VOCABULARY IN ENGLISH FOR ACADEMIC PURPOSES: A CORPUS STUDY OF JOURNAL ARTICLES

Mayumi Tsubaki

Introduction

With the growth of English as a means of communication, a great number of people feel the need to learn English. It is indeed ideal to raise overall English proficiency so that English learners can function in all situations. However, English learners in Japan, in an English as a Foreign Language (EFL) environment, need English only in particular areas for their work or study, in contrast to English as a Second Language (ESL) learners who have to use English constantly in and out of class. They usually have limited time to study English, and they do not feel a serious need to improve their English skills outside of their work or study. For example, nurses working for hospitals in Japan may have to speak English only when communicating with patients from overseas. Psychology majors in Japanese universities might be required to read journal articles that are written in English for their class assignments and research. Indeed, there are many cases where English learners are interested in improving their English skills only in order to manage their duties for work or study, for they can function well using their native or other language in other situations. This paper will analyze vocabulary in texts from professional journal articles as one specific area which EFL learners can be interested in learning because unfamiliar vocabulary is viewed as one major barrier to comprehension.

Literature Review

English for Specific Purposes

In order to deal with the issue above, I believe that English for Specific Purposes (ESP) should be adopted to reach the goals in English learning, although teaching of "general" English is commonly practiced in Japan. According to Dudley-Evans and St. John (1998), ESP is "designed to meet specific needs of the learner" (p.4), and it is contrasted with

English for General Purposes (EGP). Robinson (1991) mentioned that ESP is usually goal-oriented, and it contains "specialist language and content" (p. 4). He listed two types of ESP: -English for Occupational English (EOP) and English for Academic Purposes (EAP). Both areas, along with language research in discourse communities, received recognition and developed all over the world in the 1990's (Noguchi, 2000).

English for Academic Purposes

　　English for Academic Purposes (EAP), the focus of the current paper, is defined as "the teaching of English with the specific aim of helping learners to study, conduct research or teach in that language" (Flowerdew & Peacock, 2001). Dudley-Evans and St. John (1998) categorized EAP situations into four types: EAP in an English-speaking country, EAP in an ESL environment in which English is the official language and is commonly spoken, EAP situation where some subjects are taught in English, and EAP situations where classes are conducted in the national language. Japan belongs to the fourth type, and English is used for particular situations. Flowerdew and Peacock also explained four types of EAP environments. In their analysis, Japan, along with some countries of Western Europe, China, and Latin America, is regarded as a country which does not have any historical connection to English, and, therefore, EAP, rather than EGP, is more urgently needed and, hence, should be far more motivating. In fact, college and university undergraduate and graduate students can survive without raising their overall English proficiency, and among all EAP skills, English journal reading and writing are probably the top priorities and are the most beneficial skills to acquire in class. There are also professionals, such as other language teachers, medical professionals, and scientists, who would like to have sufficient reading skills for the purpose of getting information through journal articles. In all, language teachers should make those EFL learners familiar with academic papers so that they would have a less difficult time reading academic papers in the English language.

　　One area that learners have to be familiar with is vocabulary words that appear in journal articles. Concordancers, or computerized text analysis programs (Flowerdew, 2001), are used for investigation of vocabulary in texts. To my knowledge, only one study has investigated vocabulary which is employed in journal articles. Pickard (1995) examined 11 journal articles from one applied linguistics journal in order to examine lexical and grammatical features in citation and quotation that expert writers choose, realizing that ESL students use "say" too often. Her findings indicated that ESL students lack knowledge in citation and quotation and that data-driven exercises with computerized concordancing programs can possibly help learners to be aware of vocabulary and then to choose words in a more sophisticated manner.

More research seems to have been conducted on vocabulary and academic texts. Ward (1999) analyzed words from 5 first-year university level engineering texts and came to the conclusion that EAP engineering students can function sufficiently with knowledge of only 2,000 word families including function and general words. Sutarsyah, Nation, and Kennedy (1994) also demonstrated the usefulness of ESP courses by comparing 300,000 words of a single economics text with a variety of academic texts of a similar number of words. The result clearly showed words in economics texts come from a far smaller number of word families than words in general Academic English texts. Thurstun and Candlin (1998) described a project in which students were exposed to academic vocabulary in academic discourse with a computer concordancing program. It was found from the research that both teachers and students see the value of the data-driven exercises.

Journal articles exhibit unique discourse and vocabulary, and knowing what is in those articles will definitely help teachers and learners to deal with them. In this research, I used computer software programs as in previous research. Teaching English to Speakers of Other Languages (TESOL) journals were taken to represent English journal articles and also one area of EAP. As a result of analysis, I will also suggest what English teachers should be concerned with when helping students to develop skills to read journal articles.

Four Types of Vocabulary

The analysis of vocabulary in this paper is based on the idea that English words can be categorized into the following.

(1) high frequency words
(2) academic words
(3) technical words
(4) low frequency words.

The approximately 2,000 high frequency words are seen in most uses of the English language. The standard list of these words is the General Service List (GSL) of English Words (West, 1953) In spite of its age, it is still considered the best list (I.S.P. Nation, 1900; Coxhead & Nation, 2001). The GSL has two parts, the first 1,000 words (referred to as GS1), and the second 1,000 words (referred to as GS2). GS1 covers about 77 % of academic texts and GS2 5% (I. S. P. Nation, 2001).

The second category, academic words, contains words that frequently appear in academic texts regardless of subject areas, but it is not common in non-academic materials. Coxhead (2000) conducted a study using an academic corpus from various subject areas and made the best list of academic vocabulary of 570 words. This list does not include words from the GSL. Words such as *assume*, *establish*, and *major* are included in the list.

This is referred to as the Academic Word List (AWL), and it covers about 8.5% to 10% of the vocabulary found in academic texts (Coxhead & Nation, 2001). Therefore, English learners are strongly recommended to learn the words in the AWL in addition to those in the GSL if they have to read academic texts.

The third category is technical words, which are commonly seen in a particular subject, but not in others. The category is assumed to have up to 1,000 words and covers 5% or less of the running words in academic texts. The last and largest category is low frequency words (Coxhead & Nation, 2001). Words in this category range from the words that English learners come across with moderate frequency to the words that English speakers rarely use (I. S. P. Nation, 2001; P. Nation, 1990).

Research Questions

This study attempts to answer the following research questions:

1) How many words do English learners have to know to comprehend different sections (abstract, introduction, method, result, discussion, and conclusion) of journal articles?

2) What are the percentages of coverage by the GSL and the AWL? Are there any differences among different sections of journal articles?

3) Are there any differences in the AWL words that are commonly frequent in each section of journal articles?

4) What words outside of the GSL and the AWL frequently occur in journal articles?

Method

Materials

Ten articles from TESOL Quarterly Digital version were examined (see Appendix 1 for the list). TESOL Quarterly is a refereed journal that is well-known among TESOL professionals. Because different components and formats exist in articles, I selected ten articles of similar components. I started to look at the most recent issue and then continued till I found the total of articles with all the sections of abstract, introduction, method, result, discussion, and conclusion. Nine that satisfy the condition were selected, while one had all the sections except a conclusion section. The word count of the articles ranges from 6,355 to 11,685, and the total number of words is 81,795.

Analysis

The analysis of this corpus was carried out using VocabProfile and Range (Nation & Coxhead, 2002), computerized vocabulary analysis programs for texts. The basic idea behind the programs is that the importance of a word can be determined by its frequency and range of occurrence according to explanation of the previous version by P. Nation

(2001). In other words, words of high frequency are considered more important to learn than others. Not only words in the GSL (GS1 and GS2) and AWL but also their inflected and derived forms are available; therefore, VocabProfile shows the number and coverage of word tokens, word types, and word families (I. S. P. Nation, 2001). Token are every single word in a text, and each occurrence of the same word is counted. When counting word types, if the same word appears more than once, it is counted as one. A word family includes a headword and its inflected and reduced from. Some examples are third person singular, plural, past tense, -ing, and comparative and superlative forms. Headwords with regular affixes such as -ly, -ness, and un- should be added to a word family (I. S. P. Nation, 2001). Texts were downloaded from the digital journal and words used in tables and figures were excluded.

Results and Discussion

The number of word tokens and types overall and in each section indicated that the introduction section has the largest number of words while the method and discussion sections have the second largest number of words (see Table 1 and Table 2). There were 16 times as many word types as in the introduction and about 10 times in the method and discussion sections, compared with those in the abstract section. While it comes to word types, 5 times as many words exist in the introduction, method, and discussion sections. The introduction generally contains information on empirical evidence for the needs of the study and the theoretical background, including previous studies; it has essential information. Therefore, I can say that the introduction can be the hardest section to comprehend not only for its content but also for the number of words included. At the same time, it is reasonable to assume that English learners may also face difficulties when reading the abstract, for they have to understand a great deal of information that is explained with a great deal of the AWL vocabulary in short passages.

Table 1 Number of Word Tokens by GSL and AWL in Different Sections

	Abstract	Introduction	Method	Result	Discussion	Conclusion	Overall
GS1	1088	18064	11667	9895	12056	2708	55478
GS2	78	1095	893	593	749	153	3561
AWL	257	3711	2059	1641	2066	570	10304
Other	234	3833	2002	1792	2018	482	10361
Total	1657	26703	16621	13921	16889	3913	79704

Table 2 Number of Word Types by GSL and AWL in Different Sections

	Abstract	Intro.	Method	Result	Discussion	Conclusion	Overall
GS1	312	1198	1011	868	963	482	4834
GS2	53	290	243	180	227	75	1068
AWL	154	775	546	383	564	261	2683
Other	126	1205	763	417	703	208	3422
Total	645	3468	2563	1848	2457	1026	12007

Intro.: Introduction

When GSL (GS1 and GS2) and AWL were run against texts from each section of the journal articles, it was found that there are no significant differences among different sections (see Table 3). The GSL covers between 70.4% and 75.8% of word token in journal article texts, and the AWL covers about 11.8% and 15.5%. Among all the sections, the abstract has the least coverage of the GSL and the most coverage of AWL probably because it has the most condensed information. Interestingly enough, the combination of the GSL and AWL makes a minimum of 85.7% coverage of word tokens in the abstract and a high percentage in the other sections. That is, 15% of the words are either technical or low frequency words.

Let us look at the differences in coverage between journal articles and general academic texts. The figures obtained from the overall journal articles reveal smaller coverage of the GSL and larger coverage of the AWL and more portions of words in neither the GSL nor the AWL than those in general academic texts. As mentioned earlier, the coverage of the GSL and AWL in academic texts was 80% and 8.5%–10% of running words respectively, according to Coxhead and Nation (2002). The GSL provides coverage of 74.2% in journal articles, which is about 5% lower in comparison with general academic texts. On the contrary, the AWL provides coverage of 13.1% in journal articles, and this is 3% to 5%

Table 3 Percentage of Word Tokens by GSL and AWL in Different Sections

	Abstract	Intro.	Method	Result	Discussion	Conclusion	Overall
GS1	65.7	67.7	70.2	71.1	71.4	69.2	69.9
GS2	4.7	4.1	5.4	4.3	4.4	3.9	4.3
GST	70.4	71.8	75.6	74.4	75.8	74.1	74.2
AWL	15.5	13.9	12.4	11.8	12.2	14.6	13.1
Total	85.9	85.7	88.0	86.2	88.0	88.7	87.2

Intro.: Introduction

more than in regular academic texts. Moreover, every 7th running word belongs to the AWL in journal articles. Thus, I would like to emphasize the importance of the AWL in comprehending the content of journals.

Since the words in the GSL are easier to learn and understand than words in the AWL, I looked at the AWL words of high frequency in each section. Table 4 and 5 reveal that there are different sets of academic words in each section. The 20 most frequent AWL words from all the ten journal articles are listed in the left column, and the order is written to the left of each word. Likewise, those in the other sections that match are written in Table 4. More than 20 words are listed when there are a word or words of the same rank order. The words in the overall section and in each individual section are in bold type. We can say that close to half the words are the same in each section and in the overall list and that different sections have different sets of the AWL words. It should be also noted that most words here have meanings and concepts specific to the field of TESOL.

More than 10 % of the vocabulary words are outside of the GSL and AWL, and that means every 10th word is neither a GSL nor AWL word. As Table 5 shows, the majority

Table 4 Top 20 Frequent AWL Words in Overall Articles and Individual Sections

Overall	Abstract	Intro.	Method	Result	Discussion	Conclusion
1. task	**output**	**input**	**task**	**task**	**task**	**input**
2. focus	**target**	**focus**	**data**	**significant**	**focus**	**focus**
3. input	**input**	**research**	**target**	**tense**	**input**	**output**
4. target	**research**	**target**	**tense**	**overall**	**research**	**target**
5. research	**data**	**output**	error	**input**	**target**	**task**
6. significant	**focus**	**task**	clause	**positive**	**significant**	**context**
7. output	**task**	**structure**	**phase**	**target**	**structure**	**positive**
7. text	topic	error	**structure**	credit	**data**	**research**
9. data	error	**positive**	topic	**phase**	**specific**	goal
10. tense	previous	role	version	**text**	**phase**	**structure**
11. structure	**specific**	**significant**	**focus**	grade	similar	benefit
12. specific	adult	approach	**research**	contrast	**whereas**	**data**
13. positive	function	**contact**	**text**	**hypothesis**	strategy	theory
14. hypothesis	furthermore	issue	author	item	**context**	error
15. phase	indicate	**specific**	**output**	random	feature	feature
16. overall	**positive**	explicit	item	**specific**	**hypothesis**	modify
17. process	relevant	adult	similar	gender	role	promote
18. context	**significant**	diverse	subsequent	impact	indicate	strategy
18. whereas	**structure**	impact	design	index	initial	**text**
20. contact	**text**	affect	period	somewhat	issue	aspect
	whereas		**significant**	**whereas**	**process**	
			specific	clause	promote	

of the most frequent words outside of the GSL and AWL are technical words, which are specific to the field of TESOL. There are three types of technical words and low frequency words. The first category is concept words, such as *comprehension* and *proficiency*, which get a lot of attention from TESOL researchers and whose concepts and definitions are difficult. The second category of technical words, such as *classroom* and *pretest*, are those which are not the focus of research but which appear more as general language of educational related journals. The last category of technical words consists of acronyms that are made up of the AWL and GSL words, for example, *L1* (the first language) and *ESL* (English as a Second Language). Low frequency words are *preemptive*, and names of researchers (e.g. Doughty), which are specific to some journal articles. Hence, these words are likely to disappear if a larger number of articles are examined. English teachers have to teach the definitions and concepts of the technical words and should not put too much focus on the low frequency words.

Table 5 Top 10 Frequent Words outside of AWL and GSL in Overall Articles and Individual Sections

Overall	Abstract	Intro.	Method	Result	Discussion	Conclusion
comprehension	L2	recast	comprehension	posttest	rhetorical	feedback
ESL	rhetorical	ESL	scores	scores	L2	recasts
L2	comprehension	L2	ESL	uptake	Japanese	L2
rhetorical	feedback	feedback	proficiency	ESL	ESL	preemptive
FFES	classroom	AL	Japanese	Phil*	recasts	comprehension
Japanese	preemptive	ET	L1	FFES	LP	pronunciation
preemptive	recast	rhetorical	classroom	pretest	FFES	EFL
scores	flap	Chinese	recall	rhetorical	HP	classroom
recasts	corrective	classroom	Chinese	feedback	AP	corrective
linguistic	FFES	classrooms	recast	L2	Lyster*	immersion
	Japanese	mainstream	utterance			Japanese
	proficiency	Doughty*				uptake
	versus					

*Doughty, Phil, and Lyster are names of authors cited in journal articles.

Conclusion

In conclusion, English teaching professionals should gear their teaching toward journal article reading if English learners feel the urgent need to improve their skills. Journal articles have their own specific genre and unique group of words. Likewise, even each section of a journal article has its own genre, and it has turned out to have different sets

of words. Therefore, it seems reasonable to suggest that English teachers should choose different learning objectives and target words so that learners will be able to read different sections well. As for the different types of words I. S. P. Nation (1990, 2001) mentioned, the AWL and GSL contain important words to know. In fact, words in the AWL appear far more commonly in all sections of journal articles than they do in regular academic texts, so the need for learning those words cannot be emphasized enough and a fair amount of time should be spent in class on them. Also, English teachers should make sure that English learners know not only the common meanings, but also specific meanings and concepts of the words in the AWL list and technical words. Additionally, since corpora have made language investigation easier (Hunston, 2002), language teaching professionals should make use of them to expose learners to a variety of word usages and even to word collocation, with this data-driven approach.

References

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238.

Coxhead, A., & Nation, P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes*. (pp. 252-267). Cambridge: Cambridge University Press.

Dudley-Evans, T., & St. John, M. J. (1998). *Developments in English for Specific Purposes*. Cambridge: Cambridge University Press.

Flowerdew, J. (2001). Using small corpora to investigate learner needs: Two vocabulary research tools. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice*. (pp. 68-92). Amsterdam: John Benjamin.

Flowerdew, J., & Peacock, M. (2001). Issues in EAP: A preliminary perspective. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for Academic Purposes*. (pp. 8-24). Cambridge: Cambridge University Press.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, M. A.: Heinle & Heinle.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, P. (2001). Using small corpora to investigate learner needs: Two vocabulary research tools. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice*. (pp. 31-45). Amsterdam: John Benjamin.

Nation, P., & Coxhead, A. (2002). *Range and frequency*. [Computer software and manual] Retrieved from http://www.vuw.ac.nz/lals/staff/Paul_Nation.

Noguchi, J. (Ed.). (2000). *ESP no Riron to Jissen* [*English for Specific Purposes: theories and practices*]. Tokyo: Sanshusha.

Pickard, V. (1995). Citing previous writers: What can we say instead of 'say'? *Hong Kong Papers in Linguistics and Languages, 18*, 89-102.

Robinson, P. (1991). *ESP today: A practitioner's guide*. Hertfordshire, England: Prentice Hall.

Sutarsyah, C., Nation I.S.P. & Kennedy, G. (1994). How useful is EAP vocabulary for ESP—a corpus-based case study. *RELC Journal, 25*, 34-50.

Thurstun, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes, 17*, 267-280.

Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language, 12*, 309-323.

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

Appendix 1

The List of Journal Articles Investigated for This Study (chronological order)

Fotos, S. (1994). Integrating grammar instruction and communicative language use through grammar conscious-raising tasks. *TESOL Quarterly, 28*, 323-352.

Klingner, J. K., & Vaughn, S. (2000). The helping behaviors of fifth graders while using collaborative strategic reading during ESL content classes. *TESOL Quarterly, 34*, 69-98.

Izumi, S. & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly, 34*, 239-278

Riney, T. J., Takada, M. & Ota, N. (2000). Segmentals at global foreign accent: The Japanese flap in EFL. *TESOL Quarterly, 34*, 711-738.

Oh, S. (2001). Two types of input modification and EFL reading comprehension: Simplification verses Elaboration. *TESOL Quarterly, 35*, 69-96.

Young, C. S. & Young, G. A. Jr. (2001). Predictors of mainstream teachers' attitudes ESL students. *TESOL Quarterly, 35*, 97-120.

Ellis, R. Basturkemen, H. and Loewen, S. (2002). Preemptive focus on form in the ESL classroom. *TESOL Quarterly, 35*, 407-432.

Chu, H. J. (2003). Cultural representations for rhetorical conventions: The effects on reading recall. *TESOL Quarterly, 36*, 511-542.

Han, Z. H. (2003). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly, 36*, 543-572.

Panova, I. & Lyster, R. (2003). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*, 573-596.