AUTOMATIC AND SEMI-AUTOMATIC PROCESSES OF WORDSMITH 3.0 AS A TEXTBOOK EVALUATION INSTRUMENT: A CASE STUDY

Jayakaran Mukundan
Anealka Aziz Hussin
Universiti Putra Malaysia, Malaysia

Abstract: As checklists developed for textbook evaluation are question-able in terms of reliability and validity, other ways are being sought to bring about more systematic, efficient and objective evaluation instruments, which can provide greater insight into the strengths and weak-nesses of textbooks. With this in mind, the researchers explored the abilities of WordSmith 3.0, a concordance software, in providing some insights into the structure of textbooks. This study will provide findings on data WordSmith 3.0 generates automatically and semi-automatically, and how this information could be used in the evaluation of textbooks.

Key words: textbook evaluation, material evaluation, ESL textbooks, concordance software

Textbook evaluation can be carried out for selection purposes or to determine the effectiveness of textbooks while they are being used. Evaluating a textbook during the selection process is known as predictive evaluation. It focuses on the potential value of the textbook. Ellis (1997) calls this an "evaluation designed to make a decision regarding what materials to use". On the other hand, evaluating a textbook that is in-use is known as retrospective evaluation, and it focuses on awareness and description of what the learners are actually doing whilst the materials are being used (Ellis, 1997).

The predictive and retrospective evaluation of a textbook has to be not only systematic but efficient. The current practice which relies solely on the evaluation checklist, is not very efficient and has created doubts among teachers as well as researchers, in terms of reliability and validity of the instruments, as many checklists developed for textbook evaluation are rarely tested. Littlejohn (1998), Sheldon (1998) and Ur (1996) have reported on flaws within the evaluation checklists, which affect the reliability and validity of the instrument. It is the inefficient and subjectivity of the checklist in textbook evaluation that prompted the researchers to look at other ways to bring about a more systematic, efficient and objective evaluation instrument, which can then provide greater insight into the strengths and weaknesses of a textbook.

Due to the inadequacy of textbook evaluation methods, which are overly reliant on the checklist, this study proposes the use of concordance software as an additional instrument for both predictive and retrospective evaluation of textbooks. Concordance software allows researchers to analyze how language is used (Fox, 1998). Some of the concordance software developed for language analysis includes Concordance 3.0, HAMLET for Windows, MicroConcord (DOS), TextQuest 1.37, MonoConc Pro 2.0 and WordSmith 3.0. Mukundan (2004) suggested using WordSmith 3.0, which is one of the three instruments in his composite framework for textbook evaluation, in evaluating a textbook.

WordSmith 3.0 is an integrated programme that looks at how words behave in a text. Oxford University Press uses the tools for lexicographic work in preparing dictionaries. These tools are also useful for language teachers, students and researchers in investigating language patterns. There are three analysis tools, which are WordList, Concord, KeyWords and three utility tools, which are Splitter, Text Converter and Viewer.

The WordList tool lists all the words or word-clusters in a text and can be presented in alphabetical or frequency order. It provides detailed statistics and consistency analyses for the text. This function is to find out which word recurs consistently, study the types of vocabulary used, identify common word clusters and compare the frequency of words in different text files or across genres. It also helps to generate lists of key-words and main key-words.

The Concordancer gives a chance to see words or phrases in context. In other words, it enables the same word or phrase to be viewed in different contexts. It provides information about collocates of the search words, dispersion plots showing location of the words and cluster analyses showing repeated clusters of words/phrases.

The KeyWord tool is used to find the keyword in a text. It gives reasonably good clue as to what the text is about. It provides the dispersion plot of the key words in the text. Associates is another function of Keyword and it refers to the key-words which are associated with the key key-word and they may/may not co-occur in proximity to the key key-word. Another function, Clumps, refers to the groups of key-words associated with a key key-word. These two functions can be activated if needed.

The Splitter tool is a utility tool, which splits large files into smaller ones for text analysis purposes. The text converter tool is another general-purpose utility, which is used to edit texts, rename text files, change file attributes and to move files into new directory if they contain certain words or phrases. The final utility tool is the Viewer tool, which allows the source text to be displayed in various formats.

The study was intended to explore the abilities of WordSmith 3.0 as a textbook evaluation instrument. All the three analysis tools were used in analyzing the textbook. In order to demonstrate these abilities clearly, one secondary English language textbook was randomly selected and in this study, a Form 4 English language textbook published by Mutiara Ilmu Sdn. Bhd. was used. As not all information provided by WordSmith 3.0 was derived automatically, users need to organize the information in such a way as to produce meaningful findings about the textbook.

The presentation of the abilities of WordSmith 3.0 is based on the research questions stated below:

1. What are the automatic processes provided by WordSmith 3.0 in evaluating a textbook?
2. What are the semi-automatic processes provided by WordSmith 3.0 in evaluating a textbook?

In this study, the following terms need to be defined:

1. Automatic processes refer to the processes to generate results automatically performed by WordSmith 3.0 without requiring any human interventions in preparing the outputs.
2. Semi-automatic processes refer to the processes to generate results that need human interventions to re-organize the outputs of automatic processes in order to provide meaningful findings to answer the research questions stated.
3. Tokens refer to the running words encountered as Wordlist processes the text

files.

4. Types refer to the different words encountered as Wordlist first processes the text.
5. Text Density ratio is calculated by dividing the number of types by the number of tokens found in the text (types/ tokens). It measures the intensity of the text itself.
6. The Consistency ratio is calculated by dividing the number of tokens by the number of types found in the text (tokens/types). It measures the consistency of the introduction of new vocabulary in the text.
7. Keyness is not the most frequent words but words, which are unusually frequent in the 1000 word article. These words give reasonably good clues as to what the text is about.

METHOD

Before using WordSmith 3.0 to analyze the Form 4 textbook, several processes are to be followed. First, each page of the selected Form 4 English textbook is photocopied on a single side of a separate piece of paper. Then, the photocopied pages are scanned using a high-speed scanner, to convert them into Tagged Image File format (TIF). After that, the image files are saved in the computer hard disk before they go through the Optical Character Recognition (OCR) to convert all the TIF files into computer text files (.txt). The text files are checked for errors before saving and renaming them according to the respective units of the textbook. Now, these text files are then ready to be analyzed.

RESULTS AND DISCUSSION

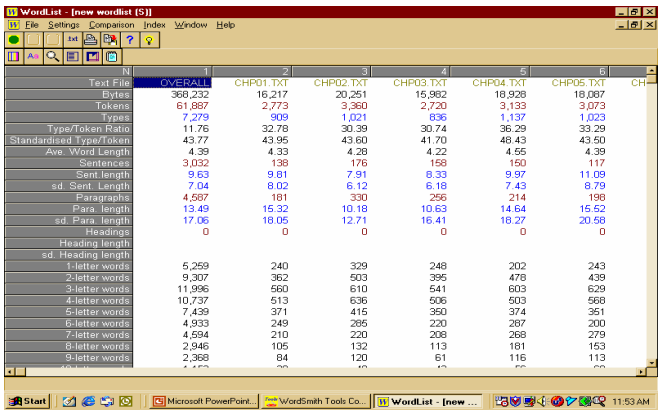
This study will illustrate the outputs of three analysis tools: WordList, Concord and KeyWord. The illustration is divided into two parts following the research questions stated earlier. The first one deals with outputs that are automatically produced by these analysis tools and the second part deals with the kinds of information that can be derived with the help of these analysis tools. The second part refers to the semi-automatic processes provided by WordSmith 3.0 in evaluating a textbook.

Automatic Processes Provided by WordSmith 3.0 in Evaluating a Textbook

Automatic processes refer to the processes to generate results automatically performed by WordSmith 3.0 without requiring any human interventions in preparing the outputs. The automatic processes of WordSmith 3.0 are WordList, Concord and KeyWord Tools.

The WordList tool provides detail statistics of the textbook, list of words in the textbook according to alphabetical and frequency order and detailed consistency analysis of the textbook. Figure 1 shows the statistics of the textbook as a whole and based on individual units in the textbook. The statistics consists of information such as Bytes (refers to the size of the file), Tokens (refers to the running words encountered by WordSmith as it processes the text), Types (refers to the different words encountered by WordSmith as it processes the text), Type/Token Ratio, refers to the text density ratio that measures the intensity of the text itself. Other information provided include: average word length, number of sentences, sentence length, number of paragraphs and paragraph length.

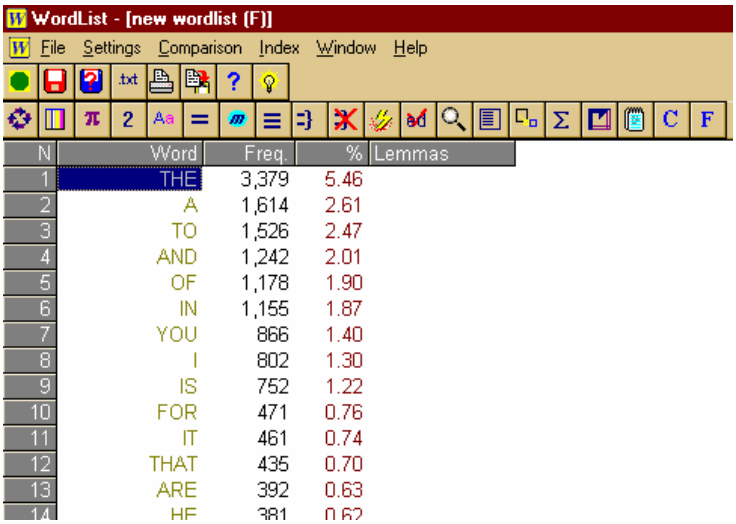
Figure 1. Detailed Statistics of the Textbook



N	1	2	3	4	5	6
Text File	OVERALL	CHP01.TXT	CHP02.TXT	CHP03.TXT	CHP04.TXT	CHP05.TXT
Bytes	368,232	16,217	20,251	15,982	18,928	18,087
Tokens	61,887	2,773	3,360	2,720	3,133	3,073
Types	7,279	909	1,021	636	1,137	1,023
Type/Token Ratio	11.76	32.78	30.39	30.74	36.29	33.29
Standardised Type/Token	43.77	43.95	43.60	41.70	48.43	43.50
Ave. Word Length	4.39	4.33	4.28	4.22	4.55	4.39
Sentences	3,032	138	176	158	150	117
Sent.length	9.63	9.81	7.91	8.33	9.97	11.09
sd. Sent. Length	7.04	8.02	6.12	6.18	7.43	8.79
Paragraphs	4,587	181	330	256	214	198
Para. length	13.49	15.32	10.18	10.63	14.64	15.52
sd. Para. length	17.06	18.05	12.71	16.41	18.27	20.58
Headings	0	0	0	0	0	0
Heading length						
sd. Heading length						
1-letter words	5,259	240	329	248	202	243
2-letter words	9,307	362	503	395	478	439
3-letter words	11,996	560	610	541	603	629
4-letter words	10,737	513	636	506	503	568
5-letter words	7,439	371	415	350	374	351
6-letter words	4,933	249	285	220	287	200
7-letter words	4,594	210	220	208	268	279
8-letter words	2,946	105	132	113	181	153
9-letter words	2,368	84	120	61	116	113

Figure 2 shows the list of words based on frequency order and Figure 3 shows the list of words based on alphabetical order. Both orders can later be sorted in ascending or descending manner.

Figure 2. List of Words Based on the Frequency Order



N	Word	Freq.	%	Lemmas
1	THE	3,379	5.46	
2	A	1,614	2.61	
3	TO	1,526	2.47	
4	AND	1,242	2.01	
5	OF	1,178	1.90	
6	IN	1,155	1.87	
7	YOU	866	1.40	
8	I	802	1.30	
9	IS	752	1.22	
10	FOR	471	0.76	
11	IT	461	0.74	
12	THAT	435	0.70	
13	ARE	392	0.63	
14	HE	381	0.62	

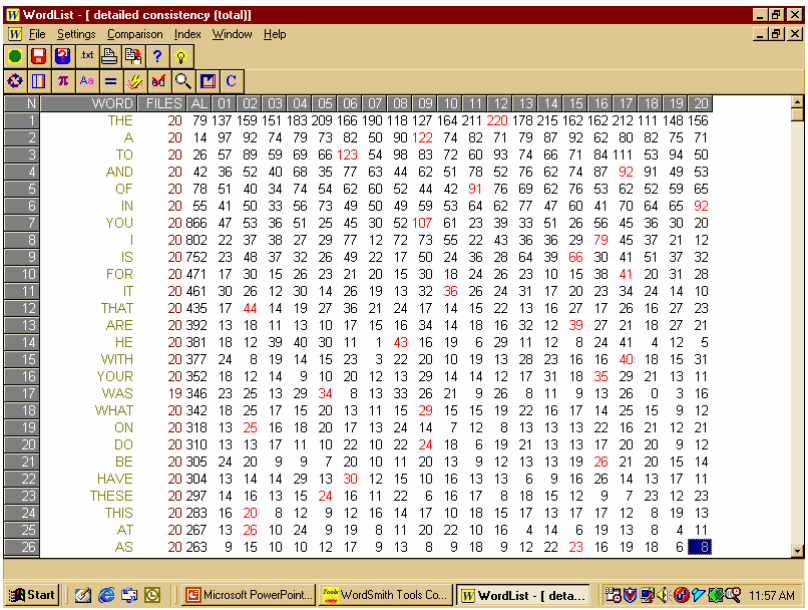
Figure 3. List of Words Based on Alphabetical Order

N	Word	Freq.	%	Lemmas
1	A	1,614	2.61	
2	ABACK	1		
3	ABILITIES	2		
4	ABILITY	5		
5	ABLE	12	0.02	
6	ABOUT	135	0.22	
7	ABOVE	30	0.05	
8	ABRASIVE	1		
9	ABRUPTLY	1		
10	ABSENCE	1		
11	ABSENT	1		
12	ABSENTEE	1		

Figure 4 shows the detailed consistency analysis of the textbook. It shows the occurrence of words in the whole textbook as well as in each unit. It can be

sorted according to alphabetical order, frequency order and its occurrence in units in the textbook.

Figure 4. Detailed Consistency Analysis of the Textbook



N	WORD	FILES	AL	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
1	THE	20	79	137	159	151	183	209	166	190	118	127	164	211	220	178	215	162	162	212	111	148	156
2	A	20	14	97	92	74	79	73	82	50	90	122	74	82	71	79	87	92	62	80	82	75	71
3	TO	20	26	57	89	59	69	66	123	54	98	83	72	60	93	74	66	71	84	111	53	94	50
4	AND	20	42	36	52	40	68	35	77	63	44	62	51	78	52	76	62	74	87	92	91	49	53
5	OF	20	78	51	40	34	74	54	62	60	52	44	42	91	76	69	62	76	53	62	52	59	65
6	IN	20	55	41	50	33	56	73	49	50	49	59	53	64	62	77	47	60	41	70	64	65	92
7	YOU	20	866	47	53	36	51	25	45	30	52	107	61	23	39	33	51	26	56	45	36	30	20
8	I	20	802	22	37	38	27	29	77	12	72	73	55	22	43	36	36	29	79	45	37	21	12
9	IS	20	752	23	48	37	32	26	49	22	17	50	24	36	28	64	39	66	30	41	51	37	32
10	FOR	20	471	17	30	15	26	23	21	20	15	30	18	24	26	23	10	15	38	41	20	31	28
11	IT	20	461	30	26	12	30	14	26	19	13	32	36	26	24	31	17	20	23	34	24	14	10
12	THAT	20	435	17	44	14	19	27	36	21	24	17	14	15	22	13	16	27	17	26	16	27	23
13	ARE	20	392	13	18	11	13	10	17	15	16	34	14	18	16	32	12	39	27	21	18	27	21
14	HE	20	381	18	12	39	40	30	11	1	43	16	19	6	29	11	12	8	24	41	4	12	5
15	WITH	20	377	24	8	19	14	15	23	3	22	20	10	19	13	28	23	16	16	40	18	15	31
16	YOUR	20	352	18	12	14	9	10	20	12	13	29	14	14	12	17	31	18	35	29	21	13	11
17	WAS	19	346	23	25	13	29	34	8	13	33	26	21	9	26	8	11	9	13	26	0	3	16
18	WHAT	20	342	18	25	17	15	20	13	11	15	29	15	15	19	22	16	17	14	25	15	9	12
19	ON	20	318	13	25	16	18	20	17	13	24	14	7	12	8	13	13	13	22	16	21	12	21
20	DO	20	310	13	13	17	11	10	22	10	22	24	18	6	19	21	13	13	17	20	20	9	12
21	BE	20	305	24	20	9	9	7	20	10	11	20	13	9	12	13	13	19	26	21	20	15	14
22	HAVE	20	304	13	14	14	29	13	30	12	15	10	16	13	13	6	9	16	26	14	13	17	11
23	THESE	20	297	14	16	13	15	24	16	11	22	6	16	17	8	18	15	12	9	7	23	12	23
24	THIS	20	283	16	20	8	12	9	12	16	14	17	10	18	15	17	13	17	17	12	8	19	13
25	AT	20	267	13	26	10	24	9	19	8	11	20	22	10	16	4	14	6	19	13	8	4	11
26	AS	20	263	9	15	10	10	12	17	9	13	8	9	18	9	12	22	23	16	19	18	6	8

The concordance tool shows the location of where certain words appear in the textbook. It allows the user to view the words as they appear in different context. It also provides the dispersion plots showing where the word appears in each unit (files). Figure 5 shows the word *advertisement* as it appears in different contexts in the textbook. Figure 6 shows the dispersion plot of words in the textbook.

Figure 7 shows an example of KeyWord output and it gives clues on what that particular unit is all about. For example, in Unit 1, based on the words below, it can be concluded that Unit 1 is about repairing a tap. Referring to the Form 4 syllabus, there is a section on how to fix things and this is the unit that teaches students on how to fix things on their own.

Figure 5. Concordance Output of the Word '*Advertisement*' in the Textbook

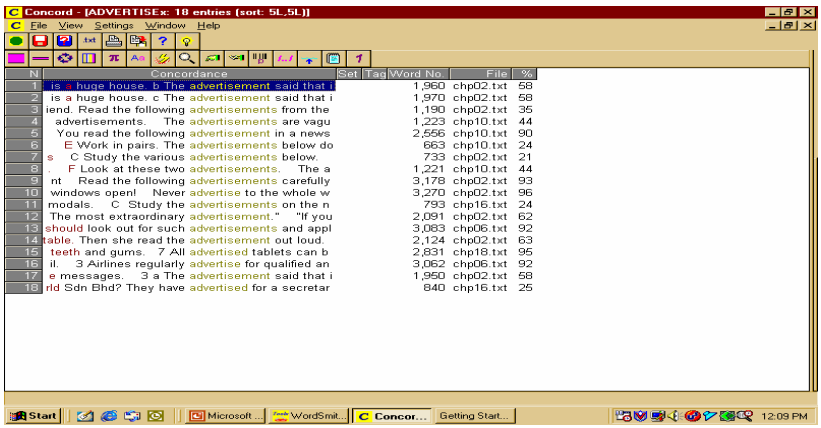


Figure 6. A Dispersion Plot of the Word '*Advertisement*' in the Textbook

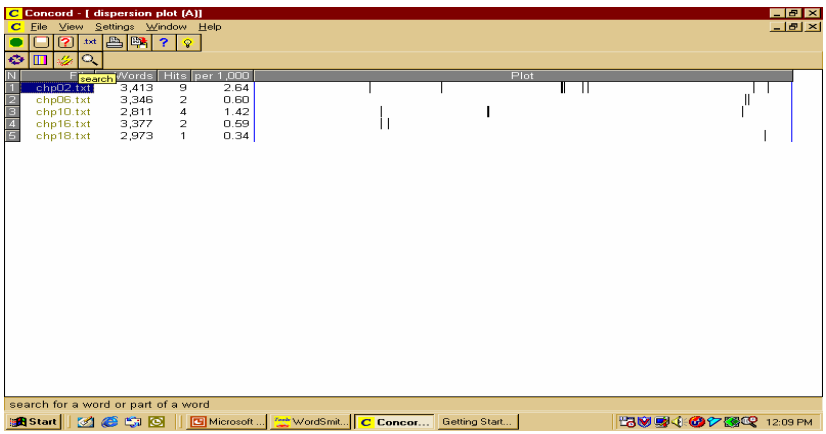
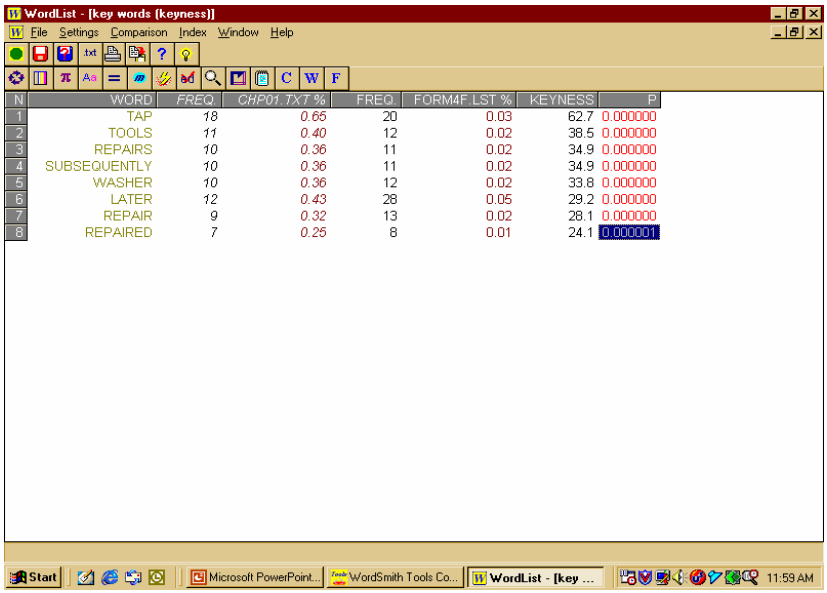


Figure 7. A Sample of KeyWord Output–Keyness



Semi-Automatic Processes of Wordsmith 3.0 in Evaluating a Textbook

Semi-automatic processes refer to the processes to generate results that need human intervention to re-organize the outputs of automatic processes in order to provide meaningful findings to answer the research questions stated. WordList, Concord and Keyword Tools are used to generate results and these results will be re-organized accordingly to produce specific findings to serve the intended purposes.

The three analysis tools, WordList, Concord and KeyWord, are used to provide the following information. The study uses the Form 4 syllabus as a guide in evaluating the Form 4 textbook. The discussion will be divided into a few sub-headings: General information about the vocabulary used in the textbook; Comparison of the vocabulary used in the textbook and vocabulary suggested by the Form 4 syllabus; The teaching of sequence connectors in the textbook; the Keyness aspects of each unit in the textbook; and gender representation in the textbook.

General Information about the Vocabulary Used in the Textbook

Vocabulary Used in the Textbook

The Wordlist Tool is used to provide this information. The proposition of types and tokens shows the density of the text (Nation, 1990). Table 1 shows the overall statistics of the textbook. The total number of the running words (tokens) found in the text is 61,887 words and the total number of different words found in the textbook is 7,279 words. The type/token ratio measures the density of the text. Every new different word is equivalent to 0.12 of the running words. The consistency ratio of the textbook is 8.502.

Table 1. Some General Information about the Vocabulary Used in the Textbook

Number of tokens (running words)	61 887 words
Number of types (different words)	7 279 words
Type/token ratio (measure density ratio)	1: 0.12
Token/type ratio (measures consistency ratio)	1: 8.502

Table 2 shows the some information about the number of running words, the number of different words, the density ratio of each unit and the consistency ratio of introduction of new words in each unit of the textbook.

Table 2. Some General Information about the Vocabulary Used in Different Units in the Textbook

Units	**Token**	**Types**	**Density Ratio (%) (Type/Token)**	**Consistency Ratio (Token/Type)**
1	2,773	909	32.78	3.05
2	3,360	1,021	30.39	3.29
3	2,720	836	30.74	3.25
4	3,133	1,137	36.29	2.76
5	3,073	1,023	33.29	3.00
6	3,291	977	29.69	3.37
7	2,744	849	30.94	3.23
8	3,051	964	31.60	3.16
9	3,594	1,016	28.27	3.54
10	2,782	903	32.46	3.08

Units	Token	Types	Density Ratio (%) (Type/Token)	Consistency Ratio (Token/Type)
11	2,888	979	33.90	2.95
12	3,229	980	30.35	3.29
13	3,313	994	30.00	3.33
14	2,912	920	31.59	3.17
15	3,227	1,048	32.48	3.08
16	3,352	1,060	31.62	3.16
17	3,711	1,054	28.40	3.52
18	2,934	864	29.45	3.40
19	2,878	945	32.84	3.05
20	2,922	872	29.84	3.35

Density Ratio of the Units in the Textbook

Density ratio refers to the compactness of the text. It is used to further illustrate the consistency in the types of words used in the textbook. Fig. 8 shows the text density ratio of each unit. It shows how each unit is fully loaded with different types of words or in other words, the amount of different types of words in each unit. The density ratio should be consistent to indicate that the different types of words are equally distributed throughout the textbook. The density ratio of this textbook is 0.12 or 12%. Figure 8 shows that there is no consistency in the density ratio of each unit in the textbook. For example, more types of words can be found in Unit 4 than in Unit 9.

Consistency Ratio of Units in the Textbook

The Wordlist tool is used to provide this information. The consistency ratio is calculated by dividing the total number of the running words (tokens) by the number of the different words (types) found in the textbook. It measures the consistency of the introduction of a new word in the text. The introduction of a new word should be done after 8.502 of running words have been covered. In other words, the introduction of every new word will only be done at every 8.502 word interval.

Figure 8. Text Density Ratio

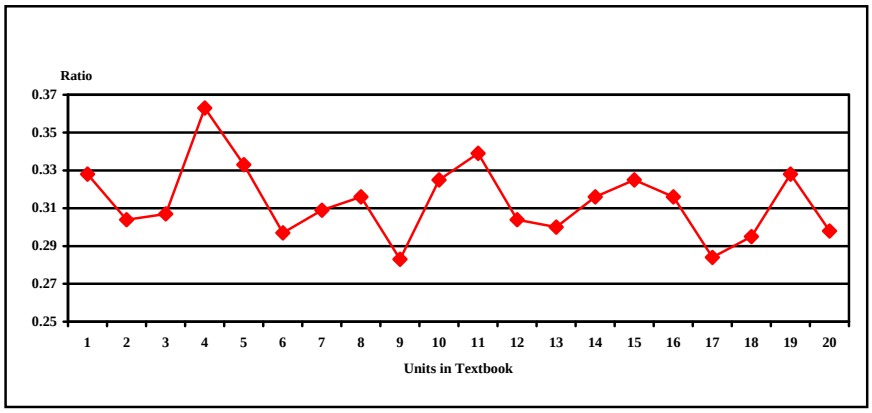
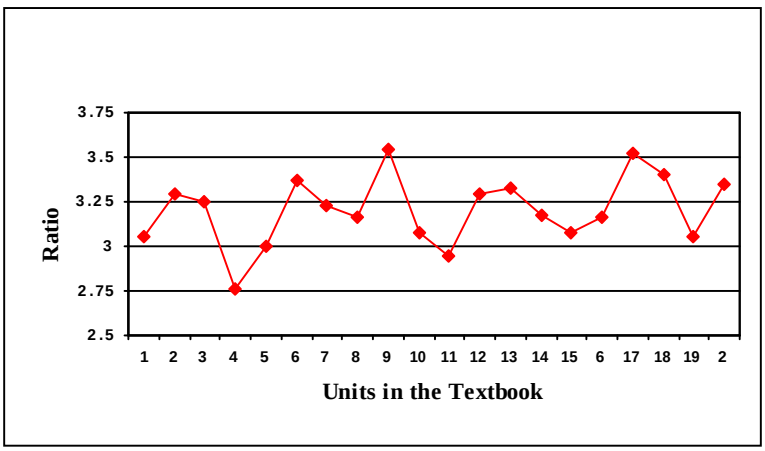


Figure 9 shows the word consistency ratio of each unit. There is no consistency in the ratio. The introduction of new words is not done at the same interval. For example, new words in Unit 4 are introduced after every 2.76 of the running word while in Unit 9 new words are introduced only after every 3.54 of the running word.

Figure 9. Word Consistency Ratio



Comparison of the vocabulary used in the Form 4 textbook and vocabulary suggested by the Form 4 Syllabus.

New, Old and Missing Vocabulary Items

There are 151 words/phrases in the suggested vocabulary list for Form 4 and 118 words (78%) out of 151 words are new words/phrases while the remaining 33 words (22%) are words, which have been introduced earlier. Words derived from the same root word, and having the same meanings are listed as one item. However, words that have the same form but have different meanings are listed as separate items. WordSmith Concordance Tool was used to check the occurrence of the suggested vocabulary items in the selected Form 4 textbook. It was found that not all vocabulary items in the suggested word list appear in the textbook. There are 28 words/phrases (18.5%) that were not used in the textbook. 24 words/phrases are new items and 4 words/phrase are old items. Table 3 shows the summary of new, old and missing words/phrases in the textbook:

Table 3. Summary of New, Old and Missing Words/Phrases in the Textbook

Contexts	Total No of suggested words/phrases	Total No. of New words/phrases	Total No. of Missing Words/Phrases in Textbook
Instructions on how to fix things	12	10	1 new word
Messages from the mass media	11	9	
Stories –self-reliance, diligence, public spiritedness	11	9	3 new words
Talks – consumerism and health care	13	10	4 new words 1 old word
Information in newspaper reports and book reports	10	7	
Information in graphs and manuals	18	12	2 new words 1 old word
Information in informal letters in the newspapers	13	12	2 new words

Contexts	Total No of suggested words/phrases	Total No. of New words/phrases	Total No. of Missing Words/Phrases in Textbook
Information in formal letters – enquiry, complaint	14	14	3 new words
Description of scenes – tourist spots in ASEAN region	10	8	1 new word
Description of events –SEA games	20	15	3 new words
Opinions on current issues – unemployment	8	6	2 new words
Description of processes & procedures – recycling materials	7	4	1 new word 2 old words
Expression of regret	4	3	2 new phrases
Social skills – interrupting, joining in and participating in a conversation	3	2	

Table 4 shows that only 2.1% of 281 words (6 words) cover half of the units in the textbook (10 units). 67% or 188 words appear once throughout all 20 units. The recurrence of words is not balanced. This shows that the words are not being recycled enough.

Table 4. Distribution of the Suggested Vocabulary in the Textbook

Total number of units covered	Total no. of words/phrases introduced	
20 units	2 words	2.1%
16 units	1 word	
10 units	3 words	
8 units	4 words	
7 units	6 words	
6 units	3 words	
5 units	5 words	
4 units	13 words	
3 units	27 words	
2 units	29 words	
1 unit	188 words (67%)	

Detailed Consistency Analysis of the Textbook

This tool provides a detailed analysis of the distribution of vocabulary in the whole textbook. Out of 7 279 different types of words in the text, only 88 words (1.21%) appear in each of the 20 units. 4.9% of 7 279 words (356 words) cover half of the textbook (10 units). 59.2% or 4311 words appear once throughout all 20 units. The recurrence of words is not balanced. The words are not being recycled enough.

Table 5. Distribution of the Vocabulary in the Textbook

Total number of units covered	Total no. of words found
20	88 (1.21%)
19	33
18	25
17	20
16	34
15	30
14	21
13	27
12	36
11	42
10	57
9	56
8	62
7	209
6	146
5	208
4	313
3	541
2	1120
1	4311 (59.2%)

The Teaching of Grammatical Items in the Textbook: Sequence Connectors

Cohesiveness in a text is important. It refers to the way information in a text is fitted together so that there is smooth and logical flow of ideas. Signal

words provide clues to the relationships between the parts of a sentence, between sentences and between paragraphs. These words are connectors of transitions from one idea to another that help make the text cohesive. One type of signal word is the sequence connector. It shows the order or sequence of ideas in a text.

There are two new sequence connectors introduced in the Form 4 syllabus, and the connectors are 'later' and 'subsequently'. The first is used to indicate a lapse of time - something that takes place after a period of time and the latter, is used to indicate the effect of an action. There are other sequence connectors introduced earlier in the previous syllabi – 'first', 'next', 'then', 'after that' and 'finally'.

The Concordance Tool in WordSmith Tools 3.0 will be used to locate the connectors in the textbook.

Units that Introduced the New Sequence Connectors

Both new sequence connectors are introduced in Chapter 1 "How Do I Fix It?. "Later" appears 12 times in Chapter 1 and the remaining 16 times in 11 different chapters in the textbook (Please refer to Table7). "Subsequently" appears 10 times in Chapter 1 and only once in Chapter 16.

The Number of Times SequenceCconnectors Appear in the Textbook

"First" appears 72 times in 21 chapters but not all are sequence connectors. 21 out of 72 are sequence connectors and they appear in 11 chapters. "Next" appears twice in the textbook, which is in Chapter 1 and Chapter 19. "Then" appears 69 times in 18 different chapters in the textbook. "After that" appears 3 times in 2 different chapters and "Finally" appears 12 times in 6 different chapters. (Please refer to Table 6 for tabulation of detailed distribution and Appendix I-VII for the distribution plots).

Recycling of the Sequence Connectors

Three out of seven sequence connectors (Then, First and Later) are well recycled. But the recycling of the remaining items need to be improved. The summary of the items is in Table 7.

Table 6. Distribution of Sequence Connectors in the Textbook

Sequence Connector	Total Hits	Chapters in the Textbook																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Later	27	12		1		1	1		1		2	3	1			1	1	1		2	
Subsequently	11	10															1				
First	21	5	1		1		1		2		1			5			1	2		1	1
Next	2	1																		1	
Then	69	5	4	7		4	1		1	6	1	2	6	12	2	1	8	2	2	1	4
After that	4	2		1										1							
Finally	12	2		1		1						4		3			1				

Table 7. Recycling of Sequence Connectors in the Textbook

Sequence Connector	Total Hits	No of Chapters	Recycling of items
Later	27	12	Sufficient
Subsequently	11	2	Very insufficient
First	21	11	Sufficient
Next	2	2	Very insufficient
Then	69	18	Very sufficient
After that	4	3	Very insufficient
Finally	12	6	Insufficient

The new and the previously learned sequence connectors are not well distributed in the textbook, except for "then", "first" and "later". These items need to be recycled well in order for students to be effective users of these items.

The Keyness Aspects of Each Unit

It compares two pre-existing word lists: the "list interested" and the "reference list" (Unit-Textbook). It is to find out which words characterize the units. The words are not the most frequent words but the words, which are unusually frequent in the 1000 word article. It gives a reasonably good clue to what the text is about. The Keyness is calculated by cross-tabulating the following:

- Its frequency in the small wordlist
- The number of running words in the small wordlist

- Its frequency in the reference corpus
- The number of the running words in the reference corpus

The p value ranges from 0 to 1 and it suggests the 1% danger of being wrong in claiming the relationship. For this text, p=0.05 (the smaller the number, the fewer the key words in the display).

The results of the analysis show that these units have significantly different words from other units. These are units where one can guess what the units are all about. Table 8 shows an example of keyness for Unit 13. By looking at these words, we can guess that Unit 13 must be about recycling and protecting the environment. It also shows that Unit 13 has a focus and the textbook could be using the thematic approach in each unit.

Table 8. Example of Keyness of Each Unit Use of Male/Female References in the Textbook

		Unit 13					
N	**Keyword**	**Freq.**	**Chp 13%**	**Freq.**	**Form 4%**	**Keyness**	**P**
1	Waste	28	0.85	28	0.05	92.3	0
2	Paper	27	0.81	36	0.06	78.8	0
3	Recycled	23	0.69	23	0.04	75.8	0
4	Recycling	16	0.48	16	0.03	52.7	0
5	Glass	16	0.48	19	0.03	49.1	0
6	Process	14	0.42	16	0.03	43.7	0
7	Materials	13	0.39	14	0.02	41.6	0
8	Refuse	11	0.33	12	0.02	35	0
9	Plants	10	0.3	11	0.02	31.7	0
10	Recycle	9	0.27	9	0.01	29.6	0
11	Used	21	0.63	82	0.13	29.6	0
12	Machine	10	0.3	14	0.02	28.5	0

WordSmith provides some basic knowledge on the ability of the writers to strike a balance and to avoid a "male-dominated" bias in the textbook. The book was analyzed to determine if there was a bias towards any particular gender by means of the density of pronouns used (he, she, her, him, his) and the nouns (man, men, woman, women, girl/s, boy/s). The results in Table 9 show that although all three writers of the book are female, the book shows strong male focus.

Table 9. Gender Bias in the Textbook

Words	Number of Units Covered	Number of Occurrence	Total Number of Occurrence
Man, Men	19	81	794
Boy / Boys	11	35	
He	20	374	
His	19	213	
Him	16	91	
Woman/Women	15	58	485
Girl / Girls	13	29	
She	20	244	
Her	19	154	

CONCLUSIONS

This study uses only three WordSmith 3.0 analysis tools. There are other tools in WordSmith 3.0 that could be used to provide more information about the textbook.

The three analysis tools managed to provide a lot of information about the textbook. The tools show that WordSmith 3.0 is able to provide information such as the total number of words in the textbook, the total number of different words used in the textbook, the total number of words per unit and the total number of different words per unit. WordSmith 3.0 is also able to measure the density ratio of the textbook and its respective units, and the consistency in introducing new words in the textbook and in its respective units.

WordSmith 3.0 can be programmed to locate words that share the same base form in the textbook. It is able to tell how many times the words appear throughout the textbook and exactly where the words appear in the textbook by showing the dispersion plot of the words. It is also able to help identify any missing words from the textbook and locate the teaching of specific grammatical features in the textbook. Besides, it is able to tell the characteristics of each unit that make it different from other units. Finally, it can provide some basic information about gender representation.

REFERENCES

Ellis, NC. 1997. Vocabulary Acquisition: Word Structure, Collocation, Word Class, and Meaning. In N. Schmitt, & M. McCarthy (Eds.). *Vocabulary: Description, Acquisition and Pedagogy*. (pp. 122-139). Cambridge: Cambridge University Press.

Fox, G. 1998. Using Corpus Data in the Classroom. In B. Tomlinson (Ed.). *Materials Development in Language Teaching*. Cambridge: Cambridge University Press.

Mukundan, J. 2004. *A Composite Framework for ESL Textbook Evaluation*. Unpublished Doctoral Thesis. Serdang: Universiti Putra Malaysia.

Littlejohn, A. 1998. "The Analysis of Language Teaching Materials: Inside the Trojan House" In B Tomlinson (Ed.), *Materials Development in Language Teaching*. Cambridge: Cambridge University Press.

Nation, P. 1990. *Teaching & Learning Vocabulary*. USA: Heinle & Heinle Publishers.

Sheldon, L. 1998. Evaluating ELT Textbooks and Materials. *ELT Journal* 42/4: 237-246.

Ur, P. 1996. *A Course in Language Teaching: Practice & Theory*. Cambridge: Cambridge University Press.