How large a Vocabulary do EAP Engineering Students need?

Jeremy Ward
Institute of Social Technology, Suranaree University of Technology, Amphoe Muang, Nakhon Ratchasima

It is commonly suggested that learners need a vocabulary of at least 3000 word families in order to begin reading efficiently, and that this vocabulary should be based initially on a general word list and then supplemented by an academic and/or technical word list. This study shows, first, that EAP engineering students with a vocabulary of only 2000 word families may have sufficient lexical knowledge to read texts; second, that there may be no need for them to start with a distinct, "general" vocabulary; and third, that using texts based on a specialist vocabulary may not overburden learners with difficult technical vocabulary. This in turn suggests that EAP reading for engineers can be undertaken at an earlier stage than it commonly is.

1. INTRODUCTION

The importance of vocabulary knowledge for effective second language reading hardly requires justification to anybody who has tried to read in a second language in which he/she is not very proficient, to anyone who, when checking text comprehension, has been asked by learners "What does this word mean?", or to anyone who has noticed the widespread learner strategy of glossing unknown words in texts (Ghadessy 1979).

Vocabulary knowledge may have received less attention during the late 1970s and 1980s, when more stress tended to be laid on top-down, schema-driven processes in reading but with the postulation of "interactive" models of reading (Carrell et al. 1988) common sense has reasserted itself (see the discussion in Grabe 1988). One claim particularly relevant to the present study is that a reader will not be able to read effectively unless he/she knows a certain percentage of the running words in a text (Liu & Nation 1985). The figure of 95% has been suggested, at which "threshold" a reader can use the reading skills he/she has acquired in the first language to read in the second language (Laufer 1988, 1992, 1997; Hirsh & Nation 1992). In a sense this is obviously an oversimplification, depending as it does on the importance, exact position, guessability and re-occurrence of the "unknown" words, but the claim is not obviously untrue. Nation (1996:11) remarks "A larger vocabulary size (sc. than 95%) is clearly better" and quotes West (1941) as recommending 98% "known" words in a reading text (Nation & Coady 1988:99). Hazenberg & Hujlstin (1993:152) express similar reservations about the 95% figure, but this study, like theirs, will take it as the benchmark.

JEREMY WARD is an assistant professor at Suranaree University of Technology, Nakhon Ratchasima, Thailand. He has also taught in Iran, Thailand and Japan, and has published two textbooks and a number of articles and reviews. His area of research interest is EAP reading for undergraduates in developing countries.

This study investigates what lexical resources are needed to achieve this 95% threshold by undergraduate engineering students reading their textbooks in English. It seems at least common sense that they need a "general" vocabulary to start with, such as West's (1953) General Service Word List, before moving on to specialist, technical or academic vocabulary (however these terms are defined). As Nation & Waring (1997:11) remark, "..Clearly the reader needs to know the 3000 or so high frequency words of the language. These are an immediate high priority and there is little sense in focusing on other vocabulary until these are well learned". Even more blithe is Jordan (1997:150), who remarks that "The vocabulary for students following EAP courses should clearly be much more advanced than the core 2000-3000 words that provide the basis of about 80% of the words likely to be encountered."

However since there is evidence (Nurweni & Read 1999) that many undergraduate engineers are some way short of this 3000-word mark; since there is a lack of evidence to show that such a general list would give the 95% coverage necessary; since there is often limited time available to facilitate student textbook reading; and since there seems on the face of it an inherent contradiction in using a general list for learners with specific purposes, it seems reasonable to ask first, if it would be possible to devise an "engineering word" list as a short cut to reading fluency, and second, if it would not be more efficient to introduce such a list to learners at the earliest possible opportunity (provided needs can be identified at an early stage).

2. METHOD

A list of the most common words in first-year engineering texts was created in the following manner.

In consultation with engineering faculty members, five extended texts were chosen from the first year courses in engineering that an engineering student at the writer's institution[1] would be required to take. These texts dealt with engineering thermodynamics, engineering mechanics, fluid mechanics, statistics & probability and mechanics of materials. The words in the texts were counted giving a corpus of approximately 1 million background engineering words. This will be referred to as "the engineering corpus". 1 million is of course the total number of running words, or tokens; there were approximately 15000 different types[2].

The words in this corpus were then grouped in word families at level 6 as described in Bauer & Nation (1993), i.e. including under the headword all inflected forms and specified derived forms but excluding classical roots and affixes and compound words. For example, the word family under the headword "USE" included the following: used / uses / using / usable / useful / usefulness / useless / users / user.

From the resulting list of word families the 3000 most common were taken to create the engineering word list. The figure 3000 was chosen for purposes of comparison with the General Service Word List (West 1953) of 2000 words, complemented by the University Word List of 836 words (Xue & Nation 1984).

This engineering list (EngList) was run against various texts to establish its predictive power - i.e what proportion of the words in the text it covered, or (as it will be termed here) predicted. Twelve 5000-word texts were selected from a variety of academic disciplines, in the following manner. An undergraduate textbook was selected in each of 12 disciplines: physics, chemistry and biology (i.e. background sciences); engineering materials, fluid mechanics and engineering mechanics (background engineering); chemical engineering, electrical engineering and mechanical engineering (specialist engineering); and economics, philosophy and psychology (humanities). A random sample of 30 pages (every 10th page starting from page 27) was taken from each book. After excluding diagrams, rubrics, footnotes, variable names, equations, proper names, units of measurement and abbreviations, the first (approximately) 5000 words of each selection were taken as the texts. This made a total corpus of 60000 running words (which I shall refer to as the interdisciplinary corpus).

3. RESULTS AND DISCUSSION

I will be referring to various lists in this paper and the reader is advised to undertake the simple task of mastering the naming system (Table 1 below) before continuing.

LIST NAME	CONTENTS
EngList	3000 most common engineering word families
EngList1	1000 most common families from EngList
EngList2	2nd most common 1000 word families in EngList
EngList3	3rd most common 1000 word families in EngList
GSL	General Service Word List (West 1953) (c. 2000 families)
GS1	1000 most common families from GSL
GS2	2nd most common 1000 families from GSL
UWL	University Word List (Xue & Nation 1984) (academic words not in GSL) (800+ families)
GSL/UWL	GSL & UWL combined

Table 1: List of lists

3.1 The predictive power of the engineering lists .

The engineering corpus contained 15000 word types; of these approximately 8000 were accounted for in the 3000 word families of EngList. However the least frequent of these 8000 - i.e. those in EngList3 - occurred less than 15 times in the engineering corpus, with those at the bottom of the list occurring only four times. EngList3 thus assumes rather less importance and it is EngList1&2 that will receive most attention here.

Table 2 below shows how much of the 1 million engineering word corpus is predicted by the EngList 1, 2 and 3 - the three lists which make up EngList. Incidentally, the relative infrequency of EngList3 items in the texts, noted above, was also reflected in the EngList3 column in Table 2.

	EngList1	EngList2	EngList 1 & 2 Total	EngList3	EngList (total)
Thermodynamics (c280000 tokens)	92.9	2.9	95.8	0.7	96.5
Fluid Mechanics (c250000 tokens)	91.7	4.2	95.9	1.1	97.0
Mechanics of Materials (c170000 tokens)	93.2	3.1	96.3	0.5	96.8
Statistics & Probability (c200000 tokens)	90.2	4.5	94.7	1.2	95.9
Vector Mechanics (c120000 tokens)	92.7	3.1	95.8	0.8	96.6

Table 2: EngList run against engineering corpus texts (i.e. the texts from which EngList was assembled) (Figures indicate percentages of tokens predicted)

It will be seen from Table 5 that EngList1 and EngList2 combined predict over 95% of tokens in all except one of the texts, in which the figure is 94.7%.

The reader should note that the figures in Table 2 are biased in favour of the predictive power of EngList, since EngList is itself derived from a corpus based on these particular texts (with diagrams, rubrics, footnotes, variable names, equations, proper names, units of measurement and abbreviations eliminated). Table 2 does however show, at least, that it is possible to predict 95% of the tokens in a large amount of text - the 1 million word engineering corpus - with a list of 2000 word families (EngList 1 & 2). The obvious question - whether EngList 1 & 2 would predict other engineering texts as well, cannot be answered definitively here. For example, the reader may have noticed that the 1-million word engineering corpus contains no text in engineering materials, computer programming or electrical engineering - three sub-disciplines that might well form part of a basic engineering curriculum - and surmise that EngList would not predict as well in these sub-disciplines. On the other hand, perhaps EngList would do so if it were based on a corpus that included them. Section 3.2 below, which gives the results when EngList is run against the interdisciplinary corpus, gives us some idea, but since the interdisciplinary corpus is much smaller then the engineering corpus the results cannot be regarded as having the same weight.

When EngList was run against a different 100,000+ word engineering mechanics text, i.e. one which was not part of the engineering corpus, it proved to be an excellent predictor (engineering mechanics being one of the sub-disciplines in the engineering corpus); see Table 3 below.

	EngList1	EngList2	EngList 1 & 2 Total	EngList3	EngList (total)
Mechanics text (c116000 tokens)	92.6	4.3	96.9	0.9	97.8

Table 3: EngList run against engineering mechanics text (116000 words) which was NOT in the original engineering corpus (Figures indicate percentages of tokens predicted)

To return to the present EngList1&2; the figure of 2000 word families is considerably lower than the 3000 suggested by Laufer (1997) and the 5000 suggested by Nation & Hirsh (1992) for general fiction, and spectacularly lower than the 11000 suggested by Hazenberg and Hujlstin (1993) for undergraduates in universities in Holland, although the latter use a more restricted meaning of "word" and appear to be dealing with students who have a much wider range of English reading requirements.

The contents of EngList1 & 2 proved remarkably different from those of GSL; 50% of the word types in EngList 1&2 (approximately 3000 out of 6000 types) do not occur in GSL.

3.2 The coverage of the engineering list over the interdisciplinary corpus

With the use of Vocab Profile software from Victoria University of Wellington in New Zealand, the texts in the 60000-word interdisciplinary corpus - the 12 different subject areas - were analysed to show the percentage of tokens (running words) that were predicted by the GSL, GSL/UWL and EngList respectively. The results may be seen in tables 4 & 5 below.

	Gs1	Gs2	UWL	Total
biology	71.3	5.2	9.4	85.9
chemistry	71.6	4.5	12.6	88.7
physics	76.7	5.0	10.9	92.5
engineering mats	68.0	8.0	10.4	86.4
engineering mech	76.8	6.2	9.0	92.0
fluid mechanics	72.8	7.1	12.2	92.1
mechanical eng	65.1	7.3	7.9	80.3
chemical eng	74.3	5.8	12.4	92.5
electrical eng	77.4	4.3	12.4	94.1
economics	77.3	5.5	9.9	92.6
philosophy	83.8	3.4	7.8	95.0
psychology	74.6	6.3	10.2	91.0

Table 4: Percentage of tokens in interdisciplinary corpus (12 x 5000) predicted by GSL/UWL

Incidentally, the figures in Table 4 confirm two claims made by Nation (1996) - that the GSL predicts approximately 80% of academic text (the mean coverage of GSL in Table 1 is 79.8%) and also that the UWL will predict another 10% (the mean coverage in Table 1 is 10.4%). However, GSL prediction is considerably better in the humanities (mean 83.6%) than in the scientific/technical disciplines (mean 78.5%).

	EngList1	EngList2	EngList 1 & 2 Total	EngList3	EngList (total)
biology	72.3	7.2	79.5	2.6	82.1
chemistry	81.4	7.8	89.2	2.5	91.7
physics	88.5	5.8	94.3	1.5	95.8
engineering mats	81.9	7.5	89.4	2.3	91.7
engineering mech	92.9	4.5	97.4	1.2	98.6
fluid mechanics	93.7	3.7	97.4	1.4	98.8
mechanical eng	78.8	5.4	84.2	1.4	85.6
chemical eng	85.8	5.9	91.7	3.5	95.2
electrical eng	90.4	6.3	96.7	1.1	97.8
economics	75.1	11.0	86.1	4.1	90.2
philosophy	78.1	9.2	87.3	2.3	89.6
psychology	76.0	8.4	84.4	3.0	87.4

Table 5: Percentage of tokens in interdisciplinary corpus (12 x 5000) predicted by EngList

The figures in Table 5 give some support to the idea of "academic English"; the only thing that all the texts have in common with EngList but not with GSL is that they come from university textbooks, i.e. are academic.

As a comparison of tables 4 and 5 shows (summarised in table 6 below), EngList 1&2 combined predict better in all disciplines than GSL1 & 2. While the difference is not enormous in the humanities, it certainly is in the scientific/technical disciplines[3] (see the two right-hand columns of Table 6 below).

	GSL/UWL	GSL	EngList1&2 total
biology	85.9	76.5	79.5
chemistry	88.7	76.1	89.2
physics	92.5	81.7	94.3
engineering mats	86.4	76.0	89.4
engineering mech	92.0	83.0	97.4
fluid mechanics	92.1	79.9	97.4
mechanical eng	80.3	71.4	84.2
chemical eng	92.5	80.1	91.7
electrical eng	94.1	81.7	96.7
economics	92.6	82.8	86.1
philosophy	95.0	87.2	87.3
psychology	91.0	80.9	84.4

Table 6: Relative predictive power in interdisciplinary corpus (12 x 5000) of EngList1&2 combined, GSL and GSL/UWL (Figures indicate percentage of tokens predicted)

UWL was of course created to remedy precisely this deficiency of GSL; with UWL added, this "general" list becomes a better predictor (than EngList1&2) in biology, all the humanities and (mystifyingly) chemical engineering.

However, as the reader will note, EngList1&2 combined predict at 95% or more in no less than three engineering disciplines - engineering mechanics, fluid mechanics and electrical engineering. Physics comes very close (94.3%). This suggests that the predictive power of EngList 1&2 is not limited to the disciplines from which the engineering corpus texts were taken. Discussion of why EngList 1&2 predict less well in the other 5000-word engineering texts will appear in section 3.3 below.

I would like therefore to suggest that reading courses should be based on a lexical syllabus such as EngList 1 & 2, where these conditions apply:

1. Students - secondary or tertiary - are or will be specialising in engineering
2. English-language engineering textbook reading is the primary aim
3. Exposure to English-language materials is limited, either because of lack of learner motivation, lack of materials, or lack of time.

In fact it would appear that such conditions are far from rare (see any number of studies from Cooper 1984, Hassan et al 1986 through Cohen 1988 to Nurwani & Read 1999)

Point 3 is crucial because, as Laufer (1997) points out, the vocabulary that students need to read effectively is a sight vocabulary, an ability to process ("knowledge of") words without giving them focal attention. This "knowledge" comes with repeated exposure and given the (point 3) constraints, will inevitably be limited in terms of breadth (how many words it consists of) ; the suggestion here is that it therefore needs to be developed in terms of a small and specific vocabulary (i.e. EngList 1&2). If, however, high-school students - even those who have chosen technical specialisations - are using general-orientation English language textbooks[4], then, in so far as they are building up a sight vocabulary at all (cf. Nurwani & Read op.cit.) it is of a general nature. What is more, when they begin their university courses in English their studies may well have a wide EAP focus that will contain much that is irrelevant to their engineering needs (Sutarsyah et al. 1994).

An impression of the irrelevance of much of GSL can be gained from the 5000 word mechanical engineering text. This was relatively poorly predicted by both GSL/UWL and EngList; although we should bear in mind that this is a "specialised" text for 3rd/4th year students, whereas EngList is based on a background, non-specialised engineering corpus.

branches caving die dirt dragging east fastened flag frozen further ladder lady laid laying lifts mid north parked railroad remind repair rescuing sights soon stolen storm surfaced toe town village waist winding

Table 7: Mechanical Engineering (5000 word text): words predicted by GSL/UWL but not by EngList1&2

The reader will notice the absence of any function (closed-system, grammatical) words from this list. Obviously the need for exposure discussed above applies strongly to function words since these are the most general words of all (in the sense of having the most general, abstract meanings as well as being the most common): but there were no function words at all in GSL, or for that matter in GSL and UWL combined, that were not also present in EngList 1&2. (the converse is also true). It is only with content (open-system) words that the differences arise.

Compare Table 7 with the words that only EngList1&2 predicted:

algebraic analytical **axial** *bracket brakes buoyant cable* **cantilever** *capsule centerline* **centroid** *chord compression concrete concurrent considerable constants cord* **cosine** *couple crane crate dam dashed deduced delete diagonal downward encountered flexible furthermore graphical hence hint horizontal idealized impending infinitesimal infinity intercept interface labelled links meter negligible orthogonal parabola parallel parallelogram patrol pitch pivot plastic platform polygon* **prismatic** *pulley quantitatively* **radially** *rear reference resultant sag* **scalar** *sectional settings shaft shear sin sine singular span static submarine submerged substitute subtract symmetry tan tend tending tensile termed thrust* **torque** **transversely** *trigonometry truck* **truss** *unknowns unstretched variables* **vector** *vehicle whereas windmill*

Table 8 : Mechanical Engineering (5000 word text): words predicted by EngList1 & 2 but not by GSL/ UWL (bold face words are "technical", see above)

The reader cannot get an exact picture from tables 7 & 8 because of differences in the word families that appear in EngList and GSL/UWL (for example *analytical* does not appear in UWL but *analytically* does: *setting* appears in UWL but *settings* does not[5]). But it does appear that it is EngList's technical orientation that brings an advantage in predictive power (as seen in Table 6).

Can it be objected, on this basis, that EngList is too technical in nature for high school, or "foundation-year" university students? The word technical here is taken to mean depending on specialist knowledge of subject matter, here of course in the applied sciences.

Table 8 above is, as noted, from a more advanced text. But even with such a text, if we ask whether these words are"....readily accessible through their use outside the field" and thus "...less obviously technical", following Nation 1993) the answer is in most cases yes. (As a technologically challenged non-scientist, this writer would need explanations of only 11 of the 96 word types listed above i.e. those in bold type). Of course everybody's choice of what is technical in Table 8 will be different, but I have tried not to flatter my own knowledge. Table 8 does not depend largely on a knowledge of specialist concepts or subject matter. EngList1&2 are not very technical. EngList1 & 2 predict better than GSL a lot of lexis which (while technical in the sense of related to science) is readily accessible to the lay reader.

3.3 The words which EngList 1 & 2 do not predict

What would the gaps be in an EngList1&2- trained student's vocabulary? He/she would, for example, be some distance from the 95% threshold with the 5000-word engineering materials text and with the 5000-word mechanical engineering text (Table 6).

What are these words that EngList fails to predict?

Table 9 below show lists of all the unpredicted words occurring 5 times or more in three science/technology-related 5000-word texts which were poorly predicted (relatively speaking) by EngList 1 & 2.

ENGINEERING MATERIALS	MECHANICAL ENGINEERING	CHEMISTRY
eutectic	*austenite*	*acid*
lattice	*bainite*	*calcium*
precipitate	*bronze*	*chloride*
covalent	*carbide*	*ions*
nucleation	*brasses*	*magnesium*
martensite	*cementite*	*ray*
quantum	*chromium*	*aqueous*
silicon	*dislocation*	*cation*
zinc	*embrittlement*	*excited*
valence	*ferrite*	*hydrochloric*
pearlite	*further*	*ion*
magnesium	*graphite*	*many-electron*
bainite	*gray*	*neutron*
chloride	*high-speed*	*nucleus*
nickel	*low-carbon*	*orbit*
ferrite	*martensite*	*paramagnetic*
crystal	*microstructure*	*phosphate*
austenite	*niobium*	*phosphoric*
strengthening	*nucleate*	*potassium*
nonequilibrium	*pearlite*	*proton*
homogenization	*plain-carbon*	*quantum*
cementite	*precipitate*	*radioactive*
sodium	*quench*	*reactive*
coexist	*stainless*	*stair*
	temper	*strontium*
	water-hardening	*subshell*
		unpaired

Table 9: Most common words in three 5000 word texts not predicted by EngList 1 & 2

Since the words in Table 9 account for about half the unpredicted tokens in the texts, we may claim that a substantial number of the words that EngList fails to predict are technical words, in the sense of depending on knowledge of the subject matter or in the sense of being related to science/ technology. This suggests strongly a), that our EngList 1 & 2 -trained student would not gain an advantage from knowing

GSL, and b) that the words he/she does not know will, being technical, be explained in the course of the chapter where they occur.

These claims can be reinforced in another way: if we take every 10th word from the complete list of all the unpredicted words in the three texts it will be seen that many of these are guessable. The first 16 words from each sample are shown in Table 10 below.

ENGINEERING MATERIALS	MECHANICAL ENGINEERING	CHEMISTRY
carbides	*austempering*	*artificial*
compacted	*billets*	*illness*
crystals	*cementite*	*penetrate*
discrepancy	*cold-drawn*	*solute*
exposure	*crack-tip*	*accident*
heat-affected	*deep-drawing*	*atone*
integrity	*detrimental*	*boron*
low-carbon	*drastically*	*chlorate*
metastable	*eutectic*	*darkened*
noncoherent	*fine-grain*	*donates*
plain-carbon	*grain-boundary*	*evolution*
propylene	*healed*	*fluorescence*
segregation	*high-temperature*	*hydrofluoric*
solidus	*impede*	*krypton*
three-dimensional	*inhomogeneous*	*manganese-four*
volume-to-surface	*irons*	*neon*

Table 10: Random sample of words not predicted by EngList 1 & 2 in three 5000-word texts

The reader will note the predominance of technical and compound words, the latter often made up of single words from EngList 1 & 2 and so guessable.

EngList 1 & 2 are technical in a sense but the words they do not predict are technical in a more specialist, subject specific way. If then we accept this "technicality" then the probability exists that words which are not in EngList1 & 2 are likely to be explained as part of the specialist subject study, as the subject matter demands it. This claim is perhaps reinforced by the fact that these unpredicted technical words tend to occur within a limited area of the text - they tend to be clustered. Here are the five words that happen to occur ten times in the engineering materials text:

quantum / many / heat / form / calculate

Of these *quantum* is the only clearly technical term. All the occurrences of *quantum* are on the same page, i.e. they occur within a narrow band of the text (6% of the total running text, actually). Compare this with the EngList-predicted "general" words *calculate* (28% band, various pages), *heat* (59%, various), *form* (81%, various), and *many* (96%, various).

Similarly, taking the 7 words which occur 16 times, the respective areas of the text which include these words are:

eutectic 9%: electrons 12%: composition 8%: each 92%: has 73%: must 72%: size 83%

If it is true that technical words are clustered, it may well be because they are new concepts that are being introduced and explained.

Our EngList1&2- trained student, therefore, will not be greatly hindered at the beginning-engineers' level either by specialist technical vocabulary or by general words not predicted by EngList1&2.

CONCLUSION

A first-year engineering student may know 95% of the tokens in many basic engineering texts with a vocabulary of only 2000 word families. This vocabulary will clearly have a technical flavour but will contain all the general words (including all function words) necessary, and deal with concepts that are in general accessible to the non-expert. The lexis that students may not know will be largely technical and contain a number of compound words; much of it will be explained in lectures or in the text. Therefore for students with the limited aim of reading their engineering textbooks in English it makes sense to start them as soon as possible on material based on lists like EngList 1 & 2. Much greater efficiency will be provided where purposes are specific and homogeneous; and if this leads to better understanding, then that in turn will have payoffs as learner attention is turned away from irrelevant lexis towards understanding the relationships between the content words and the meanings of whole sentences. If you like, they will have more of their limited attention available for interpreting function words.

Acknowledgements: the author would like to thank Dr. Paul Nation of Victoria University of Wellington for his unfailing assistance in the preparation of this article, and Suranaree University of Technology for funding the project. The author can be contacted at: jeremy@ccs.sut.ac.th

1 Suranaree University of Technology, Nakhon Ratchasima, Thailand.

2 To illustrate the meaning of tokens and types: "*to be or not to be*" contains six tokens and four types (*to*, *be*, or and *not*).

3 With the exception of biology, whose descriptive, taxonomic nature perhaps sets it apart from the applied sciences and perhaps from the natural ones too.

4 this is certainly the case in Thai state high schools, for example (Nawm, personal communication)

5 Nation (personal communication) points out that some of these may be caused by oversights in making the families in the GSL.

REFERENCES

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

Cohen, A., Glasman, H., Rosenbaum-Cohen, P., Ferrara, J., & Fine, J. (1988) Reading English for specialized purposes: discourse analysis and the use of student informants. In Carrell P., Devine J. & Eskey D. (eds) *Interactive Approaches to Second Language Reading*. Cambridge: CUP

Cooper M. (1984) Linguistic competence of practised and unpractised non-native readers of English. In Alderson, J. Charles & Urquhart, A.H. (eds) *Reading in a Foreign Language*. London: Longman

Coxhead A., & Nation P. (1998) The specialised vocabulary of English for academic purposes (unpublished ms., Victoria University of Wellington)

Cowan, J.R. (1974) Lexical and syntactic research for the design of EFL reading materials. *Tesol Quarterly* 8(4) 389-399.

Ghadessy, M. (1979) Frequency Counts, word lists and materials preparation. *English Teaching Forum* 17(1) 24-27

Grabe, William (1988) Reassessing the term "interactive" in P. Carrell, J. Devine & D. Eskey (eds.) *Interactive Approaches to Second Language Reading*. Cambridge: CUP.

Hassan M., Neilson P., & Thomas J. (1986) Course for First Year Students of the College of Earth Sciences. In D. Harper (ed) *ESP for the University* (ELT Documents 123). London: Longman.

Hazenberg, Suzanne & Hulstijn, Jan H. 1996: Defining a minimal receptive second language vocabulary for non-native university students: an empirical investigation. *Applied Linguistics* (17)2, 145-163.

Hatch, E & Brown, C (1995) *Vocabulary, semantics and Language Education*. Cambridge: CUP.

Hirsh, D. & Nation, ISP (1992) What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8(2), 689-96.

Jordan, R.R. (1997) *English for Academic Purposes*. Cambridge: CUP.

Liu, N & Nation, ISP (1985) Factors affecting guessing vocabulary in context. *RELC Journal* 16 (1) 33-42.

Laufer, B (1988) What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordmann (Eds.) *Special Language: From Humans Thinking to Thinking Machines*. Clevedon: Multilingual Matters.

Laufer, B. (1992) How much lexis is necessary for reading comprehension? in P. Arnaud & H. Bejoint (eds) *Vocabulary and Applied Linguistics*. London: Macmillan.

Laufer, B. (1997) The Lexical Plight in Second Language Reading in Coady J & Huckin T (eds.)(1997) *Second Language Vocabulary Acquisition*. Cambridge: CUP.

Nation, P & Waring, R (1997) Vocabulary size, text coverage and word lists in Schmitt N & McCarthy M, *Vocabulary: Description, Acquisition and Pedagogy*, 6-19. Cambridge: CUP.

Nation, P & Coady, J (1988) Vocabulary and Reading in Ronald Carter & Michael McCarthy, *Vocabulary and Language Teaching*. London: Longman.

Nation, P. (1993) Vocabulary size, growth and use in R.Schreuder & B. Weltens (eds.) *The Bilingual Lexicon*. Amsterdam/Philadelphia: Benjamins.

Nurweni, A. & Read, J. (1999) The English language knowledge of Indonesian university students. *English for Specific Purposes*, 18(2), 161-175.

Sutarsyah, C, Nation, P, & Kennedy, G (1994) How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal* 25(2), 34-50.

Trimble, L (1985) *English for Science and technology: a discourse approach*. Cambridge: CUP

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