『中国地区英語教育学会研究紀要』 No.34 (2004)

Absolute Method for Optimal Selection and Arrangement of the Texts in Textbook Design

Tadashi Nishida
Hiroshima University
Song Fuyun
IDEC Hiroshima University

1. Introduction

A goal of every textbook designer is to design the best (optimum) textbooks to facilitate students' learning of English. For years, research has been conducted on ways to develop teaching materials which facilitate language acquisition. Traditionally, textbook designers relied on intuition about language learning when they set out to write textbooks. They had to extrapolate which texts should be selected and arranged in the textbooks from the materials they could find, or the language they heard and read. How to select the words for a text and order the texts for a textbook mainly depended on the designers' subjective preference or assessment. As a result, the validity of the subjective extrapolation is difficult to measure and remains a problem in textbook design.

In recent years one of the most significant developments in materials design has been the great improvement of corpus. This development enables us to obtain reliable quantitative data on the frequency and range of occurrence for vocabulary and pattern examples from the corpus texts. How the language is being used can be handled and analyzed effortlessly. Up to now much research on corpus has focused on the words and their relationships with each other in context. Main achievements have been counting word frequency, presenting collocation, making word lists, and producing dictionaries. A minimum of literature, however, exists on how to exploit corpora in textbook design (Collins, 2000).

This study is an effort to systematize the selection and ordering of texts for producing textbooks so as to facilitate students' vocabulary access. A new efficient method, absolute method, is proposed to select and order the texts optimally from a corpus. To verify that the proposed method would select and arrange the texts in an optimum order, a computer program, *Corpus-Based English Textbook Design System* (*CBETDS*) (Song, 2002) is applied to select and order fifteen texts from the sixty texts of *College English* (Zhai et al., 1999). The algorithms are set up to carry out the calculation of evaluation values in a number of mathematical procedures by computer to solve the problem in the traditional design of textbooks.

2. Absolute Method

This study proposes the absolute method to select and order texts from a corpus. It uses *CBETDS* to calculate the evaluation values to select and order the texts for the design of textbooks in an optimum way. The term 'absolute' in this study means that an amount is expressed as a fixed quantity rather than referring to other factors. That is the evaluation values of the words in each text are calculated independently without considering the development of students' vocabulary learning, which is different from the relative method (Nishida & Song, 2003).

Weights to measure the degree of word difficulty for the four word lists are set. Expected values to obtain the closest evaluation values from the observed values are defined. Constraint conditions to eliminate non-optimal observed values are given.

2.1 Optimal Process

The optimization of design is a fundamental objective of virtually every designer who strives to create a system to meet a need of different levels of students. Taking the advantages of electronic storage and analysis, and the immense speed and accuracy of computer, the absolute method selects the texts from a corpus and arranges the selected texts for the textbooks in an optimum way. In the algorithms the evaluation values are calculated for the optimal approaches. The procedures to implement the optimization process are presented in Figure 1.

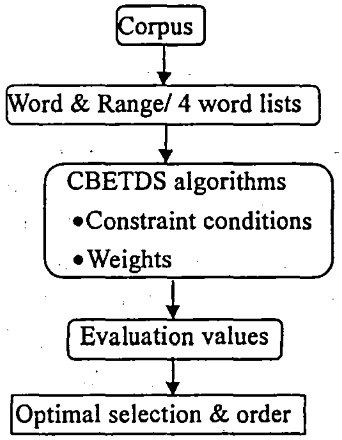

Figure 1. Optimization process

Quite different from the traditional way of designing a textbook, the optimization algorithms work on a corpus, namely, a population of candidate authentic texts. Each candidate text is run by the optimization algorithms in computer. The candidate texts have to compete to contribute to the optimization process. If the evaluation value of a text is closest to the expected value, the text will be selected into the textbook.

2.2 Evaluation Criteria – Four Word Lists

In the process of optimization, a measure of system effectiveness is needed, which often involves value judgment. The present study defines different values for different levels of words. It employs the following four word lists as criteria to calculate the evaluation values of the words in the corpus of the 60 texts in the absolute method. Word List 1 and Word List 2 include the most frequent first 1000 words, and the second most frequent words in West's (1953) *General Service List of English Words*. Word List 3 contains 570 additional words in Coxhead's *Academic Word List* (1998). Word List 4 comprises the words that do not appear in Word Lists 1, 2, and 3.

Four word lists are formed for each of the 60 texts after each text is run by *Word & Range* program (Heatley & Nation, 1996). Then how many of the words belong to Word Lists 1, 2, 3, or 4 would be given. These lists are used to evaluate the difficulties of texts by introducing the weights of the four word lists. The difficulty degrees of each text are calculated through the algorithms.

3. Algorithms for the Calculations in *CBETDS*

As vocabulary acquisition is an incremental process, it is essential in selecting the words for the textbooks, to move through the way of introducing new words gradually. This study views vocabulary building as one influential factor to arrange texts for textbooks in an optimal way. From the viewpoint of optimization theory, this study proposes an optimal method to design textbook based on corpus by processing the words in the texts absolutely. In vocabulary learning, the importance of words of different levels is different from one another. It is known that the more difficult or new words included in a text are, the more difficult the text is. Thus, it is obvious that the easier texts should be arranged first in a textbook. Since the purpose of this study is to arrange the texts for textbooks in an optimum order, it is necessary to define the order function for texts. In this study, the weights are introduced to define the degrees of word difficulty in each text.

In order to evaluate the difficulties of texts, the evaluation criterion is set for the absolute method. The evaluation function for j-th text is defined as:

$$E_j = \sum_{i=1}^{4} W_i N_i^j \qquad (1)$$

where

$W_i (i = 1,2,3,4)$ are the weights of Word Lists 1,2,3, and 4 respectively.

$N_i^j (i = 1,2,3,4)$ are the number of words for j-th text belonging to Word Lists 1,2,3 and 4 respectively.

After evaluating all the texts by Formula (1), the evaluation values for all the texts can be obtained and seen in the plotted black circles in Figure 2. Then, it is necessary to determine which texts are to be selected for a certain textbook. Obviously, for different purposes and different levels of students, the selections of texts would be different. The expected arrangement of the texts for a textbook would be the one with the evaluation values as shown by the line in Figure 2. The line represents the expected evaluation values for the texts of the textbook. The expected evaluation function for the k-th text in new text order can be defined as:

$$Ex_k = N_0 + k \times \delta \qquad (2)$$

where

N_0 : the basic word requirement for evaluation value

δ : the increase of the evaluation value between the sequent texts

k : the new text order

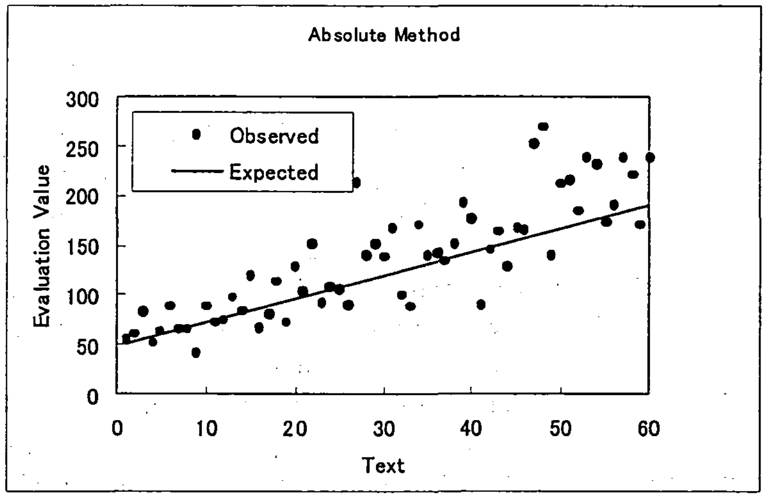

Figure 2. Absolute method

When arranging the texts in a new order, it is important to find the text whose evaluation function is the closest to the expected value under the constraint conditions. As illustrated in Figure 2, the objective of optimization is to find the texts closest to the expected evaluation line. The upward tendency line shows the accumulated tendency of the words for the texts in the textbook.

In order to select the k-th text in new text order from the texts in corpus, the formula to minimize the difference between the observed value and expected value is given in the following formulas.

$$\underset{j=1}{\overset{m}{Min}} | E_j - Ex_k | \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where

E_j : the evaluation value for the j-th text of the texts in corpus

Ex_k : the expected evaluation value for the k-th text in new text order

Constraint conditions are defined in order to give a controlling amount to the selection of texts for a textbook. Since the words belonging to Word List 4 are considered as the most difficult words among the four word lists, the constraint conditions in the absolute method is the percentage of the words belonging to Word List 4. If this percentage is large, the text is considered to be difficult. Therefore, the texts with the high percentage of the words belonging to Word List 4 will not be selected as a text for the textbook. The constraint conditions for the percentage can be written as:

$$\phi_4^j \leq \overline{\phi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

where

ϕ_4^j : percentage of the words belonging to Word List 4 for the j-th text in corpus

$\bar{\phi}$: the upper boundary of the percentage of the words belonging to Word List 4

The optimization calculation can be summarized as below:

$$\underset{j=1}{\overset{m}{Min}} | E_j - Ex_k | \quad \text{......} \quad (3)$$

subject to

$$\phi_4^j \le \bar{\phi} \quad \text{......} \quad (4)$$

where

$$E_j = \sum_{i=1}^{4} W_i N_i^j \quad \text{......} \quad (1)$$

$$Ex_k = N_0 + k \times \delta \quad \text{......} \quad (2)$$

4. Numerical Examples for the Absolute Method

Sixty texts in *College English* are stored as a small corpus in computer. The absolute method is adapted to select 15 texts for a textbook from the text corpus.

The weights of the 4 word lists in Formula (1) are defined in this numerical example as follows:

$W_1 = 0 \quad W_2 = 1 \quad W_3 = 1 \quad W_4 = 1$

Therefore, as a result, in this calculation, Formula (1) becomes the sum of the number of words belonging to Word List 2, 3, and 4 as shown in the following:

$E_j = N_2 + N_3 + N_4$

For the expected evaluation function defined by Formula (2), the basic word requirement and the increase of the evaluation value between the sequent texts are defined in the numerical example as:

$N_0 = 40$

$\delta = 10$

The upper boundary of the percentage of the words belonging to Word List 4 in Formula (4) is set to be 20% ($\bar{\phi} = 20\%$).

By using *Word and Range*, the number of the words belonging to Word Lists 1-4 for the 60 texts and their percentage are calculated, and the results are given in Table 1. The evaluation values for the 60 texts are shown in the column of E_j. The selected text number by the absolute method and the corresponding original number are also given in Table 1.

Table 1. Optimized order for the 15 texts in the absolute method

WL1	WL2	WL3	WL4	Ej	Old No.	WL1%	WL23%	WL4%	New No.
205	18	14	11	43	9	82.7	12.9	4.4	
211	22	9	20	51	4	80.5	11.8	7.6	1
185	25	19	11	55	1	77.1	18.3	4.6	
231	25	4	33	62	2	78.8	9.9	11.3	2
239	32	1	31	64	5	78.9	10.9	10.2	
246	39	7	20	66	7	78.8	14.7	6.4	
256	34	2	30	66	8	79.5	11.2	9.3	
209	30	13	25	68	16	75.5	15.5	9.0	3
207	20	10	43	73	19	73.9	10.7	15.4	
193	25	11	38	74	11	72.3	13.5	14.2	
203	27	5	43	75	12	73.0	11.5	15.5	
260	29	8	44	81	17	76.2	10.8	12.9	4
225	47	5	30	82	3	73.3	16.9	9.8	
215	27	13	44	84	14	71.9	13.3	14.7	
229	29	1	58	88	10	72.2	9.4	18.3	
243	37	16	36	89	6	73.2	15.9	10.8	
236	21	19	49	89	33	72.6	12.3	15.1	
283	22	22	47	91	26	75.7	11.8	12.6	
254	26	28	37	91	41	73.6	15.6	10.7	5
253	44	12	37	93	23	73.1	16.2	10.7	
292	36	22	41	99	13	74.7	14.8	10.5	
265	36	31	33	100	32	72.6	18.4	9.0	6
275	44	22	37	103	21	72.8	17.4	9.8	
292	39	8	59	106	25	73.4	11.8	14.8	
332	55	18	34	107	24	75.6	16.6	7.7	
266	37	22	54	113	18	70.2	15.6	14.2	7
239	51	23	45	119	15	66.8	20.6	12.6	8
287	48	2	78	128	44	69.2	12.1	18.8	
321	52	18	59	129	20	71.3	15.6	13.1	9
344	57	10	68	135	37	71.8	14.0	14.2	
261	36	44	58	138	30	65.4	20.0	14.5	
298	38	43	59	140	28	68.0	18.5	13.5	
240	44	36	60	140	35	63.2	21.1	15.8	
368	43	37	60	140	49	72.4	15.8	11.8	10
348	49	38	55	142	36	71.0	17.8	11.2	
314	31	28	88	147	42	68.1	12.8	19.1	
290	40	20	91	151	29	65.8	13.6	20.6	
252	35	40	76	151	38	62.5	18.6	18.9	11
342	58	9	85	152	22	69.2	13.5	17.2	
338	37	52	75	164	43	67.3	17.8	14.9	12
444	74	17	74	165	46	72.9	15.0	12.2	
344	59	23	86	168	31	67.2	16.0	16.8	
317	40	60	68	168	45	65.4	20.6	14.0	13
321	49	17	105	171	34	65.2	13.5	21.3	
329	57	19	96	172	59	65.7	15.2	19.2	
373	51	56	66	173	55	68.3	19.6	12.1	
339	53	55	68	176	40	65.8	21.0	13.2	
355	71	15	98	184	52	65.9	16.0	18.2	14
355	52	32	107	191	56	65.0	15.4	19.6	15
314	47	26	120	193	39	61.9	14.4	23.7	
313	48	28	136	212	50	59.6	14.4	25.9	
424	81	16	116	213	27	66.6	15.2	18.2	
311	45	117	53	215	51	59.1	30.8	10.1	
388	47	56	119	222	58	63.6	16.9	19.5	
412	82	30	118	230	54	64.2	17.5	18.4	
321	47	77	114	238	53	57.4	22.2	20.4	
361	53	71	114	238	60	60.3	20.7	19.0	
444	73	44	122	239	57	65.0	17.1	17.9	
391	97	17	137	251	47	60.9	17.7	21.3	
361	59	57	153	269	48	57.3	18.4	24.3	

WL1: Number of words belonging to Word List 1
WL2: Number of words belonging to Word List 2
WL3: Number of words belonging to Word List 3
WL4: Number of words belonging to Word List 4
Ej: Evaluation value, namely Number of words belonging to Word List 2, 3, & 4
Old No.: Original text order of the 60 texts
WL1%: Percentage of word number belonging to Word List 1
WL23%: Percentage of word number belonging to Word List 2 + Word List 3
WL4%: Percentage of word number belonging to Word List 4
New No.: New text order according to the absolute method

All the vocabulary information about the selected texts by the absolute method and the expected evaluation values are summarized in Table 2 and Figure 3. Table 2 indicates that the evaluation values of the selected texts by the proposed absolute method are quite close to the expected evaluation values. In fact, in calculation, we always select the texts whose evaluation values are closest to the expected evaluation values under the constraint conditions by optimum calculation. Obviously, the closer the observed evaluation value is to the expected one, the bigger the value is.

Table 2. Optimized and expected text order in the absolute method

New No.	WL1	WL2	WL3	WL4	Ej	Old No.	WL1%	WL23%	WL4%	Expt Ej
1	211	22	9	20	51	4	80.5	11.8	7.6	50
2	231	25	4	33	62	2	78.8	9.9	11.3	60
3	209	30	13	25	68	16	75.5	15.5	9.0	70
4	260	29	8	44	81	17	76.2	10.8	12.9	80
5	254	26	28	37	91	41	73.6	15.6	10.7	90
6	265	36	31	33	100	32	72.6	18.4	9.0	100
7	266	37	22	54	113	18	70.2	15.6	14.2	110
8	239	51	23	45	119	15	66.8	20.6	12.6	120
9	321	52	18	59	129	20	71.3	15.6	13.1	130
10	368	43	37	60	140	49	72.4	15.8	11.8	140
11	252	35	40	76	151	38	62.5	18.6	18.9	150
12	338	37	52	75	164	43	67.3	17.8	14.9	160
13	317	40	60	68	168	45	65.4	20.6	14.0	170
14	355	71	15	98	184	52	65.9	16.0	18.2	180
15	355	52	32	107	191	56	65.0	15.4	19.6	190

Expt Ej is the expected evaluation value. See Notes for Table 1.

In these numerical examples, the 15 texts are selected from the 60 texts as the texts for a textbook. It is assumed that 10 new words belonging to Word Lists 2, 3 and 4 are increased to each of the sequent fifteen texts. If the number of words belonging to Word List 4 for a text is too large, the constraint conditions would constrain that this text would not be selected because this means that the text is too difficult. In the optimization calculation, the constraint conditions for the upper boundary of the percentage of words belonging to Word List 4 are set to be 20%. In Table 1, the expected evaluation for Text 13 of the new order is 170. The closest text is Text 34 of the original order in the 60 texts, whose E_j is 171. However, the percentage of the words belonging to Word List 4 is 21.3%, which is larger than 20%. Therefore, Text 45 of the original order is selected as Text 13 of the new order.

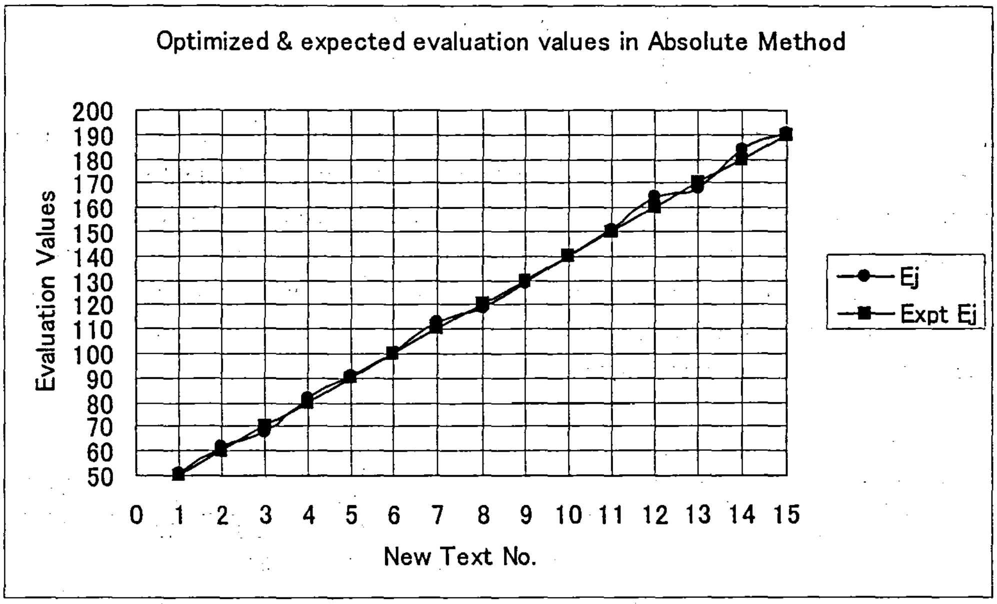

Figure 3. Optimized and expected evaluation values in the absolute method

The evaluation values of the selected texts by the absolute method and the expected evaluation values are shown in Figure 3. The line with square marks is the expected one, which is straight. The curve with circle marks is the optimized curve. This curve is the closest to the expected one. From these two curves it can be known that the proposed method is effective because the evaluation value of the texts optimally selected are quite closer to the expected ones. If there are enough texts, the two curves in Figure 3 will become the same curve.

Table 2 and Figure 3 show that the optimized order whose evaluation values are closest to the expected values increases the new words text by text in a smooth upward tendency. It is beneficial and efficient for students' access to the new words and texts in the textbooks. The results of the optimization calculation can be given in a few minutes by computer. One of the important advantages is the quick selection of the tests can be obtained simultaneously while the text order is being optimized. Figure 4 also shows that the optimized order is quite close to the expected value. However, the original order makes some big bumps between the texts. Thus, the proposed method is effective in selecting and ordering the texts in textbook design for facilitating students' vocabulary access.

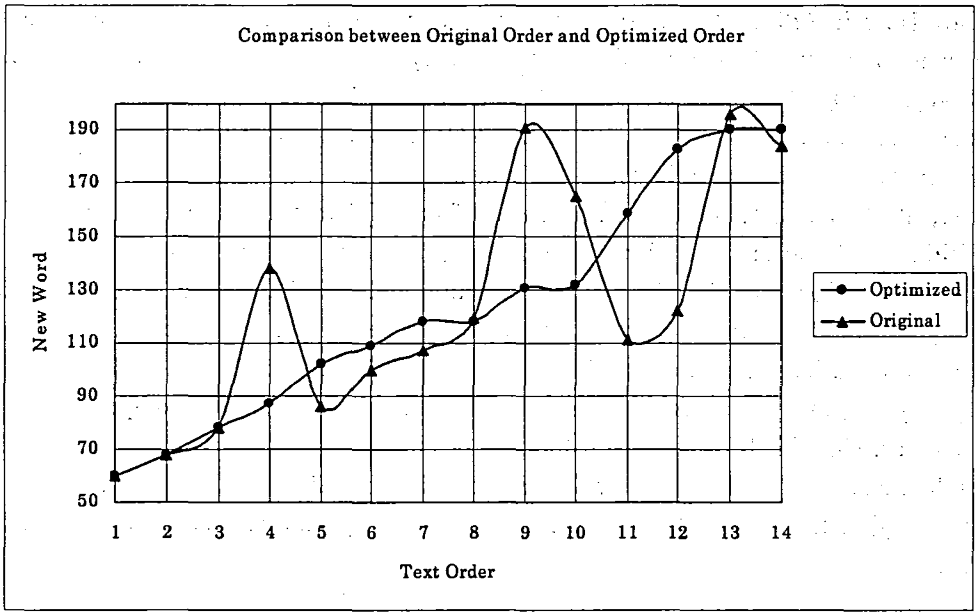

Figure 4. Comparison for optimized order and original order

5. Conclusion

The results for the numerical example operated by *CBETDS* indicate that the proposed absolute method is feasible. The optimized selection and ordering of the texts in the textbook decentralize the difficulties of the texts, and rank the texts in a smooth upward tendency according to the word difficulty of each text. The optimization applied in this study provides students with a better chance to be exposed to systemized vocabulary in the textbooks.

One of the most important features of this study is the capability of designing textbooks by using computer. This results in saving huge computing time, which accounts for the most part of the computation in textbook design optimization. Degrees of word difficulty in the texts can be yielded in a few minutes.

The absolute method can be used for quick selection when dealing with a great number of texts in a corpus. It can be used not only to reorder the materials in existing textbooks in an optimum way, but also to optimize the text order for different English courses for any level of students, if suitable word lists are available.

This study is based on the idea that vocabulary learning follows a sequential process that can be considered when designing textbooks. In other words, the present study is approached from the perspective of vocabulary acquisition. In practice, several other factors, such as topics, stylistics, syntactic complexity, influence the selection and arrangement of textbook materials. The proposed method provides an effective means for students to access new words and texts.

References

Collins, Helosia. (2000). 'Materials design and language corpora: a report in the context of distance education.' In Burnard L. and McEnery T. (eds.). (2000). *Rethink Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang.

Coxhead, A. (1998). 'An academic word list.' *Occasional Publication Number 18, LALS,* Victoria University of Wellington, New Zealand.

Nation, P. & Heatley, A. (1996). *Word & Range* [Computer software for DOS/V]. Wellington: Victoria University of Wellington. (http:www.vuw.ac.nz/lals).

Nishida, T. & Song, F. (2003). 'Optimal selection and ordering of the texts in textbook design for the facilitation of students' vocabulary access.' A paper presented at First Asia TEFL International Conference held on Nov 7-8, 2003 in Busan, Korea.

Song, Fuyun. (2002). *Corpus-Based English Textbook Design System* [Computer software program]. Higashi Hiroshima: Hiroshima University.

West, Michael. (1953). *A General Service List of English Words*. London: Longman.

Zhai, Xiangjun et al. (eds.) (1999). *College English*. Shanghai: Shanghai Foreign Language Press.