

Effects of Word Frequency on Acquisition of the Academic Vocabulary for Japanese and Chinese University Students

Fuyun SONG

Graduate School, Hiroshima University

Tadashi NISHIDA

Hiroshima University

1. Introduction

Much research has revealed that frequency is one of the important criteria for selecting vocabulary from a corpus. In practice, frequency counts have been employed to provide lists of the most frequent and widely used words of a language for teaching and learning for about half a century. Since the 1980s it has been well recognized that frequency is a valuable source of empirical information that can be used to design graded courses and reading texts, and to examine the language in depth (Biber, Conrad, & Reppen, 1994). Nation (1990) stated that the frequency of words occurring in texts encountered by learners was one of the important principles which influence the acquisition of vocabulary.

Two main approaches and factors influence vocabulary acquisition: 1) to learn new vocabulary, and 2) to access and encounter it (Ellis, 1994). In a pragmatic sense, the more we encounter the word, the more familiar we become with it. Japanese and Chinese university students are in contexts where English is a foreign language, that is, where students study English as a formal subject in school but have no immediate need for it outside of the classroom. The primary input the students receive in the language learning process is the textbook. In a certain sense, good textbooks have a great advantage of contextualizing new vocabulary items for students (Richards, 2001). Meaningful language encounters facilitate the acquisition of vocabulary knowledge, and a significant number of meaningful encounters with the target language will result in proficiency. Therefore, it is essential to explore the association between the word frequency in the textbooks used by the students and their vocabulary acquisition.

JACET 4000 Basic Words (JBW) (JACET, 1993), and Academic Word List (AWL) (Coxhead, 2000) are supposed as criteria to select the vocabulary in designing textbooks. Students' academic vocabulary knowledge is examined. The frequency information in six textbooks used by the Japanese students, and the six textbooks used by the Chinese students are counted. Association of the word frequency with the student's acquisition of the academic vocabulary is investigated using correlation and regression analysis. The differences and relationships between the frequencies in the textbooks and that in JBW and AWL are examined and measured, using *t*-test analysis.

2. Academic Vocabulary Test & Word Frequency in Textbooks, JBW, and AWL

2.1 Academic Vocabulary Test (AVT)

AVT (See Appendix 2) was designed and developed in this study to test the students' academic vocabulary knowledge. The 40 tested words were adopted from the AWL, compiled by Coxhead (1998) from a corpus of written academic English. In AVT, students were required to present their knowledge of the 40 academic words. Students' performances were evaluated by Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1997) on a scale of one to five. The selection of Scale 1 is evaluated as

gaining 1 score, and that of Scale 5 is evaluated as gaining 5 scores.

2.2 Twelve Textbooks and their Word Frequency

This study takes the word frequency in the textbooks into account to check the students' academic vocabulary acquisition. All of the texts in the 12 textbooks, including six textbooks of English used by the Japanese students and six textbooks used by the Chinese students were scanned into a computer. The word frequency was counted by the software program, Range & Word (Nation & Heatley, 1996).

The tokens, types, and word families in the 12 textbooks are shown in Table 1.

Table 1. Tokens, Types, & Families of 6 Japanese & 6 Chinese Textbooks

	Token	Type	Family	Type/token
Japanese Textbooks.	56,650	7,788	2,262	13.7%
Chinese Textbooks	64,555	7,682	2,228	11.9%

2.3 Frequency in JACET 4000 Basic Words List (JBW)

The 4000 Basic Words (1993) are the vocabulary required to acquire by the JACET Materials Research Committee when Japanese university students complete their required English courses in the period of the first two academic years. It was designed based on *The American Heritage Word Frequency Book*, *Defining Vocabulary 2000 Words in Longman Dictionary of Contemporary English*, *Computational Analysis of Present-day American English*, *Frequency Analysis of English Usage*, and *Cambridge English Lexicon*. JBW is recommended to be used as a criterion to design textbooks as what the preface in the bulletin of JACET 4000 Basic Words has described. The frequency in the 4000 Basic Words is categorized into five classifications. In this study 5 refers to the first-most frequent words, 4 refers to the second-most frequent words, 2 refers to the least frequent words in the list, and 1 refers to the words tested which are not included in the JBW List. Ten of the 40 tested words are not included in the JBW List.

2.4 Frequency in the Academic Word List (AWL) (Coxhead, 2000)

An academic word list plays a crucial role in setting vocabulary goals for language courses, guiding university students in their independent study, and assisting course and materials designers in selecting texts and developing learning activities. The academic vocabulary includes the words that students meet in a wide range of academic texts, and the acquisition of the academic vocabulary would facilitate university students' English language learning. The AWL shows learners with academic goals which words are most worth studying. Coxhead (2000) divided the AWL into 10 smaller, frequency-based sublists to aid in the sequencing of teaching and in materials development. 10 rank-ordered sublists according to decreasing word family frequency are shown in Appendix 1. The words in the first 3 sublists occur with comparatively high frequency. In this study 10 is assigned to Sublist 1 – the most frequent words, 9 is assigned to Sublist 2 – the second-most frequent words, and 1 is assigned to Sublist 10 – the least frequent words.

3. Methodology

3.1 Subjects

144 university students in their sophomore year, including 72 students at Hiroshima University in Japan in November, 2001, and 72 students at Tianjin University in China in January, 2002, were administered to conduct the Academic Vocabulary Test. The Japanese students majored in society and culture, information and behavioral sciences, and natural and social environment. The Chinese students

majored in mechanical and electrical engineering. Two groups of the students had similar proficiency in English vocabulary when they entered the university according to the curriculum in English Course – Teacher’s Manual (Suenaga, 1998), and in College English Course Syllabus (1999). Japanese students were required to have a command of 2,000 words, and Chinese students were required to have a mastery of 1,800 words after they completed the English courses in senior high school. The time available for the whole test was 40 minutes.

3.2 Procedures

First, 40 academic words in AWL were selected. Second, the frequency of the 40 tested words in the textbooks, JBW, and AWL was counted by Range & Word program. Third, Japanese and Chinese students were administered to take the AVT. Fourth, the association of the word frequency with the student's acquisition of the academic vocabulary is compared and analyzed using correlation and regression analysis. Finally, the differences and relationships between the frequencies in the textbooks and that in JBW and AWL are examined and measured, using *t*-test analysis.

Table 2. Scores of Japanese & Chinese Students, and Frequency in Textbooks, JBW, & AWL

		Score J	Freq.JT	Score C	Freq.CT	FreqJBW	FreqAWL
1	physical	3.83	8	3.56	11	4	8
2	interactive	2.63	1	2.56	3	1	8
3	environment	4.22	6	4.15	8	4	10
4	aid	2.75	2	3.51	7	4	4
5	media	3.47	6	2.99	6	3	4
6	authority	3.10	5	2.76	5	4	10
7	investigate	3.18	7	2.69	6	3	7
8	alter	2.67	4	2.93	4	2	6
9	despite	3.33	8	3.56	5	1	7
10	instruction	3.07	5	3.35	5	3	5
11	initially	2.24	4	2.25	3	3	8
12	assurance	2.40	5	2.71	2	3	2
13	furthermore	3.28	6	3.07	4	1	5
14	significance	2.75	5	3.10	7	2	10
15	contribute	2.97	4	3.17	4	1	8
16	facilitate	1.71	1	1.65	1	3	6
17	external	2.72	2	3.21	8	2	6
18	create	4.01	12	3.44	9	5	10
19	participate	3.21	4	3.74	3	2	9
20	process	3.96	8	3.57	10	5	10
21	primarily	2.86	3	2.75	4	2	9
22	evidence	3.50	7	3.15	10	2	10
23	publication	2.31	3	2.85	3	2	4
24	specific	2.88	4	3.10	5	4	10
25	select	3.81	6	3.76	7	4	9
26	specify	2.40	3	2.19	2	1	8
27	feature	3.06	5	2.76	3	4	9
28	similar	3.18	9	3.67	4	5	10
29	duration	1.43	2	1.49	1	1	2
30	vary	2.65	3	2.93	3	4	10
31	unique	3.74	4	2.68	4	3	4
32	subsequent	1.89	2	1.89	3	1	7
33	derive	1.58	1	2.22	4	2	10
34	involve	3.13	7	3.00	8	5	10
35	format	2.29	3	2.79	3	2	2
36	procedure	2.01	1	1.97	4	3	9
37	extract	1.83	1	2.00	2	1	4
38	emerge	2.35	3	1.99	5	1	7
39	supplementary	2.17	2	1.81	2	1	2
40	identify	2.85	4	3.14	4	4	10

Table 2 shows the mean scores of the Japanese and Chinese students for each of the 40 tested words in the Academic Vocabulary Test, and the frequency of the 40 tested words in the textbooks used by the students, JBW and AWL.

3.2.1 Regression and Correlation Analyses

This study uses a bivariate linear relationship to estimate a regression of the co-variation of two continuous variables between the scores of Japanese and Chinese students and the frequency in the textbooks, JBW, and AWL. Correlation and regression analyses are used to measure and show the association between the students' scores and the word frequency in the textbooks, JBW, and AWL. The direction and strength of the linear relationship between the frequency and the scores are estimated. The correlations between the scores and the frequency are shown as follows:

Table 3. Correlations between the Scores and the Frequency

	<i>r</i> value	Sig. (2-tailed)	<i>df.</i>
Score J & Freq. JT	.794	.01	38
Score C & Freq. CT	.649	.01	38
Score J & Freq. JBW	.530	.01	38
Score C & Freq. JBW	.507	.01	38
Score J & Freq. AWL	.386	.05	38
Score C & Freq. AWL	.382	.05	38

Score J: Mean scores of the Japanese students
 Score C: Mean scores of the Chinese students
 Freq. JT: Word frequency in the six textbooks of English used by the Japanese students
 Freq. CT: Word frequency in the six textbooks of English used by the Chinese students
 Freq. JBW: Word frequency in JACET 4000 Basic Words
 Freq. AWL: Word frequency in Academic Word List

Table 3 and Figure 1 show that our observed *r* value .794 is larger than the critical one .393 with a .01 significant level for a two-tailed test. This means that the word frequency in the Japanese textbooks of English have a closer relationship with the scores of the 40 tested academic words for the Japanese students than the ones in the other figures. The closer the points cluster around the line, the stronger the relationships are.

Table 3 and Figure 2 show the association between the scores of the Chinese students and the word frequency in the Chinese textbooks of English. It is found that the observed *r* value .649 is larger than the critical *r* value .393 with a .01 significant level for a two-tailed test. This indicates that there is also a significant correlation between the word frequency in the textbooks used by the Chinese students and the test scores. But the correlation .649 is smaller than the one .794.

Table 3 and Figures 3 & 4 show the relationship between the scores of the Japanese and Chinese students for the 40 tested words, and the frequency in JACET 4000 Basic Words. The observed *r* values are .530, and .507, respectively. This indicates that the frequency in JBW has some correlation with the scores of the Japanese and Chinese students, but the correlation is smaller than that of the textbooks. It can also be seen that the correlation of the frequency in JBW for the Japanese students is closer than

that for the Chinese students.

The scatter plots of 6 pairs of the bivariate relationships reflecting the actual strength of the correlation between scores and frequency are shown in the following 6 figures.

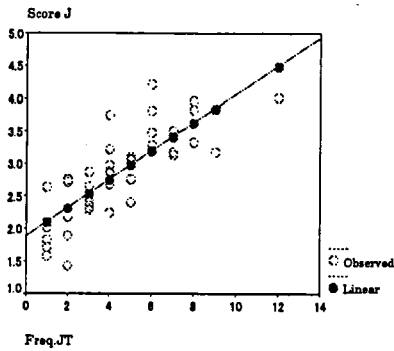


Figure 1. Score J & Freq. JT

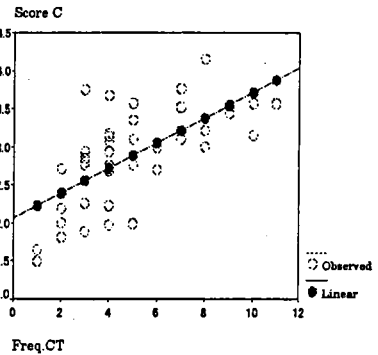


Figure 2. Score C & Freq. CT

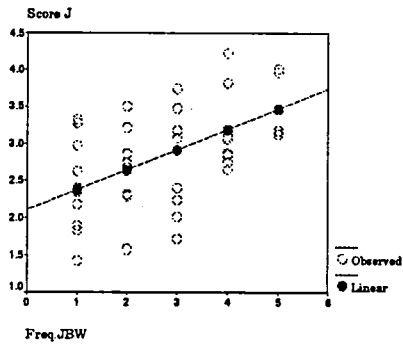


Figure 3. Score J & Freq. JBW

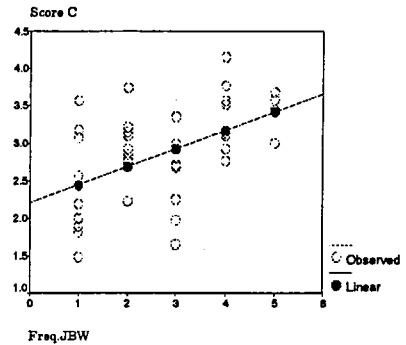


Figure 4. Score C & Freq. JBW

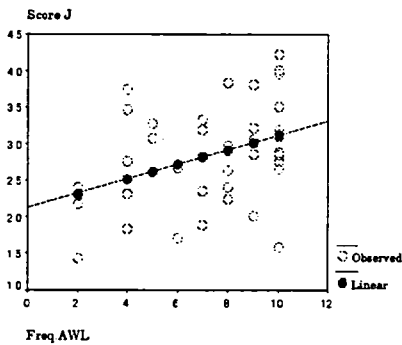


Figure 5. Score J & Freq. AWL

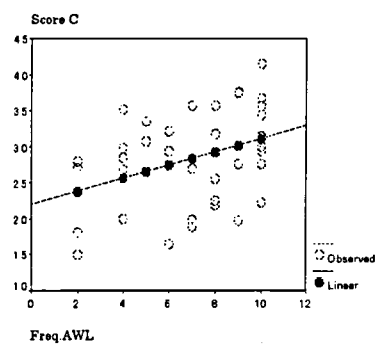


Figure 6. Score C & Freq. AWL

Table 3 and Figures 5 & 6 show that there is some correlation between the students' scores and the frequency in AWL. The observed r values are .386 and .382, respectively. The critical r value with a .05 significant level for a two-tailed test for $d.f. = 38$, is .314. Thus, our observed r values .386 and .382 are

a little bit larger than the critical r value .314. The frequency in AWL has some correlation with the scores of the Japanese and Chinese students, but the correlation is smaller than that in the textbooks and in JBW. The results show that the correlations between the frequency in AWL for the scores of the Japanese and Chinese students are quite approximate. The correlation between the scores of the students and the frequency in AWL are the smallest among the three sources of the frequency.

The analyses of correlation and regression show that among the six pairs of the correlation between the students' scores and the frequency in the textbooks, JBW, and AWL, the correlations of the scores and the word frequency in the textbooks are larger than the ones in JBW and AWL. The correlation of the frequency in JBW for the Japanese students is closer than that for the Chinese students. This reveals that the words the Japanese and Chinese students encounter directly and the frequency of the words occurring in the textbooks substantially influence their acquisition of academic vocabulary.

3.2.2 Analysis of *t*-Test & Correlation for 6 Pairs of Scores & Frequency

Six pairs of the variables, including mean scores of Japanese and Chinese students, the frequency in the textbooks, JBW, and AWL are compared and analyzed by using *t*-test. It computes 1) the differences and the correlations between the mean scores of Japanese and Chinese students, 2) the differences between the frequency of the textbooks used by Japanese students and the textbooks used by Chinese students, and 3) the differences between the frequency in the textbooks, JBW, and AWL.

Table 4. *t*-Test & Correlation for 6 Pairs of the Scores & the Frequency

		Std. D	Std. E	Correl.	<i>t</i> value
Pair 1	Score J - Score C	.38713	.06121	.835	-.282
Pair 2	Freq. JT - Freq. CT	2.15787	.34119	.637	-1.172
Pair 3	Freq. JT - Freq. JBW	2.12675	.33627	.542	-5.983
Pair 4	Freq. CT - Freq. JBW	2.21649	.35046	.487	-5.477
Pair 5	Freq. JT - Freq. AWL	2.98619	.47216	.353	5.055
Pair 6	Freq. CT - Freq. AWL	2.80007	.44273	.432	5.992

$p < .05$; $d.f. = 39$

In Pair 1, the observed *t* value is -.282. The critical *t* value for $d.f. = 39$, is 2.021 with $p < .05$. There is no significant difference between the scores of the Japanese and Chinese students, as the observed *t* value, -.282, is smaller than the critical one, 2.021. The correlation, .835, between the mean scores of the Japanese (2.84) and Chinese (2.85) students is very high, which indicates the mean scores of the two groups of the students are roughly the same. The highest score is 5, and the lowest one is 1. This result indicates that the general proficiency of the academic vocabulary of the two groups of the students is roughly similar. It is very interesting to see in Pair 2 that there is not much difference between the frequency in Japanese and Chinese textbooks. This means that the frequency for the 40 tested words both in Japanese and Chinese textbooks are approximately similar.

In Pair 3 and 4, there is difference between the frequency in JBW and the textbooks used by Japanese and Chinese students. The observed *t* values are -5.983 and -5.477, respectively, which means that there is more difference between the frequency in JBW and the textbooks used by Chinese students, than that in JBW and the textbooks used by Japanese students. In Pair 5 and 6, there are differences between the

frequency in AWL and the textbooks used by Japanese and Chinese students, with t values of 5.055 and 5.992, respectively. The results that the correlations of the frequency in JBW and AWL with that of the textbooks are comparatively lower, indicate that there are differences in the frequency information of the textbooks from that of JBW and AWL.

4. Conclusions

The frequencies of the 40 tested words in the textbooks have a closer relationship with the scores of the Japanese ($r=.794$) and Chinese ($r=.649$) students than the frequencies in the other sources. This reveals that the word frequency in the textbooks exerts greater effects on students' acquisition of the academic vocabulary as written textbooks are one of the major sources for students to encounter vocabulary. The correlation of the frequency in JBW for the Japanese students is closer than that for the Chinese students. The general proficiency of the academic vocabulary of the two groups of the students is approximately similar. Both groups of the students are having knowledge of about 57% for the 40 tested words.

The results of this study show that the words the Japanese and Chinese students encounter directly and the frequency of the words occurring in the textbooks substantially influence their acquisition of academic vocabulary. This agrees with what has been discussed: word frequency and its familiarity contribute to the acquisition of the academic vocabulary. In vocabulary teaching and learning, the frequency count ought to be one of the important variables in selecting what to teach. With the availability of the principled selection of the vocabulary for university students to learn, therefore, it is necessary to better the selection and arrangement of the vocabulary in the design of textbooks since this will exert a great influence on the students' acquisition of the vocabulary.

References

- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based issues in applied linguistics. *Applied Linguistics*, 15, 169-189.
- Coxhead, Averil. (2000). A new academic word list. *TESOL Quarterly* 34, 213-238.
- Committee of College English Course Syllabus. (1999). *College English Course Syllabus*. Shanghai Foreign Language Studies Press, China.
- Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: a review essay, *Applied Language Learning*, 5, 1-32.
- JACET 教材研究委員会 (Ed.). (1993). *JACET 4000 Basic Words*. 大学英語教育学会, Japan.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nation, I. S. P. and Heatley, A. (1996). *VocabProfile, Word & Range: programs for processing text*, LALS, Victoria University of Wellington, New Zealand.
- Paribakht, T. S. & Wesche, M. B. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 174-200). Cambridge: Cambridge University Press.
- Richards, Jack C. (2001). *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.
- Suenaga, Kuniaki. (Eds.). (1998). *English Course – Teacher's Manual*. Bun-Eido, Japan.

Appendix 1

Sublist	Items	Coverage of Academic Corpus (%)	Cumulative coverage (%)	Pages per repetition in Academic Corpus
1	60	3.60%	3.6	4.3
2	60	1.80%	5.4	8.4
3	60	1.20%	6.6	12.3
4	60	0.90%	7.5	15.9
5	60	0.80%	8.3	19.4
6	60	0.60%	8.9	24.0
7	60	0.50%	9.4	30.8
8	60	0.30%	9.7	49.4
9	60	0.20%	9.9	67.3
10	60	0.10%	10.0	82.5

Appendix 2 Academic Vocabulary Test

Direction: Read Text 1 and Text 2, and pay attention to the meaning of the 20 underlined words in each of the two texts, which also are listed in Answer Sheet. Then read the sentences in the five scales. Tick (.) one choice from I, II, III, and IV for each word.

Text 1

Part of this distancing effect also lies in the (1) physical layout of the University and its classrooms. The immovable rows of seats in the larger lecture theatres are not conducive to the development of (2) interactive learning (3) environments. Some lecturers continue to give classes without the (4) aid of (5) media in the belief that their ancient lecture notes and measures tones of (6) authority will motivate 150 or 300 recently enrolled First Years to (7) investigate a subject further at a later date. In spite of this however, several students had attended lectures where staff had (8) altered this traditional approach to large group teaching and, (9) despite the problems with room layout and class size, had provided interactive (10) instruction which the students remembered and enjoyed.

Small group or tutorial situations were a greatly preferred style of teaching. Several students mentioned that they (11) initially lacked confidence in speaking before people they didn't know, but they had gained a sense of self (12) assurance over their time at the University and most of them used tutorials to "bounce ideas off other students". (13) Furthermore, some students had extended the study group approach independently and had formed informal, self-led study groups with other Maori students. These groups hold particular (14) significance in (15) contributing to the informal rhythms of university life. They (16) facilitate the development of strong networks among Maori students, and they also provide support for students who have knowledge of their own culture and who wish to enter the 'deep' structures of learning. For those students competing against the pressures of university study, the commitment to adopt 'deep' approaches to their learning can be swayed by (17) external factors, such as

departmental ethos, teaching methodologies and an overabundance of unconnected information. The desire to (18) create knowledge from a quantity of information may be in part satisfied by students who (19) participate in self-directed study groups. Here, it may be possible for the university to lend its support to those students who are taking the time to extract an understanding of intellectual (20) processes alongside the demand to keep producing essays and degrees.

Text 2

The Burns and Mitchell technique of dating business cycles relied (1) primarily on two sorts of information: the descriptive (2) evidence from business (3) publications and general business conditions indices, and the “(4) specific cycles found in many individual series and the tendency for turning points to sometimes cluster at certain dates. Based on this information, a set of reference cycle dates were (5) selected that (6) specified the turning points in aggregate economic activity”. A key (7) feature of the Burns and Mitchell approach was to focus on the amount of cyclical co-movement or coherence among a large number of economic variables. This co-movement is the prime characteristic of their definition of the business cycle “... a cycle consists of expansions occurring at about the same time in many economic activities, followed by (8) similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle: ... in (9) duration business cycles (10) vary from more than one year to ten or twelve years ...”.

The NBER approach is based on the view that there is no (11) unique way of combining all these activities, and accordingly the business cycle cannot be fully depicted by a single measure, e.g. Burns (1969, p 13). Burns and Mitchell, and (12) subsequent NBER researchers, intended therefore, before the computer age, to provide a standard technique with a set of decision rules for (13) deriving business cycle turning points based on these two sorts of information. In practice, this (14) involved the application of a standard (15) format of filtering (16) procedures to (17) extract the turning points in each data series, and then combining this information in a judgmental way to determine a single turning point date. Other procedures, notably reference cycle indexes and coincident indexes, subsequently (18) emerged as (19) supplementary procedures for combining a large number of data series including various measures of output, production inputs, price series, monetary aggregates, etc, into a single composite index which have also been used to (20) identify turning points.

Answer Sheet

		I	II	III	IV	V
		I don't remember having seen this word	I have seen this word before, but I don't know what it means.	I have seen this word before, and I think it means _____. (synonym or translation)	I know this word. It means _____. (synonym or translation)	I can use this word in a sentence: _____. (Write a sentence.)
1	physical					
2	interactive					
3	environment					
4	aid					
5	media					
6	authority					
7	investigate					
8	alter					
9	despite					
10	instruction					
11	initially					
12	assurance					
13	furthermore					
14	significance					
15	contributing					
16	facilitate					
17	external					
18	create					
19	participate					
20	process					
21	primarily					
22	evidence					
23	publications					
24	specific					
25	selected					
26	specified					
27	feature					
28	similarly					
29	duration					
30	vary					
31	unique					
32	subsequent					
33	deriving					
34	involved					
35	format					
36	procedure					
37	extract					
38	emerged					
39	supplementary					
40	identify					