

VOCABULARY OVERLAP IN DIFFERENT NEWSPAPER ACCOUNTS OF THE SAME EVENT¹

Usana Sitthipol

DTEC Language Institute

Department of Technical and Economic Co-operation

Bangkok

INTRODUCTION

This paper describes the nature of the vocabulary overlap that is likely to occur when two newspaper accounts of the same event or topic are used for intensive reading practice in a foreign language learning classroom. Of all reading materials available to students, newspapers are of particular interest for several reasons. First, newspapers provide useful information that can be exploited by communication tasks. Secondly, newspapers give students extensive access to low frequency words and their associated grammar. Thirdly, newspapers are a source of alternative accounts of the same event, thereby increasing the likelihood that learners will have an opportunity to process second presentations of low frequency words in similar sentence contexts.

Two or more texts on the same topic can overlap in different ways. The texts may overlap quantitatively in terms of the propositions they share. One indication of this overlap is the number of low frequency words common to both texts. In addition, texts may overlap qualitatively in terms of the logical relationship between the information in each text. For example, the second text may restate, clarify, elaborate or exemplify information contained in the first text, a text may present contrary information to that presented in another, or it may be part of a running story.

Running stories are reports of an event as it unfolds over consecutive days. When Hwang (1989) compared the lexical frequency profile of four items in a series, he found that up to half of the low frequency word families in the first report were repeated three times or more in the subsequent reports, thereby reducing the vocabulary load vis-à-vis unrelated reports. The present study predicts that learners would be exposed to even greater repetition if their reading included parallel accounts of the same event from different newspaper sources.

When a new word is met a second time, it is remembered better if it is encountered in the same linguistic context as in the first exposure to the word. (See Besson and Kutas for a review of the experimental literature.) It is also remembered better if the modality remains the same, as in reading versus listening, and if the learner is engaged in the same task, as in recognizing the word during reading versus producing the word in spoken interaction.

Besson and Kutas discuss the reasons for a context similarity effect in learning new vocabulary. First, when a target word is encountered in the same context they claim that a 'core' meaning of the word is built up by the learners. Different sentence contexts, on the other hand, lead to the encoding of different aspects of a word's meaning. Although the range of a new word may be increased by different contexts, it is not strictly speaking a repetition of the first encounter. Secondly, Besson and Kutas

claim that it is easier to integrate a target word into a previously understood context than to integrate a word into an entirely new context.

Does the placement of a word in a text affect the likelihood that learners will remember the word after they have read the text? It has recently been shown by McKoon, Ratcliff and Ward (1993) that concepts placed in syntactically prominent positions are remembered better after reading in both the short and the longer term. A particularly prominent position is within the subject noun phrase of a main clause.

In summary, teachers of reading are beginning to use more than one textual source of information for a given topic to help students to read in a way that matches academic reading in the real world. Newspapers are a convenient source of such texts but to date lexical frequency information has been gathered only for unrelated texts and running stories. An important advantage of using two newspaper accounts of the same event may lie in the additional exposure that students have to low frequency words. The number of exposures needed to develop the core meaning of a word is high, between 7 and 10 repetitions (Hwang, 1989). Learning efficiency, in the sense of needing fewer repetitions to acquire a word, may be achieved if the context of subsequent presentations of a word remains essentially the same, and if the word is encountered in a position of prominence within the sentence syntax.

METHODOLOGY

Two newspapers published on the same day were used to select stories that were essentially about the same event or topic. For pedagogical purposes, answers to the following questions were sought:

- 1 Are low frequency words that are likely to be a target for a vocabulary development programme repeated in the two accounts?
- 2 Are the words repeated in similar or somewhat different contexts?
- 3 Do the syntactic structures used in the various presentations of a target word affect the relative prominence given to the word?
- 4 Does reading a second account of the same event bring a student into contact with other members of the same word family?

The actual texts chosen for study are described in Table 1.

Table 1 Texts selected for lexical analysis

News Section	<i>Bangkok Post</i>	<i>The Evening Post</i>
International news	7,000 killed as quake strikes Western India 1.10.93 (539 words)	Quake turns Indian region into graveyard 1.10.93 (717 words)
Business	Murdoch: Float of Ansett a possibility 13.10.93 (325 words)	News Corp hints at Ansett float 14.10.93 (364 words)
Sports	Hughes goal warning after Wednesday demolition 4.10.93 (958 words)	Hughes boosts United's title bid 4.10.93 (466 words)
Entertainment	Genie grants Disney a box-office wish 21.10.93 (472 words)	When cartoons aren't funny 16.10.93 (620 words)

Each account of a news item was entered into a computer by manual typing. Two computer programmes, FVORDS and the Oxford Concordance Programme 2, were then used to produce a lexical profile of each text, and to list the word families that co-occurred in pairs of texts.² Table 2 gives an example of the lexical frequency coverage of the text about the earthquake in India (*Bangkok Post*, 1 October 1993).

Table 2 An example of the lexical coverage of an international news item

Base list	Tokens (%)		Word families (%)	
One	380	(70)	25	(9)
Two	34	(6)	3	(12)
Three	48	(8)	54	(20)
Not in the lists	77	(14)	154	(57)
Total	539	(100)	267	(100)

- 2 List two: the new second 1,000 word list (Hwang, 1989)
- 3 List three: 1,000 words of academic English not in the first 2,000 words of general English (Xue & Nation, 1984).

Words outside of the three base lists include low frequency words and proper nouns.

RESULTS

Vocabulary profiles

Using the FVORDS programme, a lexical frequency profile (Laufer & Nation, 1994) was calculated for each text in order to show the proportion of vocabulary items used at each frequency level. Table 3 compares mean token numbers and standard deviations for the articles from the *Bangkok Post* with those from the New Zealand paper. The figures in brackets refer to the cumulative coverage of a text by the successive introduction of each base list.

Taken as a set, the articles from the *Bangkok Post* and *The Evening Post* have similar lexical frequency profiles. In fact the first 1,000 words accounted for about 70% of the words in each of the eight texts. The new second 1,000 words provided only a small additional coverage with about 5% of each text using words from this list. The 1,000 most frequently occurring words in school and university texts were sampled reasonably extensively in most items of news (on average 9%; SD 2.83). Words not present in the base lists accounted for on average 14% of each text. Just over half of those words were proper nouns associated with the topic of the texts. Between 90 and 95% of each text is covered by the most frequent 2,000 words of general English, the 1,000 most frequent words of academic English, and a knowledge of place names and other types of proper nouns.

Vocabulary overlap

The primary purpose of the project was to ascertain the likely overlap in alternative accounts of the same news item taken from different newspapers. Word families, not tokens, were taken as the unit of analysis.

Table 3 The relative vocabulary richness of articles drawn from two daily newspapers: mean percentages, (cumulative percentages), and standard deviations

Newspaper		1,000 Word Frequency	2,000 Word Frequency	Academic Vocabulary	Prope Nouns	Low Vocabulary
<i>Bangkok Post</i>	M	72	5	8	8	6
(%)			(77)	(85)	(93)	(100)
SD		3.70	0.82	1.71	2.22	2.75
<i>The Evening Post</i>	M	70	5	10	7	7
(%)			(75)	(85)	(92)	(100)
SD		1.30	1.26	3.70	2.07	2.63

A word family was defined as the base form of a word plus its inflected and derived forms (Hirsh & Nation, 1992). The use of a word family as the unit of analysis assumes that a person who knows the base form of a word will also know other members of the word family if he or she is familiar with the inflectional and affix systems of English.

Table 4 reports the number of word families shared between each pair of texts for different categories of low frequency words. The numbers in brackets refer to the extent to which repeated words of a particular kind are a percentage of the low frequency words of that kind in each text. For example, in the last column of Table 4, 41 low frequency words in the *Bangkok Post* version of the earthquake in India were repeated in *The Evening Post* account, either in the same form or as an inflectional or derivational form. The 41 repeated words represent just under half (47%) of the low frequency words in the *Bangkok Post* account, and a third (34%) of the low frequency words in *The Evening Post* version.

Based on the information in Table 4, the following observations can be made:

- 1 There was a greater repetition of low frequency word families for texts that described the same event than was found in the pair of texts on the same topic.
- 2 There is a tendency for the words from the academic list to be repeated more often than proper nouns or low frequency words, but this effect may be specific to the topic of the text. For example, the sports reports had a large number of repetitions of proper nouns and terms that were clearly specific to soccer.
- 3 In all, when the same events are described, half or more of the vocabulary outside the first two thousand words was repeated in the second text.

Table 4 Number of low frequency words in each text from the Bangkok Post that are repeated in a parallel text from The Evening Post (relative coverage of low frequency categories expressed as a percentage)

Section	Academic Vocabulary	Proper Nouns	Low Frequency Vocabulary	Total
International	19 (56) (46)	10 (53) (27)	12 (34) (30)	41 (47) (34)
Business	23 (88) (70)	11 (85) (69)	4 (67) (50)	38 (84) (67)
Sports	16 (39) (84)	33 (70) (85)	21 (58) (78)	70 (56) (82)
Entertainment	8 (26) (20)	2 (6) (12)	2 (6) (5)	12 (12) (12)

4 If articles of a markedly different length are read, as in the sports reports, there were repetition advantages in reading the short text first as that ensured the greatest number of repetitions.

Density of new words

In a tertiary EFL reading class, the word families in Base List 3 are in need of repetition because they represent the key words in academic English. The low frequency words in List 4 are likely to be ‘problem’ words so far as the reader is concerned and are the words that a student may very well look up in a dictionary. Hwang and Nation (1989) viewed the repetition effect of List 3 words in terms of the reduction in new word density in the second text. By their argument, a reduction in new word density should allow a reader to pay attention to List 4 words.

Table 5 shows the number of new word families from the academic word list that the students will encounter in each news account of the same event. It also shows how many new word families students will be exposed to in the second text. The percentages in brackets indicate the new word density when a student moves from the first account to the second.

Two observations can be made from Table 5:

- 1 When reading a second account, there was a reduction in the density of words from the academic word list in three of the four samples of news. For some accounts the reduction in density was considerable.
- 2 The reduction in density did not hold for texts on the cartoon topic, suggesting that the use of alternative accounts will not always reduce the reading burden.

Table 5 The density of word families outside the first 2,000 words in different news accounts of the same event

Section	Newspaper	Length of the news account (word families)	Total number of new word families ¹	Word families which are new to learn
International	<i>Bangkok Post</i>	267	34	34 (13%)
	<i>The Evening Post</i>	334	41	22 (7%)
Business	<i>Bangkok Post</i>	168	26	26 (15%)
	<i>The Evening Post</i>	186	33	10 (5%)
Sports	<i>Bangkok Post</i>	365	41	41 (11%)
	<i>The Evening Post</i>	239	19	3 (1%)
Entertainment	<i>Bangkok Post</i>	265	31	31 (12%)
	<i>The Evening Post</i>	294	48	40 (14%)

Same versus different sentence context

To what extent were low frequency words repeated in similar contexts versus dissimilar contexts in the second account of a news item? In order to answer this question, a teacher selected a sample of 58 words beyond the 2,000 level that had been repeated at least once in a parallel account.

Table 6 reports the results of a card sort of the contexts in which each target word appeared. The judgements made were between identical contexts, similar contexts, somewhat different contexts, and very different contexts.

Table 6 The extent to which the target words occur in similar contexts in both newspaper accounts

Section	Identical Contexts (%)	Similar Contexts (%)	Somewhat Different Contexts (%)	Very Different Contexts (%)
International news (n = 18)		37	16	48
Business (n = 21)	100			
Sports (n = 12)	100			
Entertainment (n = 7)				100

It would seem from Table 6 that one cannot assume that a word will necessarily be used in the same or similar context in the second account. For example, a student would have met target words in the same context for the business and sports items, but in very different contexts in the two accounts of the use of cartooning in the film industry. The international news item, however, would have exposed students to the use of target words in both similar and different contexts in a one in three ratio respectively.

Syntactic prominence

The relative prominence of 229 low frequency words in the sentence structure of the texts was examined in two ways: by recording the word class of each usage of the target word and, secondly, by recording whether the target item was part of a main clause, a subordinate clause or a prepositional phrase.

Table 7 summarizes the word class of the various instances of the target words in the texts as a whole.

Table 7 The word class of instances of the target words in the various newspaper accounts

Newspaper	Noun	Verb	Modifier	Total
<i>Bangkok Post</i>	70	12	27	109
<i>The Evening Post</i>	75	10	35	120
Total	144	23	62	229

Table 7 shows that approximately two out of every three instances of the target word families were nouns. Target words were less likely to be met as modifiers and verbs. In constructing the table it became apparent that the target words in the second text usually occurred in the same class as in the first account. A reader would meet a target word in another word class only to a limited extent.

Table 8 reports the kind of sentence structure that embedded the various instances of target words.

Table 8 The relative saliency of instances of target words in the sentence structure of each text: data recorded as frequencies and (percentages)

Newspaper	Main Clause	Subordinate Clause	Prepositional Phrase	Total
<i>Bangkok Post</i>	41(38)	30(28)	38(35)	109(100)
<i>The Evening Post</i>	52(43)	30(25)	38(32)	120(100)
Total	93(41)	60(26)	76(33)	229(100)

In the texts summarized in Table 8, the target words occurred in a range of positions in the sentence syntax. About four times in ten, the words occurred in a main clause. The remaining occurrences were in positions of less prominence, a subordinate clause or more frequently a prepositional phrase. Moreover, when a word was used in a main clause in one text it was frequently encountered in a less prominent position in the second text, and vice versa.

DISCUSSION

When two texts of the same news event or topic are used, there is considerable overlap in the vocabulary beyond the 2,000 word frequency level, thereby reducing the lexical density of the second text. The extent to which vocabulary repetitions occur in similar contexts and in prominent places in the syntax depends on the specific texts. It is possible to find texts where key vocabulary items are used in highly similar contexts and in main clauses. Other articles, however, may repeat words in less prominent positions in the syntax and/or in quite dissimilar contexts.

Based on the likely repetition of low frequency words, it is recommended that English classes have access to more than one newspaper for reading purposes. The difficulty in using newspaper items,

however, must not be overlooked. Students would need a 2,000 word vocabulary of general English, and be familiar with an additional 1,000 words of academic English, before the texts could be read at an instructional level. Even then, between 5 and 10% of the vocabulary used may be unknown to the learner.

Pedagogically, several implications can be drawn from the study:

- 1 Teachers should provide a range of newspaper items that are likely to be interesting to the students and helpful in a cultural and linguistic sense. To assist text comprehension, some form of vocabulary help may be required, such as margin notes and glossaries.
- 2 The texts should be supported by whole language activities that link skills and provide the opportunity for students to use English in pair work arrangements. Comprehension tasks should exploit the activities that become possible when two or more sources of information are used. Such tasks should require students to read across passage boundaries and use low frequency words in interaction with a partner.
- 3 The natural repetition of target words needs to be supported by the application of vocabulary development strategies such as using peer problem-solving techniques to guess unknown words from contexts (Nation & Coady, 1988).
- 4 Vocabulary learning goals should be identified by the students and progress towards achieving the goals monitored. Part of the monitoring process could include the use of a self-questioning checklist to draw attention to the context and syntax associated with new words.

CONCLUSION

The aim of this study was to assess the likely vocabulary input advantages of using parallel news accounts from a local and overseas English language newspaper. Newspapers are relatively cheap, easily available, and a potential source of short, interesting texts for advanced students in an English reading class. However, the vocabulary burden of using such texts needs to be described so that the texts can be matched with the language proficiency level of the students. At the same time, the reduction in new word density achieved by reading a second account of the same event needs to be realized by language learners and teachers.

From a survey of how the *Bangkok Post* and a New Zealand newspaper reported four news items, it was shown that the lexical frequency profile of the paper published in Thailand for the international community did not differ markedly from a newspaper published in an English-speaking country. A comparison of the word lists at different frequency levels showed that there were repetition advantages to be had in using a second account of the same event. Moreover, when a sample of key words was traced in pairs of texts, there was a tendency for the key words to be repeated in basically similar contexts, thereby enhancing their learnability.

NOTES

- ¹ This article is based on a research project completed as part of an MA in Applied Linguistics at Victoria University of Wellington
- ² FVORDS is a programme devised by Hwang Kyongho and Paul Nation (English Language Institute, Victoria University of Wellington, New Zealand) and written by Alex Heatley and Alex Ivopol of the Computing Services Centre (see Nation, 1993).
- ³ Total number of word families outside the first 2,000 words

REFERENCES

- Besson, M. and Kutas, M. 1993 'The many facets of repetition: a cued-recall and event-related potential analysis of repeating words in same versus different sentence contexts' Journal of Experimental Psychology: Learning, Memory, and Cognition 19/5: 1115-1133
- Carter, R. and McCarthy, M. 1988 Vocabulary and Language Teaching London and New York: Longman
- Hirsh, D. 1993 The Vocabulary Demands and Vocabulary Learning Opportunities in Short Novels MA thesis. Victoria University of Wellington.
- Hirsh, D. and Nation, I. S. P. 1992 'What vocabulary size is needed to read unsimplified texts for pleasure' Reading in a Foreign Language 8/2: 689-696
- Hwang, K. 1989 Reading Newspapers for the Importance of Vocabulary and Reading Skills MA thesis. Victoria University of Wellington.
- Hwang, K. and Nation, I. S. P. 1989 'Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers' Reading in a Foreign Language 6/1: 323-335
- Laufer, B. 1986 'What percentage of text-lexis is essential for comprehension?' LSP Symposium Vaasa
- Laufer, B. and Nation, I. S. P. 1994 'Lexical richness in L₂ written production: can it be measured?' Unpublished paper. Victoria University of Wellington.
- McKoon, G., Ratcliff, R. and Ward, G. 1993 'Syntactic prominence effects on discourse processes' Journal of Memory and Language 32: 593-607
- Nation, I. S. P. and Coady, J. 1988 'Vocabulary and Reading' In Carter, R. and McCarthy, M. (eds.) Vocabulary and Language Teaching London and New York: Longman
- Nation, P. 1993 'FVORDS, WORD and ZERO Programs for PCs using MS-DOS' Mimeograph. Victoria University of Wellington.
- Xue, G. and Nation, I. S. P. 1984 'A university word list' Language Learning and Communication 3/2: 215-229