RELC Journal 27,2 (1996) 1-11

USING TEXTS TO SEQUENCE THE INTRODUCTION OF NEW VOCABULARY IN AN EAP COURSE

DENICE WORTHINGTON AND PAUL NATION
Victoria University of Wellington
New Zealand

Abstract

This paper examines the feasibility of allowing the texts that are used in a course to sequence the target vocabulary of a course, in this paper the vocabulary of academic study as represented by the University Word List. It was found that such an approach would only allow the learners to meet a little over half of the 836 word vocabulary and that a three-step sequencing procedure would be needed to effectively meet all of the wanted vocabulary. The three steps would involve (1) using adapted texts to gradually introduce the very common academic words, (2) using unsimplified texts to meet the next 200-300 words, and (3) relying on extensive reading and planned decontextualised learning to meet the remaining lower frequency items.

High Frequency Vocabulary and Academic Vocabulary

There has long been recognition of the need to give attention to the vocabulary component of language courses. Most often this attention has been given to the presentation of vocabulary, that is, the techniques, activities and strategies that are used to teach it. As a result of the early work published by Thorndike & Lorge (1944) and West (1953) there has been increasing awareness of the need to give principled attention to the selection of the vocabulary that should go into a course. This awareness has mainly focused on the distinction between high frequency vocabulary and low frequency vocabulary (Nation, 1990). Although the dividing line between high frequency, wide range vocabulary and low frequency, narrow range vocabulary is largely arbitrarily drawn, there are striking differences in the return to the learner for the effort of learning these two types. The high frequency vocabulary is usually defined as a group of around 2,000 word families (West, 1953) (A word family consists of a base word, for example, *dig*, and closely related inflected and derived forms, *digs*, *dug*, *digging*, and *digger*). The West list is now very old, but recent research (Hwang and Nation, 1995) has confirmed the suitability of its size and the adequacy of the words it contains. The Hwang and Nation study used data from the LOB (Johansson, Leech and Goodluck, 1978) and Brown (Francis and Kucera, 1979) corpora

to make a new high frequency vocabulary, and this was compared with the West list. The study looked at frequency (how often the words occurred) and range (how many different sections of the corpus they occurred in). Although frequency and range are closely related it is important in a study of general service vocabulary to exclude words that have a very high frequency in one text or in a group of similar texts but which do not occur elsewhere. That is, narrow range words. The study confirmed that a general purpose high frequency vocabulary should contain around 2,000 words, and found that the majority of words in the West list were also high frequency and wide range words in the LOB and Brown lists. Replacing some of the words in the West list changed the text coverage by around 1%. Clearly the West list is still a suitable basis for course design, and the data it contains on the relative frequency of word meanings is not matched by any currently available list in spite of the large computer corpora available. The difficulty in replacing the West list is that the semantic count has to be done manually. On most written texts the 2,000 word high frequency vocabulary accounts for around 80% of the total running words (tokens) of the texts. This has been confirmed in numerous studies (Hwang and Nation, 1989; Sutarsyah, Nation and Kennedy, 1994; Engels, 1968). The West list gives even higher coverage (around 90%) of texts written for young native speakers of English (Hirsh and Nation, 1992). With informal spoken language, the coverage of running words may be closer to 95% (Schonell, Meddleton & Shaw, 1956).

For the teacher and course designer, there is great value in knowing which are the high frequency words, because this allows them to give these words the attention they deserve in the design and teaching of a course. This value has also had an effect on research for courses for English for Academic Purposes (EAP). Just as there is a high frequency general service vocabulary worth focusing on in basic courses, there is also a vocabulary that builds on this which has high or at least moderate frequency and wide range in the limited field of academic texts. This could be a very useful, deliberate focus in EAP courses. This vocabulary of 836 word families includes items like *component, define, equivalent, impact and publish,* which occur across a wide range of academic texts, and provides around 8.5% coverage of the running words (Xue & Nation, 1984). When this is added to the 80% coverage of the 2,000 high frequency word families, the resulting 89% coverage of academic texts makes an invaluable contribution to helping learners cope with these texts. The UWL (University Word List) is classified into 11 groups according to the frequency and range of the items (Nation, 1990: 235-239). The value of the UWL as a vocabulary to provide access to academic writing

has been examined in studies of diverse collections of academic texts as well as in studies of specialised corpora (Sutarsyah, 1993; Hwang, 1989).

We have looked at the selection of vocabulary. How can vocabulary be graded or sequenced so that learners are meeting manageable chunks at appropriate intervals in the course?

Sequencing Academic Vocabulary

Seeing vocabulary in its natural contexts of use provides useful information for learners, particularly regarding collocation and shades of meaning. This does not mean however that it is necessary to rely solely on context to teach vocabulary. A very clear finding from the large amount of research on decontextualized vocabulary learning is that the initial learning of very large amounts of vocabulary can be usefully done in this way in a very short time (Nation, 1982). This means that learners can quickly develop foreign language - first language connections for particular words which will then be enriched by meeting and using them in a variety of contexts.

There are several ways of sequencing the vocabulary in a course.

1 Learners can learn words from lists or exercises based on lists and then hope to meet them in texts. The UWL is divided into subgroups based on frequency and range which provides a rationale for sequencing the learning. The books of vocabulary exercises by Farid (1985) and Yorkey (1981) could be used in this way. These books use word building exercises using Greek and Latin word parts and multiple choice completion activities to help learners gain control of academic vocabulary.

2 The vocabulary could be formed into topic groups and special texts could be written to ensure the occurrence of the words. Exercises to prepare for and establish the vocabulary could also be added. This is the approach taken by Barnard (1980) in the series *Advanced English Vocabulary.*

3 Learners can meet the words as they naturally occur in texts and then work further on establishing and expanding this knowledge by directly learning the words using word cards and dictionary study.

These three approaches to deciding how to group and sequence the vocabulary each have their own difficulties. When the sequencing is based on lists ordered according to frequency and range, we cannot be sure that

the learners will soon meet the words again in texts after they have worked on the words. In the second approach where texts need to be specially prepared to provide a context for the words, there is the very strong risk of interference between the words because synonyms, opposites, free associates (Higa, 1963) and lexical sets (Tinkham, 1993) occur in the same lesson. This interference considerably increases the difficulty of learning these words as they tend to get confused with each other. In addition there may be considerable difficulty in constructing a realistic text. The third approach of dealing with words as they normally occur in texts is thus very attractive. However, this approach gives rise to several questions regarding the sequencing of vocabulary. How many texts are needed to cover all or a reasonable amount of the UWL? Will the amount of words from the UWL not previously met be roughly equal and be not too many or too few for each new text in a series?

This research looks at the answers to these two questions. The questions can be put another way which does not tie them to any particular approach to learning the UWL vocabulary. That is, is it feasible to use a series of academic texts to determine what and how much UWL vocabulary the learners meet at each stage of their EAP course? That is, does a representative series of academic texts cover most of the UWL vocabulary and does it do this in a regular gradual way?

The assumption behind this research is that the course will be designed to provide plenty of opportunities for helping learn the vocabulary. This research simply looks at the first occurrence of items assuming that this occurrence will be the excuse for spending further time on these particular items. The main focus of this article is to see what kind of planning needs to go into an EAP course to ensure that the necessary UWL vocabulary is covered. In general each word will be learned in a gradual fashion with each meeting of the word adding to the existing knowledge (Nagy, Herman and Anderson, 1985). Well planned courses must provide opportunities for several meetings with important vocabulary.

Procedure

Texts taken from the learned and scientific sections of the Lancaster-Oslo-Bergen (LOB), the Brown and the Wellington (Bauer, 1993) corpora were used in this study. These three corpora each consist of approximately 1,000,000 running words made up of five hundred texts which are each 2,000

words long. Each corpus is divided into subsections including general fiction, reportage, and learned and scientific writing which includes university texts and other academic writing. The learned and scientific section is subdivided into various disciplines, including political science, social science, natural science, technical and engineering, and humanities.

Table 1 shows the discipline areas of the texts from the learned and scientific sections of the three corpora that were used in this study. The four different disciplines were used to make sure that the occurrence of the UWL vocabulary was not the result of the particular discipline examined.

First, the occurrence of the UWL words in the combined 134 texts, each around 2,000 words long, was studied. The aim of this part of the study was to see how many texts were needed before every word in the UWL was met at least once, and to see if the words from the UWL not met in previous texts appeared at a constant rate. These 134 texts were made up of 282,646 running words. They included 796 (95%) of the 836 word families of the UWL.

Second, it was decided that it would be easily feasible to cover twelve texts in a course lasting a few weeks. So, ten different sets of twelve texts were analysed to see if the coverage of the words in the UWL was affected by the discipline studied and by the corpus they came from.

Table 1: The texts used in the study

Discipline	Brown corpus	LOB corpus	NZ corpus	Total texts
Pol. Science	15	14	15	44
Soc. Science	14	14	14	42
Tech. & Engin.	12	12	12	36
Nat. Science	-	-	12	12
	41	40	53	134

Two-thousand word texts were used because they were the text units of the corpora used in this study. For an EAP course a 2,000 word text for intensive reading is long, although it closely resembles the target reading

that the learners will have to do. If such texts were used in a course, each text would need to be worked with several different times, perhaps over the course of a week.

Results

Table 2 is based on a study of ten series of twelve texts drawn from those listed in Table 1. Table 2 shows that clearly it is not feasible to expect to meet all or most of the words in the UWL in a series of texts in an EAP course of modest length, say 12 weeks.

Table 2: Cumulative totals of previously unmet UWL word families in 10 series of 12 texts

Tex	1	2	3	4	5	6	7	8	9	10	11	12
Pol.Bro	64	145	190	255	280	315	353	385	403	418	431	453
Pol.Bro	87	162	190	231	279	305	330	352	369	391	401	417
Pol.NZ	51	92	141	164	195	248	291	325	342	353	368	385
Soc.Bro	125	199	265	303	353	374	395	410	423	443	453	458
Soc.Lob	95	139	184	209	233	295	318	339	366	380	395	411
Soc.NZ	66	125	171	203	227	261	284	312	340	350	357	370
Tec.Bro	56	103	164	253	289	310	334	350	364	387	393	406
Tec.Lob	73	132	170	197	227	261	260	274	312	334	347	363
Tec.NZ	59	112	152	211	241	269	299	318	334	364	372	387
Nat.NZ	66	140	165	190	224	243	256	273	290	301	323	332
Cumulative	74.2	134.9	179.5	221.6	254.8	285.7	312.0	333.8	354.3	372.1	384.0	398.2
Total	22.4	30.7	33.7	39.9	845.7	742.4	43.0	43.6	39.8	41.1	39.0	39.3
sd												

Table 2 shows that in the Pol.Bro texts (political science texts taken from the Brown corpus), 64 UWL words occurred in the first text in the series. The words in the first text combined with the previously unmet words in the first text meant that 145 different UWL word families were met in the two texts.

Table 3, which is based on the same texts as Table 2, shows that there is an initially rapid and then steady decline in the number of previously unmet UWL words in a series of texts. The number of UWL words starts out high with around 74 UWL word families occurring in the first text, and 60 UWL families which were not in the first text occurring in the second text in the series. These numbers would be overwhelming for learners with no previous knowledge of the UWL. After 5 or 6 texts the number of previously unmet words drops to a more manageable level.

Table 3: Average number of previously unmet UWL words in 10 series of 12 texts with standard deviations

Text	1	2	3	4	5	6	7	8	9	10	11	12
New UWL words	74.2	60.7	44.6	42.1	33.2	30.9	26.3	21.8	20.5	17.8	11.9	14.2
sd	22.4	14.5	12.7	21.6	9.3	16.0	8.7	7.2	7.9	6.6	4.7	4.6

Almost 200 of the total 836 UWL words are met in the first three texts and 300 in the first six texts. The number of new UWL words met in the 12th text is well under half those met in the sixth text and one-fifth of those met in the second text.

The reason for this patterning is that the UWL, like any frequency based list, is made up of words that differ in their frequency and range. A relatively small group of words will occur very frequently in most kinds of texts. In the UWL this includes words like area, require, analyse and process. There will be a larger group of words that typically occur less frequently and with a narrower range of occurrence. While we can be sure that learners will meet most of the high frequency group if they continue to study and use the language, the learners would have to work through a large number of varied texts to meet most of the other words in the list. This effect is most striking in the first 2,000 words of English with the most frequent 1,000 words covering around 77% to 80% of most texts and the second 1,000 words covering only 5% to 8% of most texts. The parts of the UWL also behave in this way, with the words from sub-list 1 occurring more frequently and thus providing much more coverage than the words from sub-list 10.

The size of the standard deviations in Table 3 for the occurrence of previously unmet words shows that there is wide variation between individual texts in the number of UWL families they contain. The number of UWL word families in the first text in each of the ten series studied ranged from 51 to 125 (see Table 2).

The study of the 134 texts showed that by the 12th text almost half of the total UWL words were met. By the 45th text close to 700 words (83%) of the UWL were met. The additional coverage from studying more texts was very small, with the last 12 texts (Texts 122-134) adding only eight additional UWL words.

There are thus several difficulties involved in using the natural occurrence of vocabulary in texts to determine the quantity and sequencing of vocabulary.

1 An impossibly large number of texts would be needed to cover all of the vocabulary of the UWL. If texts were used as a means of sequencing vocabulary, it would be possible to do this for only a part, say 50%, of the UWL. Other ways of meeting the remaining vocabulary would have to be devised. These might include adaptation of texts, learning from lists, using specially prepared exercises, or simply leaving it somewhat to chance by encouraging extensive reading.

2 A very large amount of unfamiliar UWL vocabulary is met in the first three or four texts. This is far too much to be usefully dealt with in a few lessons and so there would need to be vocabulary learning preparation before meeting these texts.

A Proposal for Sequencing Academic Vocabulary

It is clear from the difficulties involved in using texts to sequence the introduction of vocabulary that there would need to be a three step approach to sequencing.

1 First, the learners would need a gradual introduction over about 5 texts to the high frequency, wide range 100 to 200 items in the UWL. This could be done by judicious selection or partial simplification of academic texts. The glossing could be done outside the text by the addition of glossaries at the side of the page or at the end of the text (Jacobs, Dufon and Fong, 1994), or a form of elaboration could be used where the words are explained in the text itself (Long and Ross, 1993) The partial

simplification would involve the replacement or glossing of UWL words not in the first 100-200 items in addition to the replacement of some of the words not in the first 2,000 and UWL. At this step the sequencing is based on frequency and range.

2 Then, about 12 or more unadapted texts could be used to cover a further 200 or 300 items resulting in coverage of about half of the UWL. At this step the occurrence of vocabulary in texts determines the sequencing of the vocabulary.

3 Then, because the unknown vocabulary load of the texts would not be so heavy, learners could be encouraged to do large amounts of extensive reading of academic texts, both within their subject areas and outside these areas. This could be accompanied by decontextualised learning of UWL words, and study through formal exercises such as those involving word parts. At this step both frequency and range information, and occurrence in texts are used independently of each other to determine the sequencing of the items to be learned.

The assumption behind this piece of research has been that the occurrence of vocabulary in texts is the initial opportunity to meet the words which would then need to be learned to some degree so that they were not unknown items when met in subsequent texts. It is not sufficient to assume that simply meeting the items in a text would be enough to ensure learning. This meeting would have to be accompanied or followed up by intensive study and opportunity for use, so that the knowledge of each item of vocabulary would be cumulatively enriched.

References

Barnard, H. (1980). Advanced English Vocabulary. Rowley: Newbury House.

Bauer, L. (1993). Manual of information to accompany the Wellington corpus of written New Zealand English. Department of Linguistics, Victoria University of Wellington.

Farid, A. (1985). *A vocabulary workbook.* Englewood Cliffs: Prentice Hall.

Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior 2,* 170-175.

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

Hwang K. (1989). Reading newspapers for the improvement of vocabulary and reading skills. Unpublished MA thesis, English Language Institute, Victoria University of Wellington.

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

Jacobs, G.M., Dufon, P., and Fong Cheng Hong. (1994). L1 and L2 vocabulary glosses in L2 reading passages: their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading.* 17, 1, 19-28.

Long, M. and Ross, S. (1993). Modifications that preserve language and content. In M.Tickoo (ed.) *Simplification: Theory and Application.* RELC Anthology Series No. 31: 29-52.

Nagy, W.E., Herman, P., and Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly,* 20, 233-253.

Nation, I.S.P. (1982). Beginning to learn foreign vocabulary: a review of the research. *RELC Journal,* 13, 14-36.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* New York: Heinle and Heinle.

Schonell, F.J., Meddleton, I.G. & Shaw, B.A. (1956). *A study of the oral vocabulary of adults.* Brisbane: University of Queensland Press.

Sutarsyah, C. (1993). The vocabulary of Economics and academic English. Unpublished MA thesis, English Language Institute, Victoria University of Wellington.

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

Tinkham, T. (1993.) The effect of semantic clustering on the learning of second language vocabulary. *System* 21, 371-380.

Thorndike, E. & Lorge, I. (1944). *The teacher's word book of 30,000 words.* Columbia University: Teachers College.

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

IN PRAISE OF LINGUISTIC PROBLEM-SOLVING

JAMES M BOURKE
Universiti of Brunei Darussalam

Abstract

The article looks at problem-solving as a means of internalising second language grammar. Problem-solving in this context refers to a type of learning that is conscious, inductive, and heuristic. In the problem-solving framework, as outlined in this paper, grammar learning is thought of in terms of two complementary processes, viz. 'rule-getting' and 'rule-using'.

The problem-solving procedure commences with problem-posing, and proceeds through data analysis, hypothesis forming and testing, and frequent rule re-structuring, to the final rule formulation. This is followed by 'rule-using', a dynamic process in which rules are constantly being reformulated and form-function correlations are being worked out.

Contrary to Sheen (1992), who dismisses problem-solving as "a seductive hypothesis", the author argues that linguistic problem-solving should play a key role in second language learning, especially in the case adult L2 learners.)

Linguistic Problem-solving:

The use of problem-solving as a communicative interactive technique in second language learning has long been advocated by writers such as Prabhu (1987: 137) , Sadow (1983: 115-120), and Klippel (1984: 102-114). This type of task-based problem-solving has more recently been elaborated by Skehan (1993), and Estaire and Zanon (1994). Communicative problem-solving tasks of the following type are commonplace.

Example:
Problem: You and six other teachers have been asked to form a committee to organise an essay-writing competition for secondary schools in your region.

At your first meeting agree on:
1) objectives
2) categories
3) length of essay
4) theme