From analysis of error in learner writing to customized EAP instructional materials

Richard Smith
International University of Japan

Abstract

This paper presents reasons why EAP programs in Japanese universities should consider lexical needs analysis that is based on an analysis of error in learner writing within the university rather than needs analyses that rely solely on the word frequency counts and concordances of native corpora. The major reason is the time constraints faced by many EAP programs in Japan. While learner corpora have recently appeared, the paper argues that these corpora lack sufficient size and segmentation or sufficient analysis of error to allow them to be a substitute for a lexical needs analysis in a particular EAP setting in Japan. The university setting in which the analysis of error in learner text was conducted is described, and key features of its analysis and findings are presented. The relation between the findings and the decisions about the selection and use of customized instructional materials is explained, and some examples of these materials are presented.

1. INTRODUCTION

English for academic purposes (EAP) instruction at universities in Japan often faces time constraints because it tends to operate at the margins of core content programs. There is considerable evidence that acquiring "productive" (Nation, 2001: 24-30) vocabulary knowledge that is necessary for the composition of effective written or spoken text requires more time and effort than acquiring "receptive" vocabulary knowledge that is necessary for effective reading and listening (Nation, 2001: 30-33); this evidence suggests that the time constraints on EAP instruction at Japanese universities are particularly acute for the "productive" vocabulary knowledge development. One of the main goals of EAP programs in Japan and other countries is helping learners compose clear and coherent written text; however, since few students in EAP programs at Japanese universities are above intermediate level and since instruction time is limited, there is considerable pressure on the programs to develop efficient ways of reaching this goal. Helping learners achieve this goal in a relatively short period of time necessarily involves focusing learners' attention on critical skill and knowledge sets; with respect to productive vocabulary knowledge, the critical knowledge concerns the ideas that the learners in the EAP program are expected to present and develop for the written genres and task types that the program regards as predominant within the university. Identifying this critical productive vocabulary knowledge should start with a database. However, the decision about what kind of database should be used is not a simple one.

This paper begins by discussing two types of lexical data base that have drawn considerable attention in recent years: native corpora and learner corpora. After identification of some of the drawbacks of these two types of database for EAP programs in Japan, a different kind of corpus that has a narrower function is described. The "corpus" is a collection of learner created texts compiled within a single university for the sole purpose of creating instructional and reference materials (henceforth summarized as "instructional materials") that help learners to avoid or to repair errors related to lexis in their writing. In referring to this collection of texts, the phrase, "learner corpus," will be avoided since it implies that the text collection and any related text analysis exists in the form of an electronic database. The learner texts and their related text analysis in this collection comprise photocopies that were made after comprehensive error feedback had been written on the original paper texts by the instructor. Following the description of the learner text collection, the data analysis stages that led from the initial identification of lexis-related errors to the identification of the more frequent and salient among these errors are described. Finally, the criteria that were used to prioritize the conversion of the error findings into instructional materials for use in the university's EAP program are identified and explained.

2. NATIVE CORPORA AND LEARNER CORPORA

2.1 Native Corpora

In recent years, studies of native corpora have begun to replace the intuitions of textbook and syllabus writers about patterns of L1 language use. Besides providing concordances of these use patterns, their major contribution to L2 learner lexical needs analysis has been the identification of the frequency of occurrence of words in L1 text. Examples of these native corpora and associated word frequency lists include the corpus from which West (1953) compiled the General Service List (GSL), the British National Corpus (BNC) (see *http://www.natcorp.ox.ac.uk*), the Bank of English (see *http:// www.titania.bham.ac.uk/docs/ svenguide.html*), which is the database for the COBUILD dictionary project, and The Michigan Corpus of Academic Spoken English (Simpson, Briggs, Ovens, & Swales, 2002). A very useful development for EAP programs has been the publication of the Academic Word List (AWL) by Averil Coxhead (2000); the list comprises 570 word families that occur frequently across a wide variety of academic textbooks and texts and that are not represented among the two thousand words of the GSL. Combining the AWL with the most frequent words in the general word lists results in a list that represents a large

majority of the word occurrences in academic texts. For example, the GSL and the AWL combined covers 85.5% and the newer BNC 2000 (the 2000 most frequent words) and AWL combined covers 86.5% of all word occurrences in academic texts (Nation, 2004*: 8).* A more specialized academic corpus exists in the form of the British Academic Written English Corpus (see *http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/*), which comprises just over 2,500 pieces of academic writing by British university students.

The impact of these native corpora on lexical needs analysis and lexical materials development has been growing steadily during the last decade. Vocabulary testing instruments designed for public domain use are now based on the frequency counts obtained from these corpora (e.g., Daller, Milton, & Teffers-Daller, 2007). Two academic English vocabulary textbooks published in recent years are *Academic vocabulary in use* (McCarthy & O'Dell, 2008) and *Focus on vocabulary* (Schmitt & Schmitt, 2005). The former is based on various native corpora, the AWL, and one learner corpus. The latter derives its word selection from the AWL and its concordance examples from the New Longman Corpus. There has also been a general impact on the scholarly orientation of EAP scholars and practitioners. For example, in the 2005 and 2006 issues of the *Journal of English for Academic Purposes*, 40% of the articles concerned corpora, most of which were native corpora; this high percentage was a result of steady increase in the number of such articles over the previous two decades (Thompson, 2006: 248).

2.2 Learner Corpora

Another exciting and even more recent development has been the compilation of learner corpora. These offer the promise of balancing the focus on frequency counts that is associated with the native corpora with a focus on learner error and other developmental features of learners' acquisition and use of lexis. The largest among these corpora are multi-lingual collections of learner-generated written text. Among the largest are the Cambridge Learner Corpus (see *http://www.cambridge.org/elt/corpus/learner_corpus2.htm*), which comprises 20 million words, the Longman Learners' Corpus (see McEnery, Xiao, & Tono, 2006: 66), with 10 million words, and the International Corpus of Learner English (see Granger, 2003), which has 3.7 million words. Monolingual collections include the Hong Kong University of Science and Technology's (HKUST's) Corpus of Learner English (see McEnery, Xiao, & Tono, 2006: 67), with 25 million words, and the Japanese EFL Learner (JEFLL) Corpus (see *http://jefll.corpuscobo.net/*), with one million words.

These corpora and other smaller learner corpora have attracted considerable research interest (Gilquin, Granger, & Paquot, 2007). Analyses of various types of learner corpora have yielded some useful insights into learner writing, including learners' use of lexis. Contrastive interlanguage analysis (Gilquin, 2000/2001) has been used to identify L1 influence on patterns of lexis as well as on other features of learners' writing. Several studies (e.g., Flowerdew, 1998; Granger & Tyson, 1996) regarding learners' use of connectors have been conducted. Hyland and Milton (1997) published an influential article on Cantonese speakers' control or lack of control of hedging expressions. The studies of corpora compiled from the texts of multinational groups of learners also show that learners from different L1 backgrounds share common developmental linguistic features while differing in other linguistic features (Gilquin et al., 2007: 323). Nonetheless, despite Flowerdew's (2001) plea that learner corpora should be used to balance native corpora for syllabus and materials design, there have been few applications of data from learner corpora to textbooks or other published materials. *Longman's dictionary of common errors* (Turton & Heaton, 1996) and the *Cambridge learner's dictionary* (2004), which includes learner error notes taken from the Cambridge Learner Corpus, stand out as two of the few published attempts to collate and codify the information from a learner corpus.

3. DRAWBACKS OF THESE CORPORA FOR EAP PROGRAMS IN JAPAN

3.1 Drawbacks of Native Corpora

A significant drawback of the native corpora for many EAP programs in Japan is that it is difficult to construct from them, or from their textbook products, the lexical components of an EAP syllabus that can be covered within significant time constraints. Systematic coverage for productive purposes of just the 570 word families in the AWL would require a lot of time. In the Japanese university context, only undergraduate English major programs that have a four-year or three-year horizon and frequent class sessions would find this approach feasible.

The alternative to a classroom-centered approach would be a student-centered approach in which the students took responsibility for their own lexical learning outside the classroom. This would relax the time constraints, but it would encounter new difficulties regarding information access and management. The academic vocabulary textbooks based on the native corpora frequency counts have to squeeze in a large number of words within their covers; for example, *Academic vocabulary in use* covers about 1,400 lexical items, which means that information about their use tends to be thin: few words attract more than a single

in-context example. Concordances that add detail to such thin information can be difficult for learners, particularly lower proficiency learners, to utilize efficiently and effectively. While a systematic attempt has been made to alleviate the learners' access and management problems in the area of lexical information necessary for improving reading skills (Barker, 2008), alleviating the same problem in the area of lexical information necessary for improving writing and speaking skills has not yet been attempted.

The final drawback of the native corpora for EAP lexical needs analysis in Japan is the growing consensus among applied linguists that native models of language behavior may not necessarily be appropriate or efficient for the English speakers and learners in different parts of the world who do not use English primarily to communicate with native speakers (Jenkins, 2006; Seidlhofer, 2007). McKay (2003) has pointed out that one of the lexical implications of teaching English to learners who are disconnected from, or uninterested in, the cultural context of the native models is a need to consider critically the value for these learners of stock phrases and idioms whose use value outside their native culture of origin is marginal. This observation applies with particular force to EAP students in Japan, for most of whom the goal of studying English is an instrumental one of communicating clearly for a defined purpose to a global academic audience.

3.2 Drawbacks of Learner Corpora

Because they contain information about the developmental features of learner lexis such as lexical error, L1 influence, and lexical absence, learner corpora hold the promise of helping EAP practitioners to identify the degree of difficulty that learners face in using appropriately and accurately important lexis such as the lexis in the AWL. At this stage, however, the learner corpora are too raw and too idiosyncratic to be of major value for EAP lexical needs analysis in Japan. Only one of the corpora identified above, HKUST's Corpus of Learner English, has been error tagged, but the tagging is incomplete, and the texts are not representative of Japanese university student writing. The other text collections are also unrepresentative of Japanese university student writing; they tend to be collections of examination scripts that have been produced by international students from several countries who have a wide variety of proficiency levels. The two collections that are exceptions to this generalization are the International Corpus of Learner English, which comprises non-examination texts produced by advanced level students who represent twenty-one different L1 backgrounds, and the JEFLL collection, which consists of texts produced by Japanese middle and high school learners. However, these two collections have only a limited

relevance to EAP pedagogical applications in Japan. Neither collection is error tagged, and neither collection represents intermediate level writers.

4. ANALYSIS OF ERROR IN LEARNER WRITING AT THE UNIVERSITY

In the absence in the public domain of relevant, analyzed, and accessible learner corpora, the author decided to conduct a needs analysis by analyzing the errors that were related to lexis in the writing of the learners at his university. While the analysis might lack the comprehensiveness and some of the rigor of a full-scale and formal learner corpus compilation and analysis, it would have the advantage of obtaining data about errors related to lexis from writing samples that directly represent the writing produced by the whole cohort of learners at the university. For the purpose of this study, "lexical errors" were defined as errors that relate firstly to word choice, word meaning, collocation, and register constraints and secondly to the grammatical patterns in which words occur and to non-register constraints on the use of words such as transitive and intransitive verb constraints. This definition draws heavily on Paul Nation's identification of "what is involved in knowing a word" (Nation, 2001: 27). For the sake of convenience, from this point in the paper lexical errors that inhabit a lexical dimension (word choice, meaning, collocation, and register) are called "lexical errors" (LEs) and lexical errors that inhabit two or more dimensions (lexical *and* grammatical or idiomatic) are called "errors that have a lexical dimension" (ELDs). Although the needs analysis would eventually involve the compilation of a paper collection of over a thousand pages of learner generated text, the aim of this exercise was not to create an analyzed "corpus" to which the learners would have access. The time constraints faced by the EAP program and by the learners and observation of the difficulties that the university's learners have experienced in accessing, managing, and making sense of public domain corpus and concordance information led to the decision that the aim should be to create instructional materials that distilled the most important information from the analysis of LEs and ELDs in the text samples. These instructional materials would be designed to help learners either to avoid lexical error in their writing or to repair the lexical error in subsequent writing drafts.

Before the description of the data, methodology, and findings of this error analysis, some information about the learners and their institutional setting is provided.

4.1. The institution

The author's educational institution is the International University of Japan, a two-year graduate-only English-medium university in Japan. The university confers degrees in five

social science subject disciplines, which range from International Relations to International Management. The student body is both small and international: among the 160-170 students who join the university each year, between 75% and 85% come from outside Japan, mostly from countries in South East Asia. While a majority of the students are exempted from the requirement to enroll in EAP classes, a significant minority who fail to satisfy the exemption criteria are required to enroll in the classes. A large majority of these enrolled students can be categorized as having general English proficiency levels that are below high intermediate level: their test scores are in the 450-550 TOEFL and IELTS 4.0-6.0 ranges. There are three types of time constraints on the EAP classes: firstly, the enrollment requirement only applies to first-year students; secondly, because of resource and other constraints, the more proficient among the enrolled students are granted exemption at the end of each term; thirdly, the maximum duration of the EAP program is 135 hours of class sessions.

5. DATA AND METHODOLOGY

5.1 The Database

The analysis of error that is described in this paper comprises photocopies of 1,300 pages, or just over 200,000 words, of paper text written by students enrolled in the first-year required EAP classes. The small ratio to words to pages is due to the formatting guidelines issued to the students and to the existence of many pages that contain only a small number of words. All of the texts are first drafts of homework assignments. The text was produced over a period of seven years by two hundred different students who were enrolled in the author's class sections. Since the author's university is small, the EAP program is centrally coordinated in detail, and each class section represents a cross-section of the EAP student body, these texts are representative of the writing of the EAP student body. All of the texts that were collated and analyzed were written in response to three task types: data commentary, summary, and essay. The data commentary and summary tasks were based on source data and texts drawn from the five academic subject disciplines taught at the institution; the essay tasks involved expository writing in response to *cause and effect*, *comparison and contrast*, and *argumentation* prompts.

The photocopies of the first draft texts comprise not only the original text but also the author's comprehensive feedback on the language use in these first drafts. The provision of comprehensive error feedback is a standard part of the writing feedback process within the author's EAP program. The bulk of this feedback is in the form of editing abbreviations that

identify language use errors and instances of language use awkwardness. The error abbreviations are standard abbreviations that have been adapted from the error abbreviations presented in the textbook, *Writing clearly: An editing guide* (Lane, & Lange, 1993: xx-xxi).

5.2 Methodology: Four Stages of Error Analysis

The error analysis comprised four stages. In Stage 1, over a period of seven years the author accumulated the portfolio of first draft texts to which error feedback had been added. In Stage 2, the initial error tagging was elaborated and refined in order to make it more amenable to the collation of LEs and ELDs. In Stage 3, a small group of readers was asked to make judgments about the salience of a sample set of these errors. In Stage 4, an attempt was made to count the frequency of the LEs and the ELDs. All of these stages were designed to provide important information for the implementation of the final goal: the creation of customized instructional materials.

The photocopying of these annotated texts was originally designed to be part of an administrative exercise. When the exercise came to a natural end after five years, the author decided that the accumulated photocopies represented a valuable database that should not be thrown away and should be augmented. As a result, Stage 2 of the error analysis began in a tentative way in the sixth year; formal analysis began in the seventh year, at which point the accumulation of annotated text copies stopped.

At the beginning of Stage 2, the error abbreviations that identify LEs and ELDs were subdivided when this was thought to be helpful for the error analysis. These error abbreviations most closely related to errors that have a lexical dimension are the following: conn - incorrect or missing connector; wc - wrong word choice; wf - wrong word form; wo - incorrect or awkward word order; nonidiom - not expressed this way in English; and, in some cases, ss – which means a sentence structure problem. Errors that were designated “conn” errors were subdivided into clause connector errors and noun phrase connector errors. Clause connectors tend to be associated with grammatical structures that have weak lexical co-occurrences; they are also treated in detail by the writing and grammar textbooks assigned to the EAP courses at the university. For these two reasons, they were excluded from further analysis. Errors that attracted the “wc” tag were subdivided into word meaning and collocation errors. Errors that attracted the “nonidiom” tag were divided into word choice, lexico-grammar, and fixed expression errors. These new tags were added to the photocopied texts in pencil. A new tag, “vtv,” was added in pencil for errors that concern the semantic relation between verb transitivity and volition in certain classes of English verbs such as

verbs that express change. An example of such an error is the following sentence, "The government increased inflation." Some of the more purely grammatical error tags such as "ss" (sentence structure) were reviewed and marked as lexico-grammar errors if the co-occurrence of the structure and the lexis was considered significant. An example of this re-designation is "ss" tags that involve the classes of adjectives that take *it* as an anticipatory subject (e.g., "It is essential for the government to eliminate corruption.").

An important outcome of Stages 1 and 2 was the recognition that a lot of error identification is a subjective process. While a clear majority of the word choice ("wc") errors were unambiguously word meaning or collocation errors for which the solutions were word choice replacements, many of the LED errors could yield more than one plausible error identification. A representative example concerns the re-designation of "ss" error tags as lexico-grammar error tags. The error sample, "The government is important to raise taxes," was originally tagged as an "ss" error with an additional marginal note directing the writer to check "important" in a good learner dictionary. Alternatively, the error could have been construed as a purely lexical error if the words, "is important," had been tagged as word choice errors for which a verb such as "needs" would represent a correct replacement. An impartial linguistic analysis would have recognized both of these interpretations and perhaps one or two others. In such cases, the author abandoned objective linguistic analysis in favor of pragmatic analysis that was oriented to favoring the error solution that the author considered the most elegant and accessible to the learner. The error sample, "the government is important to raise taxes," was therefore tagged as a lexico-grammar error, and the eventual "solution" in the customized pedagogical materials is presented as a lexico-grammar solution (the classes of adjectives that co-occur with *it* as an anticipatory subject). The author's judgments about the elegance and accessibility to the learners of different solutions were aided by frequent discussions with the writers about these error issues during individual writing conferences; these conferences were regularly scheduled between the submissions of the first and second drafts.

In Stage 3, four colleagues of the author, who comprised two members of the university's EAP program and two members of the university's content programs, were asked to make judgments about the salience of all the errors that were tagged in a randomly selected sample of 75 pages of text. The aim of this exercise was to give some weight to error significance as well as to error frequency in the compilation of the customized instructional materials. The readers were asked to classify the tagged errors according to three sets of descriptors: (1) "the error did not impede my comprehension of the tagged words and the

surrounding text;" (2) "the error forced me to read the tagged words and the surrounding text more than once before I was confident that I had comprehended them correctly;" and (3) "after several readings, I was still unsure about the meaning of the tagged words and/or the surrounding text." The relatively small sample of 75 pages reflected the limited amount of time that the four volunteer readers were able to dedicate to the exercise. Those LE and ELD errors that were consistently classified as members of the third category were prioritized for treatment in the customized instructional materials. Examples of third category errors that were salient yet not very frequent include the use of vague summary nouns such as *point, condition,* and *thing* and preposition choice errors in change expressions such as the confusion of *to* and *by* (e.g., "The budget deficit increased to 15%.").

In Stage 4, an attempt was made to identify the frequency of the errors as error frequency would be one of the criteria used to prioritize the creation of the customized instructional materials. It soon became clear that counting errors in a paper database manually is a laborious and time-consuming task. Another constraint on error counting concerned the identity of the error. While a majority of the LE errors were countable because they were single word errors that had a clear identity, many of the more complex ELD errors could not be identified with a particular word. For example, ELD errors such as "the two differ between themselves" can be resolved elegantly by the solution, "the two differ <u>from</u> <u>each other,</u>" yet "differ," although used correctly, appears to be contributing in a significant way to the genesis of the adjacent errors. Because of the time constraint, the author decided to conduct counts of error frequencies in a sample 200 pages of text selected to represent the three main writing genres. Because of the second constraint, the counts of the more complex ELD errors were simplified by labeling errors that recurred in certain word strings as members of a heuristic category; in the case of the error sample cited above, it was arbitrarily assigned to a heuristic category identified as "similarities and differences." A small number of the LE counts obtained from the sample text were checked by counting the error frequency in the entire database. For example, the LEs that were classified as "change" LEs (e.g., *raise, rise, reduce, fall*) amount to 510 errors in the total database of just over 200,000 words, a high number that is consistent with the sample count finding. A common but less frequent LE, the use of "mention" instead of "state" or "discuss," occurs 75 times in the total database, which is a ratio similar to the ratio in the sample count. This description of the error "counting" process serves to underline the fact that error frequency identification that is conducted under these conditions and for instrumental (as opposed to disinterested) purposes is a pragmatic art whose results may not be generalizable outside the institution.

6. ERROR ANALYSIS FINDINGS

The most general finding is that instances of obvious L1 interference were limited to Japanese learners' use of *katakana* expressions and to Indonesian and Chinese learners' difficulties with verb tenses, the latter of which are not a lexical issue. Probably because East and South East Asian mother tongues have no cognate relationship to English, the learners tended to experience difficulties with the same lexical issues even though the specific errors sometimes varied. This aspect of the learner error analysis suggested that it would be possible to develop pedagogical and reference materials that would be relevant to all of the learners.

The more specific findings can be divided into two groups: (1) findings that are accessible in the research literature and in published pedagogical materials; (2) findings that are harder to find in these two domains.

Examples of findings that have been presented in the research literature include the following:

- collocations (Sinclair, 1997)
- absence or underuse of hedging expressions (Hyland & Milton, 1997)
- informal register (Granger & Rayson, 1998)
- adverbial connectors (Altenberg & Tapper, 1998)
- summary nouns and text cohesion (Hoey, 1991: 69-70)

Pedagogical treatments of all these issues can be found in widely available publications. Collocations are treated in detail in a number of specialist learner dictionaries such as the *Cambridge advanced learner's dictionary with CD ROM*. The other issues are covered in mainstream EAP textbooks such as *Academic writing for graduate students* (Swales & Feak, 2004: 27, 32-38, 125-134) and *Academic vocabulary in use* (McCarthy & O'Dell, 2008: 28-33, 82-83, 116-119).

Findings that are less accessible in the literature and in published textbooks include the examples below. These sorts of findings were prioritized. The symbol [LD] indicates that the item caused the learners to commit frequent or fairly frequent errors. The symbol [RD] indicates that the error that resulted caused the reader significant comprehension difficulty.

- the inconsistent semantic relation in English between verb transitivity and volition; in some semantic areas such as "change" there is a consistent relation between the two (~~*The government increased inflation.*~~ *Inflation increased*.) [LD] [RD]

- the lexico-grammar of certain classes of adjectives that use *it* as an anticipatory subject (e.g., ~~*My country is difficult to attract FDI*~~*. It is difficult for my country to attract FDI*. ~~*The government is essential to reform the tax system*~~*. It is essential that the government reform the tax system*) [LD]
- in the published materials, notably dictionaries, collocational patterns tend to be listed in order of frequency of occurrence in a native corpus; other orders of presentation may be optimal for learners; for example, the verbs that collocate with the noun *strategy* can be better understood by learners if they are presented in a time order that begins with *thinking about a strategy* and ends with *abandoning the strategy* [LD]
- the misuse of verbs that have academic discourse functions – e.g., *investigate, analyze, examine, identify* [LD] [RD]
- misuse of prepositions in the expression of some abstract concepts such as "change" – e.g., "The exchange rate has decreased in 15% since last year." [RD]

7. FROM ERROR ANALYSIS TO INSTRUCTIONAL MATERIALS

7.1 Two Prioritization Principles

Two principles guided the decisions about which errors should be prioritized for treatment in the customized instructional materials. The first principle was that the most frequent and most salient errors should first be considered for treatment. The second principle was that error areas that the author regarded as being adequately covered in the learners' assigned textbooks and dictionaries and other accessible public domain materials should be assigned a low treatment priority. In effect, it was decided that the learners would be delegated responsibility for finding out the meaning and use information about words that were "easier" either because in the learner database they appeared to be relatively non-problematic or because the learners already had unimpeded access to adequate treatments.

7.2 Three Groups of Prioritized Lexis

In conformity with these two principles, priority in the choice of instructional materials creation has been given to the lexis that tends to cause learner error and is not covered adequately in textbooks, dictionaries, and other public domain materials to which the learners have access. This lexis has been divided into three groups according to pragmatic pedagogical criteria; these three groups are presented in order of treatment priority from the highest to the lowest.

The first group consists of *patterns* of lexis tends to cause learner error that is frequent and/or highly salient and can be addressed efficiently in the classroom. An example of this group is the relation between transitivity and volition in the "change" area of lexis. Other lexico-grammatical issues that are fairly well covered in the published textbooks are added when such an addition is convenient. Thus, the quantifiers, *number* and *amount*, which co-occur often with countable and uncountable nouns when expressing "change" are presented in the same materials that present the relation between transitivity and volition in the "change" area of lexis. This first group is represented by a range of instructional materials including modules for in-class use, practice exercises, review quizzes, and customized online reference materials. The main aim of the materials used in class is to help learners avoid error related to the use of lexis and the main aim of the reference materials is to help learners in their second or subsequent writing drafts repair the errors related to the use of lexis.

The second group consists of *patterns* of lexis that tend to cause learner error but less frequently and/or with a little less salience than the patterns of lexis in the first group. An example of this group is the lexico-grammar of certain classes of adjectives that use *it* as an anticipatory subject. Because of time constraints, the instructional materials that focus on this area are not explicitly earmarked for intensive classroom work although the instructor may utilize them in class if a critical need is perceived to exist. Instead, the materials are primarily designed as online reference materials to which learners are directed when the need arises. This designation, however, is not a fixed one. If the error is frequent or salient among a particular group of learners, the instructor can bring the materials into the classroom.

The third group consists of lexis that tends to cause LEs that can be treated in relative isolation from adjacent LE and ELD issues. Examples of this group include specific collocational issues and lexical pairs that tend to cause LEs (e.g., *adapt* and *adopt*). While some of the instructional materials are incorporated into syllabus packets that are designed for in-class or homework use, most of them exist in the form of customized online materials to which the learners are directed when a specific and immediate need arises. Thus, they tend to be utilized for error repair rather than for error avoidance.

Small extracts that illustrate the instructional materials created for these three groups of lexis can be found in Appendices A, B, & C.

7.3 A Virtuous Circle

It has been the author's experience that most EAP learners are motivated by the prospect that they can find solutions to their problems. Creating a wide array of instructional

materials that is targeted at the sorts of lexical and lexis-related problems that learners are likely to encounter represents one of the solutions. The intimate relation between writing, lexis, and lexico-grammar means that solving lexical and lexis-related problems also contributes to the solving of other problems in the learners' writing, which enhances the learners' confidence when they tackle subsequent writing tasks. This virtuous circle represents an optimal environment for learners' productive vocabulary growth.

8. CONCLUSION

This paper makes a case for customizing the design of certain instructional materials within a university's EAP curriculum by analyzing the error in learner writing produced by the learners' predecessors at the university. This argument assumes that the profile of the learners and the conditions under which they create EAP text are relatively stable, which is the case at the author's host university. The reasons advanced in the paper concern the lack of alternatives when there is a need to identify critical lexical information because time constraints are pressing. Error identification and evaluation necessarily involve subjective judgments, and learners' proficiency levels and written genres and task types will vary across institutions; as a result, there will be limits to the extent to which EAP programs in Japan can simply adopt the error analyses and the instructional materials outcomes of other programs. However, EAP programs should not feel discouraged by the perception that a lot of time and effort is required to conduct the analyses and create the materials. The author spent several years on his project because the author worked mostly alone. A collaborative effort among several program members should require months rather than years. Finally, the time and effort invested will yield a double reward because the process will be at least as enlightening for those who conduct it as it will be for the learners who will benefit from its results.

REFERENCES

Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80–93). London: Addison-Wesley, Longman.

Barker, D. (2008). *Learning English vocabulary*. Tokyo: Back to Basics Press.

Cambridge advanced learner's dictionary with CD ROM (3rd ed.) (2009). Cambridge: Cambridge University Press.

Cambridge learner's dictionary (Japanese semibilingual version (2nd ed.). (2004). Tokyo: Cambridge University Press and Shogakukan Inc.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Daller, H., Milton, J. & Treffers-Daller, J. (Eds.). (2007). *Modelling and Assessing Vocabulary Development.* Cambridge: Cambridge University Press.

Flowerdew, L. (1998). Integrating expert and interlanguage computer corpora findings on causality: Discoveries for teachers and students. *ESP Journal, 17*(4), 329–345.

Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 363–379). Amsterdam: Benjamins.

Gilquin, G. (2000/2001). The integrated contrastive model: Spicing up your data. *Languages in Contrast, 3*(1), 95–123.

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes 6*, 319–335.

Granger, S. (2003). The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly 37*(3), 538-546.

Granger, S., & Rayson, P. (1998). Automatic lexical profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119–131). London: Addison-Wesley, Longman.

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes, 15*, 9–29.

Hoey, M. (1991). *Patterns of lexis in texts*. Oxford: Oxford University Press.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6*(2), 183–205.

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157-181.

Lane, J., & Lange, E. (1993). *Writing clearly: An editing guide* (1st ed.). Boston, MA: Heinle and Heinle.

McCarthy, M. & O'Dell F. (2008). *Academic Vocabulary in Use*. Cambridge: Cambridge University Press.

McEnery, A, Xiao, R, & Tono, Y. (2006) *Corpus-based language studies: An advanced resource book.* New York: Routledge.

McKay, S. (2002). *Teaching English as an international language: Rethinking goals and approaches*. Oxford: Oxford University Press.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I.S.P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 3-14). Amsterdam: Benjamins.

Schmitt, D. & Schmitt, N. (2005). *Focus on vocabulary: Mastering the academic word list.* New York: Longman.

Seidlhofer, B. (2007). *Explaining English as a Lingua Franca*. Oxford: Oxford University Press.

Simpson, R., Briggs, S., Ovens, J., & Swales J. (2002) *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: University of Michigan Press.

Sinclair, J. (1997). Corpus Evidence in Language Description. In Wichmann, A et al (Eds.), *Teaching and Language Corpora*. (pp. 27-39). New York: Longman.

Swales, J. & Feak, C. (2004). *Academic writing for graduate students*. Ann Arbor, MI: University of Michigan Press.

Thompson, P. (2006) Call for proposals. Special issue on corpus-based EAP pedagogy. *Journal of English for Academic Purposes 5*(3), 248–249.

Turton, N. & Heaton, J. (1996) *Longman Dictionary of Common Errors*. *New Edition.* Harlow: Longman.

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

Appendix A

Examples of instructional materials for the **first group of lexis**.

Printed reference information and class activities:

The lexico-grammar and transitive/intransitive semantics of *change* verbs and nouns: short extracts. *Number* and *amount* in descriptions of change.

Some examples of transitive verbs in sentences (look at the preposition choices!)

The government **has increased** income taxes by 10%
The government **has increased** income taxes from 7% to 10%
The discount store **will reduce** prices by up to 5%
OPEC has decided to **keep** oil prices **at the existing level** of $19 per barrel
The Central Bank **will increase** its baseline interest rate from 4.5% to 5.0%.

Some examples of intransitive verbs in sentences (look at the preposition choices!)

Interest rates in general **will rise** by 0.5%
Because of deflation, prices **will fall** by up to 5%
Oil prices **will remain at the OPEC level** of $19 per barrel

1 *Number* and *Amount*

Examples:
"The **number** of people claiming social security benefits has recently increased."
"The **amount** of money in the economy increased by 3% during June."
"The money supply contracted last month."

Which of the following nouns have to be preceded by *number* or by *amount* when the topic is **quantitative** change?

- children
- money
- investment in bonds
- consumers
- population
- money invested in bonds
- efficiency
- units produced per man hour
- coins
- consumption

2 Distinguishing <u>Purposes</u> from <u>Results</u>

2a What's the difference in meaning between sentence (a) and sentence (b)? Which sentence best represents economic reality?

(a) The economy's growth rate has increased.

(b) The government has increased the growth rate.

Appendix B

Example of instructional materials (online information) for the **second group of lexis**.

The lexico-grammar of certain classes of adjectives that use *it* as an anticipatory subject (an abbreviated sample):

***It* + *be* + adjectives which can be followed by a *that* clause.**

In these two adjective groups, the verb in the *that* clause does not require *should,* but the sentence usually begins with *it*.

Group c: most of these adjectives are emotive, including the emotive adjectives ending in *ing* :

annoying, curious, depressing, disappointing, irritating, logical, shocking, surprising etc.

"**It** is **annoying that** we have to pay higher taxes for better government services."

Group d: these adjectives relate to truth or knowledge

apparent, certain, clear, evident, likely, obvious, possible, true, well-known etc.

"**It** is **likely that** the prime minister will win the election."

In this adjective group, the verb in the *that* clause requires *should* or the subjunctive verb and the sentence usually begins with *it*.

Group e: these adjectives relate mainly to modality

appropriate, crucial, essential, imperative, important, necessary, vital etc.

"**It** is **essential that** the company market [subjunctive verb] its products immediately."

"**It** is **essential that** the company **should** market its products immediately."

Alternatively, the following sentence pattern can be used:

"**It** is **essential for** the company to market its products immediately."

Appendix C

Example of instructional materials (online information) for the **third group of lexis**.

Collocations: ***strategy***. The collocations are presented in a time order that begins with the birth of a strategy and ends with its death.

---

strategy (n) Although this noun is simple, please note the types of verbs that are collocated with it. These verbs are listed in the same order they might be used from the start of a strategy to its finish.

- **think about a strategy** or **borrow/adopt a strategy**
- **outline the strategy**
- **develop the strategy**
- **formulate the strategy, present it to others and discuss it**
- **set out the strategy** = **formulate the strategy and present it to others**
- **revise the strategy**
- **finalize the strategy**
- **implement the strategy/ put the strategy into practice / translate the strategy into reality**
- **make adjustments to the strategy**
- **make radical changes to the strategy**
- **phase out the strategy / abandon the strategy**

Note that native speakers do NOT write:
~~**make**~~ **a strategy**
~~**do**~~ **a strategy**

Also note that the same types of verbs can be collocated with nouns which have similar meanings: **plan**, **approach**, **program**, **recommendations**