering texts. However, this approach might be best suited to those learners who have highly specialized goals from the early stages of their study, which is not the case with ESP learners at the graduate school level in Japan.

The polysemy of cryptotechnical words is only one of several factors influencing the difficulty that students will have when learning a word. Awareness on the part of teachers and syllabus designers of the features that make a word difficult to learn will affect their decisions regarding the presentation, practice, and testing of vocabulary. Even a brief glance at the Pharmacology Word List reveals a number of ways that the words in the list differ in their apparent learnability: there are the long words that are typical of a medical discipline (e.g., *electrophysiological, immunoreactivity*), but there are also many short words (e.g., *gut, kit*). There are words which intuitively appear to be problematic for learners to spell or pronounce, and others which would seem to present no difficulty at all. Laufer (1989b, 1997) suggests another category of words that may also affect learnability — "deceptively transparent" words. These are words which seem to provide clues to their meaning, but actually do not; for example, in *outline, out* does not mean *out of*. An example from the Pharmacology Word List might be *uptake*, which in fact means the act of taking something *in* rather than *up*. Another feature Laufer believes contributes towards learning difficulty is "synformy" (similarity of lexical forms, e.g., *comprehensive/comprehensible*); in pharmacology, perhaps pairs such as *mediate* and *medicate* will be confusing for learners. All of the above factors are worthy of investigation in future lexical studies of specialized texts.

It will also be instructive to look into the words that were *not* found in any list, in order to determine just how much of a problem these might pose to the learners. What kind of

words are they, and how important? Do they carry crucial information? Are most of them strictly technical words, as would appear at first glance, which are closely related to learning the subject and thus might present conceptual, rather than linguistic difficulties? Even if the words are unknown, many of them are of Greco-Latin origin and are made up of regular affixes which can fairly easily be learned. There are, in fact, several regularly-occurring prefixes such as *acetyl-*, *chemo-*, *hydroxy-*, *hyper-*, *hypo-*, *pre-* and *super-* which would certainly be important to know. And if many of the terms not in the lists are the names of proprietary drugs, for instance, then perhaps they can be treated in a similar manner to proper names — it is, after all, not necessary to learn the names of all the characters in a novel beforehand in order to be able to follow the plot!

Conclusion

A main aim of the present study was to create a list of words, using frequency and range criteria, which are representative of the field of pharmacology, and which, in conjunction with the most frequent words and the Academic Word List, will enable learners to reach the threshold for satisfactory reading comprehension. Despite the abovementioned reservations, the study has, I feel, achieved some success in this: a list was created which achieved 13% coverage of the corpus of journal articles, and almost 15% coverage of a pharmacology textbook corpus — figures which compare very well with the coverage provided by the AWL. The study has also identified several factors, deserving of further investigation, which could affect the difficulty level of the words in the list and hence their priority in teaching.

Future research might include carrying out a similar study using a larger corpus; the existing pharmacology corpus consists of only 185,000 words, and so is relatively small, even for a specialized corpus. Ideally, the corpus should consist of at least 100 articles, or around 300,000 words. Further studies could also experiment with different range and frequency criteria, as those used in the present study are not necessarily the most appropriate. It would be desirable, too, to carry out a parallel study on an expanded pharmacology textbook corpus in order to further explore the similarities and differences between the two genres.

References

Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language 4* (2), 91-105.

Bauer, L. & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6* (3), 1-27.

Chujo, K. & Utiyama, M. (2006). Selecting level-specific vocabulary using statistical measures. *System 34*, 255-269.

Cobb, T. & Horst, M. (2001). Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.) *Research Perspectives on English for Academic Purposes* (pp. 315-329). Cambridge: Cambridge University Press.

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

Farrell, P. (1990). *Vocabulary in ESP: a Lexical Analysis of the English of Electronics and a study of Semi-technical Vocabulary.* CLCS Occasional Paper 25, Dublin: Trinity College.

Francis, W. N. & Kucera, H. (1982). *Frequency Analysis of English Usage.* Boston: Houghton Mifflin Company.

Fraser, S. (2005). The lexical characteristics of specialized texts. In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.) *JALT 2004 Conference Proceedings.* Tokyo: JALT.

Hirsh, D. & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8,* 689-696.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language 13* (1), 403-430.

Hwang, K. & Nation, I. S. P. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System 23* (1), 35-41.

Japan Association of College English Teachers (2003). *JACET List of 8000 Basic Words.* Tokyo: JACET.

Konstantanis, N. (2006). An attempt to create a Business Word List. Presentation given at the 16th Vocabulary Acquisition Research Group Network Conference, University of Wales, Swansea.

Laufer, B. (1989a). What percentage of text-lexis is necessary for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316-323). Clevedon: Multilingual Matters.

Laufer, B. (1989b). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review, 6,* 10-20.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.) *Vocabulary and Applied Linguistics.* (pp. 126-132). London: Macmillan.

Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In N. Schmitt, and M. McCarthy (Eds.) *Vocabulary: Description, Acquisition and Pedagogy* (pp. 140-153). Cambridge: Cambridge University Press.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes 25,* 235-256.

Nation, I. S. P. & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt, & M. McCarthy (Eds.) *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.) *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 3-13). Amsterdam: John Benjamins.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63* (1), 59-82.

Neal, M. J. (2003). *Medical Pharmacology at a Glance.* Oxford: Blackwell Science.

Nelson, M. (2000). *A Corpus-Based Study of Business English and Business English Teaching Materials.* Unpublished thesis, University of Manchester, Manchester, England.

Salager, F. (1985). Specialist medical English lexis: Classificatory framework and rhetorical function — a statistical approach. *EMP Newsletter, 2,* (2), 5-17.

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus-based study. *RELC Journal, 25,* 34-50.

Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language, 12,* 309-323.

West, M. (1953). *A General Service List of English Words.* London: Longman, Green & Co.

Xue, G. & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication 3,* 215-29.

Appendix: Headwords of the *Pharmacology Word List*

ABBREVIATION
ABNORMALITY
ABOLISH
ABSORB
ABUSE
ACCUMBENS
ACETONITRILE
ACETATE
ACETYLCHOLINE
ACID
ACTIN
ACTIVATE
ACUTE
ADENOSINE
ADENYL
ADENYLATE
ADIPOSE
ADJUVANT
ADMINISTER
ADRENALINE
ADRENOCEPTOR
ADVERSE
AFFERENT
AFFINITY
AGONIST
AIRWAY
ALBUMIN
ALCOHOL
ALIQUOT
ALLELE
ALVEOLAR
AMINE
AMINO
AMMONIUM
AMOXICILLIN
AMPHETAMINE
AMPLITUDE
AMYGDALA
AMYLOID
ANALGESIA
ANALOG
ANESTHETIZE
ANGINA
ANGIOTENSIN
ANOVA
ANTAGONIST
ANTERIOR
ANTIBIOTIC
ANTIBODY
ANTI-HUMAN
ANTI-INFLAMMATORY
ANTINOCICEPTION
ANTIOXIDANT
AORTA
APOPTOSIS
AROMATIC
ARSENIC
ARTERIOLE
ARTERY
ARTHRITIS
ASSAY
ASTHMA
ATENOLOL
ATRIUM
ATROPINE
ATTENUATE
AUTONOMIC
AXON
BACTERIA
BASAL
BASELINE
BATTERY
BIOAVAILABILITY
BENZODIAZEPINE
BIOASSAY
BIOCHEMISTRY
BIOLOGY
BIOSYNTHESIS
BLADDER
BLOCKADE
BLOCKER
BLOT
BOVINE
BUFFER
CAFFEINE
CALCIUM
CALIBRATE
CANCER
CANDIDATE
CANNABIS
CAPSAICIN
CAPSULE
CARBACHOL
CARBON
CARCINOGEN
CARDIAC
CARDIOVASCULAR
CARTILAGE
CATALASE
CATALYST
CATION
CAUDAL
CAVITY
CELL
CENTRIFUGE
CEREBELLUM
CEREBRUM
CERVICAL
CESSATION
CHROMATOGRAPHY
CHAMBER
CHEMOTACTIC
CHEMOTHERAPY
CHLORIDE
CHOLESTEROL
CHOLINERGIC
CHRONIC

CIRCULATION
CLEARANCE
CLEAVE
CLINICAL
CLONE
CLUSTER
CO-ADMIMISTRATION
COCAINE
COEFFICIENT
COGNITION
COLLAGEN
COLUMN
COMPARTMENT
COMPETITIVE
COMPLEXES
COMPLY
CONCENTRATIONS
CONCOMITANT
CONDITIONING
CONDUCTION
CONFIGURATION
CONSECUTIVE
CONTRACTION
CORD
CORONARY
CORRELATE
CORTEX
CORTISOL
COUNTER
CREATININE
CROSSOVER
CUE
CUMULATIVE
CURRENTS
CYCLASE
CYSTEINE
CYTOCHROME
CYTOKINE
CYTOSOLIC
CYTOTOXIC
DEAMINASE
DEFICIENCY
DEFICIT
DEGRADE
DEPENDENCE
DEPLETE
DEPOLARIZE
DERMIS
DIABETES
DIAGNOSE
DIAMETER
DIASTOLE
DIET
DIFFERENTIAL
DIFFUSION
DILTIAZEM
DILUTE
DISCREPANCY
DISMUTASE
DISORDER
DISSECT
DISSOCIATE
DISSOLVE
DISTIL
DONATE
DOPAMINE
DORSAL
DOSE
DOWNREGULATION
DRUG
DYE
DYSFUNCTION
DENSITY
EFFICACY
ELECTROPHYSIOLOGICAL
ELECTRODE
ELECTRONIC
ELEVATE
ELICIT
ELUTE
EMBRYO
EMISSION
ENCODING
ENDOGENOUS
ENDOTHELIN-1
ENDOTHELIUM
ENZYME
EPINEPHRINE
EQUILIBRATE
ERADICATE
ERYTHROCYTE
ERYTHROMYCIN
ESCALATE
ESOMEPRAZOLE
ESTER
ETHANOL
EVOKE
EXCITATION
EXCRETE
EXERT
EXOGENOUS
EXON
FASTING
FEED
FEMORAL
FENTANYL
FETUS
FEXOFENADINE
FIBROSIS
FLUID
FLUORESCENCE
FLUX
FRACTION
FRONTAL
FUNCTIONALITY
FUSE
GABA
GANGLION
GASTRIC
GASTROINTESTINAL
GEL
GENE
GENOTYPE
GESTATION
GLIBENCLAMIDE
GLUCOCORTICOID

GLUCORONIDATION
GLUCOSE
GLUTAMATE
GLUTATHIONE
GLYCEMIC
GRADIENT
GRAPH
GUINEA
GUT
HALF-LIFE
HALF-MAXIMAL
HALF-TIME
HAPLOTYPE
HEPATIC
HEMOGLOBIN
HEPES
HEROIN
HETEROLOGOUS
HIGH-AFFINITY
HIND
HIPPOCAMPUS
HISTOLOGY
HOMEOSTASIS
HOMOGENIZE
HOMOZYGOUS
HORIZONTAL
HORMONE
HORN
HYDROCHLORIDE
HYDROCORTISONE
HYDROGEN
HYPERALGESIA
HYPERPOLARIZE
HYPERSENSITIVITY
HYPERTHERMIA
HYPERTROPHIC
HYPOXIA
HYPERTENSION
ICE-COLD
ILEUM
IMAGING
IMMUNOREACTIVITY
IMPAIR
IMPLANT
INCISION
INCUBATE
INDOMETHACIN
INDUCER
INFECT
INFILTRATE
INFLAMMATION
INFLUX
INFUSE
INHALE
INHIBITOR
INJECT
INOTROPY
INSENSITIVE
INSULIN
INTACT
INTAKE
INTERESTINGLY
INTESTINE
ION
IPSILATERAL
ISCHEMIC
ISOFORM
ISOPROTERENOL
ITRACONAZOLE
KETAMINE
KIDNEY
KINASE
KINETIC
KIT
KNOCKOUT
KREBS
LABORATORY
LAMINA
LANE
LATENCY
LATERAL
LIBITUM
LIGAND
LIGATION
LINEAR
LIVER
LOCALIZE
LONG-TERM
LOW-DOSE
MACROPHAGE
MAGNITUDE
MALARIA
MALONDIALDEHYDE
MAMMAL
MARIJUANA
MARKEDLY
MATRIX
MAXIMAL
MEDIAN
MEDIATOR
MEDICATE
MEMBRANE
MESENTERY
METABOLISE
METHACHOLINE
METHAMPHETAMINE
METHANOL
METRIC
MICROBE
MICROMOLAR
MICROSCOPE
MICROSOME
MITOCHONDRIA
MOBILE
MODULATE
MOLECULE
MONOCYTE
MORPHINE
MOUNT
MURINE
MUSCARINE
MUSCLE
MUTATION
MYOCARDIUM
MYOCYTE
NAIVE

NAUSEA
NEGLIGIBLE
NEURON
NEUROPATHIC
NEUROTRANSMITTER
NECROSIS
NERVOUS
NICOTINE
NIFEDIPINE
NITRIC
NOCICEPTIVE
NOCTURNAL
NOMENCLATURE
NORADRENALINE
NOREPINEPHRINE
NOVEL
NUCLEOTIDE
NUCLEUS
OBESITY
ONE-WAY
ONSET
OOCYTE
OPEN-LABEL
OPIOID
OPTIMAL
ORAL
ORBITOFRONTAL
ORGANIC
OVERLOAD
OXIDATION
OXYGEN
PACEMAKER
PARASYMPATHETIC
PATCH
PATHOGEN
PATHOLOGY
PATHWAY
PATIENTS
PEAK
PEANUT
PEDIATRIC
PELLET
PENETRATION
PENTOBARBITAL
PEPTIDE
PERCENTILE
PERFUSE
PERIPHERY
PERMEABILITY
PEROXIDASE
PEROXIDE
P-GLYCOPROTEIN
PHARMACEUTICAL
PHARMACOKINETIC
PHARMACOLOGY
PHENOTYPE
PHOSPHATE
PHOSPHORYLATION
PHYSIOLOGY
PINACIDIL
PLACEBO
PIPETTE
PLASMA
PLOT
POLYMORPHISM
PORE
POSTERIOR
POSTULATE
POTASSIUM
POTENT
POTENTIALS
PREFRONTAL
PREGNANCY
PRENATAL
PRETREATMENT
PRESCRIPTION
PREVALENCE
PRIMATE
PROBABILISTIC
PROBE
PROFILE
PROLIFERATION
PROLONG
PROMINENCE
PROSTAGLANDIN
PROTEASE
PROTEIN
PROTEOLYTIC
PULMONARY
PURIFY
PUTATIVE
PYRUVATE
QUANTIFY
QUANTITATIVE
RADIOACTIVITY
RANDOMIZE
REACTIVITY
REAGENT
RECOMBINANT
RECORDINGS
RECEPTOR
RECRUIT
REDOX
REGIMEN
REGRESS
REINSTATE
RELAPSE
RELAXANT
REMODEL
RENAL
RESIDUAL
RESIDUE
RESISTANT
RESPIRATION
RETICULUM
RHYTHM
RODENT
ROUTINE
SALINE
SALIVA
SATURATE
SCIATIC
SCINTILLATION
SCORE
SECRETE
SEGMENT

SELECTIVITY
SELF-ADMINISTRATION
SENSITIZE
SENSORY
SEROTONIN
SERUM
SESSION
SETTINGS
SHORT-TERM
SIDE-EFFECT
SIGMA
SIGNALING
SIMULTANEOUS
SLICE
SODIUM
SOFTWARE
SOLUBILITY
SOLVENT
SPATIAL
SPECIES
SPECTRUM
SPINE
SPLICE
SPONTANEOUS
STEROID
STIMULATE
STRAIN
STRAND
STRATUM
STREPTOMYCIN
SUBCUTANEOUS
SUBSTRATE
SUBTLE
SUBTYPE
SUBUNIT
SULINDAC
SUPERFUSION
SUPERIOR
SUPERNATANT
SUPEROXIDE
SUPPRESS
SURGERY
SUSCEPTIBLE
SYMPTOM
SYNAPSE
SYNERGISTIC
SYNTHESISE
SYSTEMIC
SYSTOLE
TEMPORAL
TERTIARY
TETRODOTOXIN
THERAPY
THERMAL
THIOPENTAL
THORAX
THRESHOLD
THYROID
TISSUE
TOLERATE
TONE
TOXICITY
TRACT
TRANSCRIPTION
TRANSDERMAL
TRANSDUCE
TRANSFECTION
TRANSIENT
TRANSLOCATION
TRANSMEMBRANE
TRANSPLANT
TRIPLE
T-TEST
T-TYPE
TUBERCULOSIS
TUMOR
TYRODE
UNCLEAR
UNPAIRED
UNTREATED
UPTAKE
URINE
VAGAL
VASCULAR
VASOCONSTRICTION
VASODILATION
VASORELAXATION
VEIN
VELOCITY
VENOM
VENTRICLE
VERBAL
VERIFY
VERSUS
VIABILITY
VITRO
VIVO
VOLTAGE
VOMIT
WARRANT
WASHOUT
WAVELENGTH
WHOLE-CELL
WILD-TYPE
WITHDRAWAL
ZEISS

要　約

専門分野におけるコーパスを利用したワードリストに作成とその有効性

サイモン・フレイザー
広島大学外国語教育研究センター

近年，英語教育においてESP（特定の目的のための英語）の持つ役割への関心が高まってきている。日本においても，学術的な目的だけでなく，専門分野において，将来的に必要となってくる英語を，大学院レベルで教授していくべきであるとの見解が，EFL（外国語としての英語）教育に携わる教員間で強まっている。この中で，特定の目的のための言語を広範囲に教授するためには，その目的にかなった単語の選定を行い，学習者に提供していく必要がる。

そのため，本研究では，専門分野のコーパスを利用したワードリストを作成することにより，頻度と分布範囲に基づき，601のワードファミリーからなるリスト（PWL）を作成した。この比較的小規模な簡易ワードリストは，コーパス全体の13%をカバーし，略語等を含めると17%のカバー率となる。最頻出2000語，アカデミックワードリスト，及びPWLの総カバー率（84.1～87.6%）は十分な読解のために必要とされる95%レベルよりも低いカバー率ではあるが，特定の目的を持つ学習者にとって，専門分野のワードリストがいかに有効であるかを明確に示している。

最後に，ワードリスト中の単語自体が学習の難易度に影響を及ぼす可能性があり，教授の際に重要となってくるいくつかの要因について，更なる調査の必要性を指摘した。