Breaking Down the Divisions between General, Academic, and Technical Vocabulary: The Establishment of a Single, Discipline-based Word List for ESP Learners

Simon FRASER
Institute for Foreign Language Research and Education
Hiroshima University

Few language learners or teachers would disagree that vocabulary acquisition is a fundamental component of any language course. This is no less true for learners of English for Specific Purposes (ESP), for whom the most serious difficulties are almost certainly lexical, particularly at intermediate and advanced levels of reading (Ulijin, 1984). The importance of an early emphasis on L2 vocabulary has been stressed by Meara (1995), and according to Laufer (1989) massive vocabulary expansion should be one of the major goals in any course in Language for Academic Purposes. The question remains, though, as to exactly what kind of vocabulary should be taught.

This paper reports on a study which aims to quantify the vocabulary required by students of pharmacology. The findings are of particular interest, as the medical field, with its notoriously "difficult" lexis, has largely been neglected in work of this kind.

FREQUENCY LISTS IN ESP

Previous studies have had some success in creating specialized word lists which can help ESP learners reach satisfactory levels of reading comprehension (e.g., Baker, 1988; Farrell, 1990; Fraser, 2007; Hirsh & Nation, 1992; Konstantakis, 2006; Salager, 1983; and Sutarsyah, Nation & Kennedy, 1994). It has been suggested (Laufer, 1989, 1992; Hirsh & Nation, 1992) that a lexical coverage of 95% in a text is the threshold for learners to achieve adequate reading comprehension. Specialized word lists are usually learned in conjunction with general purpose and academic word lists in order to reach this kind of coverage. Konstantakis (2006), for instance, used frequency and range criteria to devise a 561-word Business Word List. His list, together with the most frequent 2,000 words and Coxhead's (2000) 570-word Academic Word List (AWL), provided greater than 93% coverage of a business English corpus. My own 601-word Pharmacology Word List (Fraser, 2007), however, brought total coverage of a corpus of pharmacology journal articles up to only 88%.

The usefulness of the AWL is, in fact, increasingly being called into question. The assumption that the AWL is representative of a wide range of disciplines may be flawed: Hyland & Tse (2007) found that only 36 AWL word families in their own corpus were evenly distributed across science, engineering, and social science subcorpora. It may be, too, that the AWL is far from complete in representing the academic words used in certain genres. Chen & Ge (2007) found that only 51% of AWL items are used in medical research articles, and that

many high frequency academic words are rarely used in such articles. What all this means is that learners may spend a lot of time on words they don't need to know.

THE PRESENT STUDY

The method adopted here, therefore, was to construct a single list based on frequency and range which does not distinguish between general, academic, and specialized vocabulary, and to determine the coverage that different frequency levels provide. In the field of engineering, Ward (1999) was able to create a 2,000-word family vocabulary list that provided over 95% coverage of texts. The main aim of this study was to establish whether a similar list could be compiled for a medical discipline such as pharmacology.

Creating a Pharmacology Word List

A specially created corpus of 100 pharmacology research articles (360,000 running words) taken from a wide international selection of pharmacology journals was used to construct a list of 2,000 word families based on frequency and range. A frequency list of words in the Pharmacology Corpus was produced using Wordsmith Tools (Scott, 1999), and the words in the list expanded into their word families (Level 6, Bauer and Nation, 1993). Two lists of words were created: the 1,000 most frequent word families, and the second 1,000 most frequent word families. The words had to appear in at least two articles, and the vast majority appeared in far more than this. The resulting Pharmacology Word List (PWL) was run against Nation's RANGE computer program (available at http://www.wuw.ac.nz/lals/) to determine its coverage of the pharmacology corpus.

Coverage Provided by the List

Tables 1 and 2 show that this list gives a far higher coverage of the Pharmacology Corpus than does a 2,000-word general word list (The General Service List, West, 1953) coupled with the 570-word Academic Word List (Coxhead, 2000).

Table 1. Coverage of the Pharmacology Corpus by the Most Frequent 2,000 Words and the Academic Word List

Word list	Coverage of corpus (%)
Most frequent 1,000 words (GSL)	56.5
2nd 1,000 most frequent words (GSL)	4.5
Academic Word List	9.5
TOTAL	70.5

Table 2. Coverage of the Pharmacology Corpus by the Pharmacology Word List

Word list	Coverage of corpus (%)
Most frequent 1,000 words (PWL)	82.3
2nd 1,000 most frequent words (PWL)	6.8
TOTAL	89.1

In order to see whether this level of coverage could be replicated using different pharmacology corpora, the RANGE program was run on a 60,000-word corpus created from a pharmacology textbook (*Medical Pharmacology at a Glance*, Neal, 2003), and on a 250,000-word corpus extracted from the web using the WebBootCaT toolkit (Baroni and Bernadini, 2004, software available at http://www.sketchengine.co.uk/). WebBootCaT works by using "seed words" selected by the compiler and sending permutations of these terms to the Yahoo! search engine. The thirteen words chosen as seed terms were those I considered to be the most salient in the PWL (e.g., *receptor, drug, antagonist*). The corpus that was created consisted of journal papers, magazine articles, extracts, and abstracts.

Although the coverage falls short of that found by Ward (1999) with his engineering list, the figures of well over 80% for all pharmacology corpora are very respectable, supporting the general applicability and validity of the list (see Tables 3 and 4).

Table 3. Coverage of Engineering Corpora by Ward's Engineering Word List

Corpus	1st 1,000 words (%)	2nd 1,000 words (%)	Total 2,000 words (%)
Engineering	92.1	3.6	95.7
Mechanics	92.6	4.3	96.9

Table 4. Coverage of Other Pharmacology Corpora by the Pharmacology Word List

Corpus	1st 1,000 words (%)	2nd 1,000 words (%)	Total 2,000 words (%)
Pharmacology	82.3	6.8	89.1
Pharm. textbook	78.7	7.6	86.3
Pharm. BootCaT	74.3	7.6	81.9

Sub-dividing the word lists into groups of 500 words is illuminating (see Table 5): we see that well over 70% coverage is given by as few as 500 words. It will certainly be well worth our while to focus on this group of words, especially as the next 1,500 words increase coverage by only 16%. What, then, are the characteristics of these most useful words, and what kind of difficulties are they likely to present learners with?

Table 5. Coverage of the Pharmacology Corpus at Different Frequency Levels

Word list	Coverage of corpus (%)
Most frequent 500 words	73.0
2nd 500 most frequent words	9.3
3rd 500 most frequent words	4.1
4th 500 most frequent words	2.7
TOTAL	89.1

The Most Frequent Words

One goal of this research is to create a list which will provide comprehensive coverage. Once we have discerned what the most frequent words are, the list will also function as a starting point for investigations into the characteristics of the most useful pharmacology words and factors affecting their learnability.

Table 6 shows the most frequent 100 words in the PWL. The first ten words or so differ little from the most frequent words in a general list, but the following words are those that can be said to be at the "core" of pharmacology, and tell us quite a lot about the field. After extracting the grammatical/function words, the top ten words are *et al, effect, activity, study, use, cell, receptor, increase, treatment,* and *inhibition.* What is striking is that almost all of these words will be familiar to the layperson. This implies that many of the most frequent words will be found in general or academic lists, which is indeed found to be the case (see Tables 7 and 8).

Table 6. The Top 100 Words in the Pharmacology Word List and Frequency of Occurrence

1. THE	19337	35. INDUCE	1024	69. THEIR	584
2. OF	15785	36. DRUG	988	70. FUNCTION	578
3. BE	12956	37. PATIENT	987	71. OTHER	575
4. AND	11457	38. DOSE	935	72. BASE	563
5. IN	10277	39. SIGNIFICANT	915	73. DEPENDENT	561
6. TO	7032	40. DAY	911	74. PRODUCE	560
7. A	6266	41. RAT	909	75. MIN (MINUTE)	558
8. WITH	3902	42. LEVEL	835	76. BOTH	556
9. FOR	3461	43. CONTROL	830	77. LOW	549
10. THAT	3010	44. RESULT	822	78. TIME	537
11. BY	2756	45. DIFFERENT	803	79. REDUCE	533
12. THIS	2663	46. WHICH	799	80. BUT	527
13. AS	2279	47. MG (MILLIGRAM)	789	81. THERAPY	517
14. ET AL	2087	48. OUR	759	82. TEST	514
15. ON	1979	49. ALSO	757	83. FIND	510
16. EFFECT	1827	50. EXPRESSION	738	84. ALL	503
17. OR	1753	51. HIGH	724	85. MEAN	497
18. FROM	1752	52. AFTER	711	86. ADDITION	495
19. ACTIVITY	1727	53. PRESENT	706	87. REPORT	484
20. AT	1716	54. DATA	697	88. OBSERVE	482
21. STUDY	1674	55. GROUP	663	89. ADMINISTRATION	472
22. USE	1647	56. NO	663	90. STIMULATION	472
23. CELL	1429	57. PROTEIN	651	91. ASSOCIATE	465
24. AN	1308	58. CALCIUM	650	92. SUGGEST	459
25. RECEPTOR	1303	59. MAY	647	93. MEASURE	458
26. INCREASE	1260	60. RELATE	643	94. EACH	455
27. NOT	1180	61. HUMAN	633	95. REGULATION	455
28. IT	1156	62. BETWEEN	630	96. DO	452
29. TREATMENT	1125	63. ANALYSIS	624	97. INDICATE	449
30. INHIBITION	1118	64. MOUSE	621	98. BLOOD	448
31. SHOW	1040	65. COMPARE	616	99. VARIOUS	448
32. CONCENTRATION	1039	66. MODEL	603	100. FOLLOW	446
33. RESPONSE	1037	67. THAN	598		
34. FIGURE	1024	68. EXPERIMENT	595		

Table 7. "General" Words in the Pharmacology Top 100

Word family	Frequency	Word family	Frequency
1. EFFECT**	1827	24. EXPERIMENT*	595
2. ACTIVITY**	1727	25. BASE**	563
3. STUDY	1674	26. DEPENDENT**	561
4. DIFFERENT	1647	27. PRODUCE	560
5. USE	1647	28. MIN (MINUTE)	558
6. INCREASE	1260	29. LOW	549
7. TREATMENT**	1125	30. REDUCE**	533
8. SHOW	1040	31. TEST	514
9. FIGURE**	1024	32. FIND	510
10. PATIENT*	987	33. MEAN**	497
11. RAT	909	34. ADDITION	495
12. CONTROL**	830	35. ASSOCIATE**	465
13. RESULT	822	36. SUGGEST	459
14. MG (MILLIGRAM)	789	37. MEASURE	458
15. EXPRESSION**	738	38. BLOOD*	448
16. HIGH	724	39. VARIOUS	448
17. PRESENT**	706	40. FOLLOW	446
18. GROUP	663	41. AGE	439
19. RELATE	643	42. VALUE**	422
20. HUMAN	633	43. CHANGE	420
21. MOUSE	621	44. INCLUDE	396
22. COMPARE	616	45. DETERMINE	394
23. MODEL**	603		

* "Lay-technical" words with a technical flavour, but likely to be known by the layperson
** "Cryptotechnical" words which have an additional, more specialized meaning

When we look closely at the words that are also found in general lists, we see that a sizeable proportion of them have technical meanings in addition to their commonly known meanings. We might label these polysemous words "cryptotechnical" because their specialized meanings are not always apparent. Words such as *expression* and *control*, for instance, have meanings in pharmacology quite removed from their general ones. Similarly, a number of the words which are found in the AWL also are used with a quite different sense in pharmacology (see Table 8).

Table 8. AWL Words in the Top 100

Word family	Frequency	Word family	Frequency
1. INHIBITION*	1118	7. FUNCTION	578
2. CONCENTRATION*	1039	8. ADMINISTRATION*	472
3. INDUCE*	1024	9. REGULATION*	455
4. SIGNIFICANT*	915	10. POTENTIAL*	428
5. DATA	697	11. CHANNEL*	417
6. ANALYSIS	624	12. ROLE	405

* Cryptotechnical words

If we turn our attention to those high frequency words which are not found in either general lists or the AWL (Table 9), we see that perhaps only one of these words – *receptor* – is what might be considered strictly technical; the others are mainly either lay-technical (e.g., *drug*) or cryptotechnical (e.g., *stimulation*).

Table 9. Words in the Top 100 Not Found in General Lists or AWL

Word family	Frequency	Word family	Frequency
1. ET AL	2087	7. PROTEIN*	651
2. EFFECT**	1827	8. CALCIUM*	650
3. CELL*	1429	9. ADDITION	493
4. RECEPTOR	1303	10. REPORT*	484
5. DRUG*	958	11. STIMULATION**	472
6. DOSE*	935	12. INDICATE**	449

* Lay-technical words
** Cryptotechnical words

Tables 10 and 11 show that many of the words that are also found in the GSL or AWL are used with a specialized meaning in pharmacology. Learners who "know" these words will not necessarily be aware of their additional uses.

Table 10. Cryptotechnical GSL Words in the Pharmacology Top 500

Word family	Frequency	Word family	Frequency
1. EFFECT	1827	17. RATE	320
2. ACTIVITY	1727	18. BLOCK	305
3. TREATMENT	1125	19. SAMPLE	302
4. FIGURE	1024	20. SENSITIVE	261
5. CONTROL	830	21. TRIAL	253
6. EXPRESSION	738	22. PRESSURE	219
7. PRESENT	706	23. POPULATION	177
8. MODEL	603	24. ORDER	157
9. BASE	563	25. RESISTANT	146
10. DEPENDENT	561	26. SMOOTH	137
11. REDUCE	533	27. DELIVERY	126
12. MEAN	497	28. FAILURE	125
13. ASSOCIATE	465	29. TERM	121
14. VALUE	422	30. GROWTH	114
15. TABLE	383	31. LOCAL	114
16. ACTION	350	32. FATTY	103

Table 11. Cryptotechnical AWL Words in the Pharmacology Top 500

Word family	Frequency	Word family	Frequency
1. INHIBITION	1118	13. REACTION	264
2. CONCENTRATION	1039	14. CONTRACTION	254
3. RESPONSE	1037	15. COMPOUND	237
4. SIGNIFICANT	915	16. INTERACTION	204
5. ADMINISTRATION	472	17. TARGET	201
6. REGULATION	455	18. TRANSPORT	180
7. POTENTIAL	428	19. STRESS	166
8. CHANNEL	417	20. DISTRIBUTION	158
9. BIND	359	21. CULTURE	153
10. RELEASE	346	22. STABLE	143
11. SOLUTION	317	23. VEHICLE	133
12. RELAXATION	294	24. MEDIUM	105

Cryptotechnical Words: Backing up Intuitions with Concordance Evidence

It is being suggested that the cryptotechnical words found in the AWL and GSL are of particular importance, as it is easy to miss or ignore the fact that these "basic" words may have quite different meanings in pharmacology. GSL words are especially worthy of our attention, as they are likely to be considered to be already known and disregarded completely. However, just because certain words have the potential to be used with a technical meaning, it does not necessarily follow that they *are* being used in that way in the corpus. I therefore decided to see if my intuitions about cryptotechnical words could be confirmed by concordance evidence. Are the words really cryptotechnical, and if so, is the specialized meaning the major one, or is the general meaning more common? Are there, in fact, cases where the words are used only with their general senses? To determine the ways in which the words are actually used, their most frequent collocations were examined. In nearly all instances it was found that the word was indeed being employed with its technical meaning, although there are a number of interesting exceptions. Some examples of the most commonly found clusters are given below.

GSL Cryptotechnical Words

Words used primarily with technical sense (21/28 = 75%):

ACTION

Used to describe the function of a drug: *action(s) of [drug], mechanism of action*; also used with *potential* to mean a change in membrane potential occurring in excitable tissue: *action potential(s).*

BLOCK

Found with the specific pharmacological meaning "prevent the action of a drug": *blockade of [receptors], channel blocker(s), beta blocker(s).*

CONTROL

Found to be mainly used with the meaning "comparison group used in a scientific study": *control group, control conditions, control mice*; some instances of the word being used to mean "regulation" were also found: *controlled by, tumour control.*

DELIVERY

This was always found with the meaning "administration or transfer of a drug": *delivery of [drug], insulin delivery, synaptic delivery.*

EXPRESSION

Found mostly to mean "detectable effect of certain protein molecules": *gene expression, protein expression, receptor expression.*

FAILURE

As expected, this is used to refer to the situation where a body organ stops working correctly: *heart/cardiac failure, respiratory failure;* but also *treatment failure.*

FATTY

This refers to any of the numerous compounds of carbon, hydrogen, and oxygen that are soluble in organic solvents, and was found almost exclusively in the combination *fatty acid(s);* but also *fatty degeneration, fatty fish, high fat.*

LOCAL

Has the meaning "involving only a restricted part of a living organism": *Localization of [receptor], local tissue, localised to;* very rarely referring to a particular place: *local hospital, local slaughterhouse, local animal ethics committee.*

MEAN

Most commonly used with the statistical sense "average of a set of numbers": *the mean, the means, mean SD*; also found, though, occurring in the phrase *by means of.*

MODEL

Found mainly referring to an idealized representation of a condition or process: *model of [clinical condition], pharmacokinetic model*; but also frequently used to mean an animal that is used to mimic a pathological condition: *animal model(s).*

POPULATION

Exclusively used to refer to a set from which samples are taken for statistical measurement: *population spike, study population, population distribution(s).*

PRESSURE

Used with the meaning "force at which a fluid (usually blood) is pushed through an organ or tissue": *blood pressure, perfusion pressure, arterial pressure.*

RATE

Used with the meaning "speed at which a biochemical or physiological process occurs within a particular time period": *eradication rate(s), heart rate, flow rate.*

RESISTANT

This was found with the meaning "not harmed or affected by the actions of an organism or drug": *resistant to [drug/drug effect], [drug] resistant.*

SAMPLE

Found to be used with its medical/pharmacological meaning "specimen taken for medical analysis": *blood sample, urine sample*; also (statistical) to mean a selection of items or individuals taken from a population: *sample size.*

SENSITIVE

Found with the meaning "responding to the effect of a drug": *[drug name] sensitive, sensitive to [drug/drug effect].*

SMOOTH

This word was found to occur almost exclusively in combination with *muscle: smooth muscle, smooth muscle cells.*

TABLE

Unsurprisingly, used exclusively to mean "systematically arranged list of data": *Table 1, shown in table, given in table.*

TREATMENT

As expected, used with the meaning "medical or surgical management of a patient": *treatment of [condition], treatment with [drug], treatment group(s).*

TRIAL

As expected, found to refer to the process of testing a drug in order to assess its effectiveness: *clinical trial(s), controlled trial(s), phase III trials.*

VALUE

Found with its statistical sense of being a particular quantitative determination: *IC50 value(s), control value(s), p value.*

Words used with both general and technical senses / unclear (4/28 = 14%):

GROWTH

Found to refer to the increase in size or development of living things in general, but also used in the following more specialized ways: *tumor growth, cell growth, growth of [tissue]*; also common are the technical *growth factor(s)* and *growth hormone.*

PRESENT

Mainly used with general senses: *present study, present in, data are presented*; few examples of medical usage, e.g., *present with [drug]*; surprisingly rare to find the meaning "appear for examination or treatment": *case presentation, present with symptoms, present with clinical signs of arthritis.*

REDUCE

Found mostly with the meaning "to make smaller in size or amount": *significantly reduced, reduced by 70%, reduce the risk*; much more rarely used with a biochemical meaning: *reduced glutathione; reduced beta-endorphin.*

TERM

Found almost exclusively in the expressions *long term, short term*; also common is *in terms of*; more rarely, specialized uses are found: *term for treatment, nonlinear term.*

Words used primarily with general sense (i.e., probably not cryptotechnical (3/28 = 11%):

ASSOCIATE

By far the most common use is in *be associated with*; also common are *associated risks, associated effects*; there are no instances of the chemical meaning as in *associated bonds*; however, there is a rarely found specialized meaning: *drug cue associations; association rate constant.*

BASE

Most commonly found as part of the phrasal verb *based on*; but also *baseline level, baseline flow*; no instances of the pharmaceutical meaning (chief ingredient of a

mixture) or the chemical meaning (a cation that unites with an anion to form a salt).

DEPENDENT

Found mainly with the meaning "determined or conditioned by something else": *concentration dependent, dose dependent, voltage dependent*; interestingly, and surprisingly, only rarely used to mean "addicted", as in *drug dependent individuals.*

Cryptotechnical Words: AWL

Words used primarily with technical sense (19/23 = 83%):

ADMINISTRATION

Exclusively used with sense "giving a drug, especially as a remedy": *drug administration, oral administration, route of administration.*

BIND

Used without exception to mean "combine by chemical forces": *binding site(s), DNA binding, binding assays.*

CHANNEL

Refers to "ion channel", which is a pore in a cell membrane that permits the selective passage of ions: *calcium/Ca2 channels, K+ channels, channel antagonists.*

COMPOUND

Used to refer to a chemical substance consisting of two or more elements: *antioxidant compound, antibacterial compound.*

CONCENTRATION

Found with the meaning "relative amount of a substance in a chemical solution": *concentration dependent, serum concentrations, concentration response curve.*

CONTRACTION

Used to refer to the tightening of a muscle or tissue: *induced contraction(s), contraction of [tissue], maximal contraction.*

DISTRIBUTION

Found to refer to the pattern of occurrence of a substance within or between cells, tissues, or organisms: *volume of distribution, tissue distribution*; also used with the statistical meaning "relative numbers of individuals in a particular category": *population distribution.*

INHIBITION

Found with the meaning "restraining of a function of a bodily organ, process, or enzyme": *inhibition of acetylcholine release, synthase inhibition, concentration dependent inhibition, platelet inhibition.*

INTERACTION

This is used to mean the process by which two different drugs affect each other: *interaction(s) between [drugs], interaction of [drug and drug], drug interactions.*

MEDIUM

This refers to a substance that something grows in, exists in, or moves through: *calcium free medium, perfusion medium, culture medium.*

REACTION

Found to mean "bodily response to a drug stimulus": *adverse drug reactions, reaction time.*

REGULATION

Used with the meaning "control of bodily processes by drugs": *regulation of [food], cardiovascular regulation, regulated by [pharmacological mechanism], regulated pathways.*

RELAXATION

Always found to mean "lessening of tightness of a muscle or tissue": *relaxation of [muscle], induced relaxation(s), relaxation in [tissues].*

RELEASE

Used to refer to a situation in which a drug or chemical is spread into the surrounding area: *release of [drug], Ca2 release, cytokine release.*

RESPONSE

Used with the meaning "change in the activity of an organism as a result of stimulation": *response to [stimulus/drug], dose response, concentration response, inflammatory response.*

SOLUTION

Invariably used with the chemical/pharmaceutical sense of "liquid mixture": *bath solution, solution containing [chemical], salt solution.*

TARGET

Has the meaning "tissue or organ upon which a drug exerts its action": *single/multi target drug, target receptors, target cells*; also less frequently used to mean "goal" or "objective": *therapeutic target.*

TRANSPORT

Used with the meaning "movement or transfer of a biochemical substance, especially a drug, in a biological system": *transport of [drug], [drug] transporter(s), [drug] transport.*

VEHICLE

Used exclusively to mean "substance used as a medium in the administration of a drug": *treated with vehicle, drugs or vehicle controls, [drug] vehicle.*

Words used with both general and technical senses / unclear (4/23 = 17%):

POTENTIAL

The cryptotechnical meaning of this word – "voltage" – is the most frequent: *action potential, holding potential, membrane potential*; but also used with the more general

meaning of future possibility: *the potential for [side effects], the potential benefits/ problems.*

SIGNIFICANT

Mainly used with the statistical meaning of being caused by something other than chance: *significant difference, significantly lower [activity]*; also occurs with the general sense of being very large or important: *a significant improvement.*

STABLE

Mostly used to refer to a condition or illness which is not deteriorating: *chronic stable angina*; also found with the more general meaning of "unchanging": *stable perfusion pressures.*

STRESS

This word was found to be used with its general (or perhaps lay-technical) sense of psychological strain: *mental stress*; also used to mean a particular kind of force: *oxidative stress.*

The examination of GSL and AWL words assumed to be cryptotechnical has shown that the vast majority (48/51, or 94%) are indeed used in pharmacology articles with their specialized meanings, and many of them are used exclusively in this way. Often, the meaning of a word is closely related to its core general or academic sense (e.g., *delivery, regulation, release, distribution, resistant, stable*). Sometimes, though, the words are used with meanings which are quite different in pharmacology or medicine (e.g., *channel, bind, medium, vehicle*). The meaning in pharmacology is frequently an extension of the general meaning, but with a word such as *medium*, the technical sense is truly hidden. A small number of the words appear to be used equally with their general and specialized senses (e.g., *potential, order, significant*); a few, surprisingly, are not found with their expected technical meaning (e.g., *dependent*). There is also a group of words that we might label "crypto-scientific": these can be found across a wide range of disciplines, not only the medical field, and include words used in statistics such as *mean* and *population.*

There are, then, several different sub-categories of cryptotechnical words, which will probably present learners with different degrees of difficulty. It would be interesting to see whether the distance between the core and technical meaning has any bearing on the perceived difficulty of a word.

Fully Technical Words

These are the words that are used almost exclusively in pharmacological/medical contexts. Table 12 shows, perhaps contrary to expectations, that a good number of the most frequent of them will be familiar to some extent to the layperson. They are, on the whole, at least pronounceable, which should be of some reassurance to the ESP teacher who lacks expert knowledge in the subject.

Table 12. Fully Technical Words in the Top 500

Word family	Frequency	Word family	Frequency
1. RECEPTOR	1303	25. VITRO	151
2. AGONIST	353	26. ATP	149
3. NEURON	348	27. CYTOKINE	147
4. ANTAGONIST	327	28. OPIOID	147
5. PLASMA	297	29. PLATELET	140
6. OXIDATION	247	30. MACROPHAGE	140
7. ACETYLCHOLINE	245	31. HYPERTENSION	131
8. ENZYME	238	32. PHYSIOLOGICAL	129
9. ASSAY	224	33. ADENOSINE	127
10. MEMBRANE	210	34. SUBSTRATE	120
11. SYNAPSE	199	35. ISCHEMIC	119
12. KINETIC	194	36. ANTIBODY	116
13. PEPTIDE	192	37. SUBUNIT	114
14. CHLORIDE	191	38. GLUTAMATE	113
15. TRIGLYCERIDE	189	39. KINASE	113
16. VIVO	182	40. RENAL	112
17. LIGAND	181	41. NUCLEUS	110
18. PGE	175	42. PANCREAS	109
19. PH	173	43. DNA	108
20. MOLECULAR	170	44. SALINE	108
21. OXIDE	160	45. AORTIC	104
22. SERUM	156	46. RENIN	102
23. VASCULAR	152	47. ARTERIOLE	100
24. ENDOTHELIUM	151	48. HEPATIC	100

Multiword Items

In our discussion thus far we have focussed on single words, but with the advance of corpus linguistics there has been increasing awareness of the importance of multiword items (see, e.g., Gardner, 2007). We have seen that many cryptotechnical words are found in frequently occurring clusters, and if these combinations of words function in the same way as single words (as with *smooth muscle*), we ignore them at our peril.

In order to carry out a preliminary investigation into potentially salient multiword items I decided to use the Sketch Engine (available at http://www.sketchengine.co.uk/). This program is able to extract keywords and key expressions by statistical comparison of the corpus of interest with a reference corpus. The final list was compiled from the overlapping items obtained for two WebBootCaT pharmacology corpora (each comprising 250,000 words). Table 13 shows the multiword items (excluding those containing function words) that can be considered to occur with unusual frequency in the corpora.

Table 13. Multiword Items (Derived from WebBootCaT Pharmacology Corpora)

adverse effect	*low concentration*
blood-brain barrier	*low dose*
body weight	*nervous system*
clinical trial	*our study*
control group	*present study*
control study	*previous study*
coronary artery	*receptor agonist*
drug concentration	*receptor antagonist*
drug interaction	*receptor blockade*
drug reaction	*side effect*
endothelial cell	*significantly different*
et al	*smooth muscle*
ex vivo	*spinal cord*
growth factor	*standard error*
growth hormone	*statistically significant*
high concentration	*time course*
high dose	*vagus nerve*
in vivo	

Only twelve of the 35 multiword items might be judged to be functioning strictly as single units (i.e., the whole could be said to be greater than the sum of the parts), and even then the meaning is probably discernible if the words making up the unit are known:

> *et al, coronary artery, ex vivo, in vivo, growth factor, growth hormone, nervous system, side effect, smooth muscle, spinal cord, standard error, vagus nerve*

It would appear that there are fewer of these than perhaps might be expected (although of course a larger corpus, and indeed my own pharmacology corpus, might tell a different story). This does not mean, though, that we should ignore collocations; on the contrary, collocational behaviour is an important part of what it means to know a word, and may be a significant indicator of lexical technicality (Ward, 2007). It would therefore be helpful to provide learners with a glossary of the most useful words which explains their meanings and shows a) their use in context with their most common collocates, and b) the more usual meaning(s) of cryptotechnical words.

CONCLUSION

It has been shown that it is possible to create a single list of words, comparable in size to the General Service List, which provides learners of a medical discipline with a large proportion of the words they will need in their academic reading and writing. Coverage of the Pharmacology Corpus is short of the magic 95% believed to be necessary for satisfactory reading comprehension, but it is respectably close to 90%. What is noteworthy is that coverage of over 70% is given by as few as 500 word families, many of which are found in general lists or the Academic Word List. This shows that a word list can be of a much more

manageable size than 2,000 words and still be of potentially great benefit to learners. The next step could be to narrow the focus of the Pharmacology Word List still further by making wide range a requisite, and excluding function words and the general words that will be unproblematic to learners at this level.

The list also serves as a starting point for detailed investigations into the characteristics of the most useful pharmacology words. As we have seen, surprisingly few of the words in the list are the strictly technical terms that will be familiar only to the subject specialist, and many more are potentially confusing crpyptotechnical words with hidden technical meaning. Future research might take a closer look at the problems caused by polysemy, and at other factors affecting learnability such as word length, sound-script congruence, and morphology. The most frequently occurring collocations and multiword units also deserve a more thorough examination.

Most importantly, the findings of this study question the usefulness of the AWL for medical ESP learners, and suggest that it is difficult to support the view that ESP vocabulary should be separated into lists of general, academic, and technical words. Rather, there should be a move away from the general towards specialization at an early stage.

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.

Baroni, M. & Bernardini, S. 2004. BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC 2004*, 1313-1316. Lisbon: ELDA.

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

Chen, Q. & Ge, G-c. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes* 26, 502-514.

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.

Fraser, S. (2007). Providing ESP learners with the vocabulary they need: Corpora and the creation of specialized word lists. *Hiroshima Studies in Language and Language Education, 10*, 127-143.

Gardner, D. (2007). Validating the construct of *word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics 28* (2), 241-265.

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.

Hyland, K. and Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly 41*(2), 235-253.

Japan Association of College English Teachers. (2003). *JACET list of 8000 basic words.* Tokyo: JACET.

Konstantakis, N. (2006). An attempt to create a business word list. Presentation given at the 16th Vocabulary Acquisition Research Group Network Conference, Swansea.

Laufer, B. (1989). What percentage of text-lexis is necessary for comprehension? In C. Lauren & M. Nordman (Eds.) *Special language: From humans thinking to thinking machines,* 316-323. Clevedon: Multilingual Matters.

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

Meara, P. (1995). The importance of an early emphasis on L2 vocabulary. *The Language Teacher Online,* 19: 02.

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

Salager, F. (1983). The lexis of fundamental medical English: Classification framework and rhetorical function – a statistical approach. *Reading in a Foreign Language 1,* 54-66.

Scott, M. (1999). *Wordsmith Tools* (Version 3.0). Oxford: Oxford University Press.

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

Ulijin, J. M. (1984). Reading for professional purposes: psycholinguistic evidence in a cross-linguistic perspective. In J.M. Ulijin & A.K. Pugh R (Eds.) *Reading for professional purposes: Studies and practices in native and foreign languages.* Heinemann.

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

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes, 6,* 18-35.

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.

要　約

用語分類の壁を突き崩して：ESP学習者のための分野別統合ワードリストの構築

フレイザー，サイモン
広島大学外国語教育研究センター

先行研究においては，EAPやESP学習者用に，一般用語とAcademic Word List (AWL)を共に使用することで，読解力向上に有益となる専用ワードリストの作成にいくらかの成功を収めている。しかし，最頻出ワードリストとAWLの両方に，専門的な意味を隠れ持つ多義性の「Cryptotechnical Word」が多数見受けられた。そして，これらCryptotechnical Wordを，学習者は既知の単語と受け止め，内包されている専門的な意味合いに気付かない可能性が指摘される。特にAWLにおいて，学習者は必ずしも覚えておく必要のない単語学習に多大な時間を費やしてしまう可能性がある。

このような非効率的な学習法を回避し，本当に必要と思われる単語のみを決定する一つの方法として，一般，Academic，Technicalといった用語分類を避けた統合ワードリストの構築が挙げられる。これは語の頻度と範囲に基づいて作成されるものであり，本稿はこのワードリストの構築について述べている。薬理学の文献から36万語に及ぶコーパスを作成し，十分な読解のために必要とされる95%レベルをカバーしているかどうかの検証を行った。そして，本コーパスの中でCryptotechnical Wordの使用例を調べた結果，わずかに例外は見られるものの，その殆どが専門的な意味を持つ単語として使用され，その多くが，連語として高頻度で確認された。これは薬理学におけるマルチワードユニットとコロケーションの役割について考察の必要性を示すものである。