Using a Learner Corpus to Develop Learner Language Profiles

Nicolai Struc
Nicholas Wood

Abstract:

This paper describes the development and subsequent analysis of a cross-sectional learner corpus comprising written samples from students in the first, second and third years of the English language writing program at Reitaku University, Japan. Students completed two writing tasks, a narrative and an argumentative essay, which were then analyzed to determine whether gains were made in the areas of fluency, lexical richness, grammatical accuracy and use of rhetorical/cohesive devices. Gains were observed in all of these four areas, the process of annotating the raw data revealed areas which were problematic for student writers but were unelaborated by the analysis employed in this research. These observations illuminate the limitations of this study and highlight directions for further analysis.

Introduction

Background

While the field of corpus linguistics has proven invaluable for

linguistic inquiry, the development of the sub-field of learner corpus studies has only more recently been utilized as a tool by SLA researchers and language teaching professionals (see McEnery , Xiao & Tono, 2006; Granger, 2004). As Granger (2004) has noted, the principal applications of corpus-derived learner language data are in materials and syllabus design and classroom methodology. As the technology for analyzing large amounts of electronic texts has become more accessible through the increasing availability of software such as Wordsmith Tools, Antconc, Range etc. it has become easier for small scale research with modest funding to endeavor in corpus-based inquiries to these ends.

The aim of the present study has been to compile a corpus comprising writing of students at three points in a university-level L2 English writing program in order to reveal a profile of the learners' writing in terms of fluency, lexical richness, grammatical accuracy and use of cohesive devices. It was hoped that the analyses and resulting profiles would reveal differences in language use in these broad areas of writing, and thereby inform ongoing development and adjustment of the writing curriculum.

Granger (Granger & Petch-Tyson, 2002), in surveying the current state of learner corpus studies, notes that its primary function has been to describe learner language and the subfield itself can be described as a "linguistic methodology which is founded on the use of electronic collections of naturally occurring texts, viz. corpora" (p.4). The analysis often focuses on frequency and patterns of distribution. Studies such as Hinkel (2002) have comprehensively described the writing of learners of English from a variety of L1 backgrounds (e.g., Japanese, Chinese and Korean) and compared the frequency and distribution of elements of writing of these nonnative speaker (NNS) groups with those of native speakers (NS).

In addition to comparing NS/NNS language, corpora have also been used to examine the differences between groups of NNS of the same L1 at different skill levels in order to illuminate patterns of second language

acquisition. The development of these types of corpora has been evident in a number of learner corpus projects in Japan, including Tono's JEFLL Corpus (see Abe & Tono, 2005) of Japanese junior high school, senior high school and university students and the SILS Learners' Corpus at Waseda University (see Muehleisen, 2006). Abe and Tono's (2005) study sought to investigate patterns of grammatical accuracy across proficiency levels, while one application of the SILS corpus is to study the development of students' writing, examine specific vocabulary and grammar use to provide a research-driven approach to the creation of course materials. Millar and Lehtinen (2008) have developed a learner corpus at Kanda University of International Stodies to highlight L1 transfer and lack of awareness of academic writing conventions.

The present study attempts to reveal a profile of various characteristics of learner writing by collecting samples of student writing at various points in the writing program. Samples were collected from first year students preceding the onset of writing instruction, second year students preceding the onset of the second year of instruction, and third year students, again, preceding their third year of writing instruction. While the total corpus consists of all the collected data, the sub-corpora of first (29,213 words) and third year students (29,488 words) are compared here as a starting point, reserving the second year data for future inquiry. The entire portion of corpus used in the present study, thus consists of 58,701 words.

In order to sample a variety of writing, two writing samples were elicited from each student: the first a narrative, and the second an argumentative essay. These writing samples were collected under standardized conditions in order to ensure comparability. These samples were then submitted to analysis either in combination or separately depending on the particular focus of inquiry.

The research questions, which guide the interrogation of the learner corpus, are as follows:

1. Fluency: Does the quantity of writing produced before and after two years of writing instruction change? If so, by how much?
2. Lexical richness and sophistication: Does the lexical richness in different vocabulary frequency levels brought to bear on the writing tasks before and after two years of writing instruction change?
3. Grammatical accuracy: Does the grammatical accuracy in past tense/aspect in writing produced before and after two years of writing instruction change?
4. Textual cohesion: Does the frequency and distribution of use of sentence connectors in writing produced before and after two years of writing instruction change?

Fluency and Lexical Richness

In order to address fluency, the present study employs a comparison of the token counts after equalizing the writing samples provided by first and third year writers to determine whether an increase was evident. Lexical richness, is the second, and more complex issue, which is approached by comparing the type-token ratios observed in the two writing samples of first and third year students and further subdivided into lexical richness in the General Service List (GSL) first and second thousand most frequent words (West, 1953) and Academic Word List (AWL) (Coxhead, 2000) as measured by Nation's RANGE software (Nation, 2002). Laufer (1994) has pointed out that lexical quality and writing quality are interconnected and cited Astika's research which showed that among various components of writing, including content, organization, language use and mechanics, vocabulary accounts for 83.75% of variance in holistic scores. Laufer's research examined writers' Lexical Frequency Profile (comprising the proportion of words in three categories: GSL 1, GSL 2 and the University Word List) and lexical variety as measured by type-token ratios. Similarly, the present study sought to reveal lexical profiles for first and third year writers and compares the lexical richness in three lexical

frequency groups. As Daller and Phelan (2007) point out, this type of word-list based approach can illuminate degree of lexical sophistication in addition to lexical diversity or richness.

There has been discussion of which words should comprise General Service wordlists (Nation & Kyongho, 1995) or what the content of an academic word list should be (Hyland & Tse, 2007). However, when subjected to corpus analysis, the General Service List of the first 2000 most frequent words (West, 1953) showed coverage of 75% in non-fiction and 90% in fiction texts (Nation & Kyongho, 1995). Similarly, Coxhead's (2000) Academic Word List (AWL) when applied to a 3.3 million word corpus of academic texts from a wide variety of disciplines and showed 10.6% coverage (Hyland & Tse, 2007). Thus, these lists may be applied with a certain degree of confidence when considering productive vocabulary as a measure of lexical richness.

Grammatical Accuracy

For grammatical accuracy, only the narrative writing sample was annotated for correct and incorrect past tense usage and counts were made using Wordsmith Tools 4 (Scott, 2004) and submitted to an obligatory occasion analysis (see Brown, 1973; Pica, 1984). The obligatory occasion analysis technique has been used in many grammatical morpheme acquisition studies (e.g., Andersen, 1978; Pica, 1984) and considers learner acquisition as an accuracy rate exceeding 80 or 90% (depending on the study), the percentage indicating the proportion of correct instances of usage to the total number of occasions where use of the target grammatical morpheme is called for. Pica (1984) further develops this by including cases of oversuppliance (i.e., providing the grammatical morpheme when it is inappropriate which provides a more conservative measure of accuracy. While language acquisition per se is not the focus of the present study, the measure of accuracy can still be used in tracking changes in student writing. Only the narrative writing task was submitted to this analysis since fictional narratives typically exhibit use of past tense.

Textual Cohesion

Academic writing can be characterized by the employment of a wide variety of discourse and text-rhetorical features, which among others may include the use of coordinating/subordinating conjunctions, sentence-level conjunctions, exemplification and hedges. As one indicator of textual cohesion, argumentative writing samples were annotated for instances of use of sentence-level transitions to determine their frequency and distribution in first and third year writing. For this aspect of the study, Hinkel's (2002) study and methodology has been both informative and instructive. In a comprehensive literature review, Hinkel highlights the influence of Confucian rhetorical tradition in the writing of Chinese, Japanese and Korean writers. Among these, brevity is valued, the responsibility for textual clarity rests more with the reader than the writer and supporting one's contentions with authority of tradition rather than one's own ideas. Hinkel succinctly contrasts Confucian and Aristotelian rhetoric: "…factual validation of opinions and propositions, logical arrangement of ideas in text to demonstrate their connectivity from one to another and to the text's purpose, and discussion and debate of ideas. None of these foundational constructs of rational argumentation are found in rhetorical traditions outside those based on the Greco-Roman rhetoric and philosophy" (p.33). It is these differences, that perhaps inform and drive the content of EFL academic writing instruction and its emphasis on developing use of sentence-level transitions, hedges and so on.

In Hinkel's study, she notes Davidson's study which asserts that the presence and use of coordinators, among other cohesive devices, can be used as a statistically reliable measure of discourse cohesion in L2 writing assessment. In her analysis of Japanese academic writing, Hinkel observes the frequency and range of sentence-level-transitions, logical-semantic conjunctions and exemplification in Japanese L1 English texts are consistently higher in comparison with their NS counterparts. Indeed, this appears to be true of writers from a wide variety of NNS from different L1 backgrounds and thus may indicate their overuse as a function of

interlanguage and/or an artifact of the curriculum of academic writing courses. In the present study, comparing the frequency or distribution of these cohesive devices with those of NS would likely reveal similar patterns, so instead, the two groups are compared to determine whether writing instruction has contributed to any change in this aspect of writing. Again, for the present study, only the argumentative writing samples were annotated for sentence-level conjunctions, logical-semantic conjunctions and exemplification since the narrative writing task would not require their inclusion to the same degree.

Methodology

1. Population

The Learners' Writing Corpus comprises writing samples elicited from first, second and third year English-major students at the College of Foreign Studies, Reitaku University during their first English Writing Classes of the 2008/09 academic year in April 2008. Placement in streamed classes was determined for first year students by previously achieved TOEIC scores, and for second year students by a combination of past class test results and end-of-year TOEIC scores. Samples of writing were collected from 8 first year classes (191 students), 5 second year classes (144 students) and 4 third year classes (131 students). The groups compared in the present study are the first and third year students. Granger (Granger & Petch-Tyson, 2002) note that institutional status rather than performance on standardized tests is often used to categorize students in learner corpus research. In addition to the year in program, students' TOEIC scores as measured within 4 months preceding data collection is provided below:

Population Data by Year in Program and TOEIC Score

Year in program	Date	N	Mean	SD	Min	Max
1	04/2008	192	356.25	85.41	110	720
3	12/2007	139	498.30	109.84	295	905

2. English Writing Program

The core aim of the English writing program is to equip students with the skills and knowledge to enable them to produce an academic research paper at the end of their third year and a graduation thesis at the end of their fourth and final year. While teachers are given latitude in the choice of text and pedagogic methodology, they are expected to ensure that students achieve specific goals by the end of each semester.

Broadly speaking, the goals for the first year are for students to acquire basic grammatical literacy, familiarity with basic parts of speech, the ability to construct a correct, full sentence, the ability to write a variety of sentence structures (simple, complex and compound), and the ability to organize and write a variety of paragraph types (e.g. cause and effect, comparison and contrast), with the use of topic and supporting sentences, and transitions to aid textual cohesion. Second year goals are to establish the conventions of academic research writing, extend vocabulary, develop sentence complexity, consolidate paragraph construction and textual cohesion, develop the skills to organize and produce a variety of essay types (e.g. descriptive, narrative, process), introduce the use of formal/academic language, and to develop the skills for paraphrasing and summarizing. The overall goals of the third year are to consolidate the skills and knowledge acquired in previous years and to enable students to prepare and produce a 5 to 10 page research paper. In all three years, tuition, course materials and writing are in English.

3. Data Collection

All Writing Class teachers were provided with instructions and materials for administering data collection. The task papers given to each student included (1) a short questionnaire in L1 (Japanese) asking for personal details (gender, age, nationality, home prefecture and first language) and about language experiences (learning, travel and duration of

stays abroad), and (2) prompts in L1 and English to elicit two samples of the student's written English.

Following completion of the questionnaire, students were allowed 20 minutes to write a response to each prompt, the first prompt aiming to elicit a past-tense narrative (Sample 1) and the second to elicit an argumentative piece, with opinions and supporting reasons (Sample 2). The prompts were first conceived of in English but later translated to Japanese and provided to students only in Japanese so that the language in the prompts would not influence the writing. The prompts are as follows:

Prompt 1- Narrative: "Imagine two friends went shopping together last week. One friend returned home happy, the other friend returned home sad. Write a story about what happened. You have 20 minutes."

Prompt 2- Argumentative: "Studying English abroad. Please write reasons for and against studying English in another country. You have 20 minutes."

During the 40 minutes allowed for writing, no speaking was permitted, nor was the use of reference materials such as dictionaries. The same procedure was followed in each of the 17 classes involved.

The questionnaires and samples were anonymous, and the students only required to write their student university numbers. On completion, all the task papers were collected by the teacher and returned to the researchers.

4. Transcription

The collected writing samples were photocopied and separated by Sample (1 or 2), Year and Class. Selected third and fourth year students were employed to digitally transcribe the samples, and in training were instructed to type each sample using a prefabricated Word template, and to code and save it as both a Word file and a Text file in a designated folder on a USB flash drive. The template provided tags for the student number (<head></head>), text (<body></body>) and for each sentence (<s></s>).

Transcribers were also given set protocol for how to render unidentifiable words, spaces left in the text and other features that required editorial attention. The overall aim, however, was for the typed versions to digitally replicate the textual features (apparent sentence organization, spellings, capitalization and punctuation) of the original handwritten samples. The completed digital transcriptions and originals were returned to the researchers to be edited, proofread and tagged.

5. Editing

The digitized samples were proofread (checked against the handwritten originals) by the authors and where necessary amended for accuracy.

6. Annotation/Tagging for Grammatical Accuracy (Sample 1)

The focus of research when examining Sample One (narrative) was to evaluate the ability of language learners to supply correctly inflected past tense verbs. The research aimed to examine not only the accuracy of inflected verbs when correctly positioned in subject-verb constructions (obligatory occasions) but also omission (instances when the suppliance of a verb was obligatory but absent) and oversuppliance (instances when the suppliance of a past tense inflected verb was unnecessary or inappropriate).

Determining numerically the instances of correct suppliance, omission and oversuppliance provided the basis for what Pica (1984) termed "target-like use analysis", and an estimation of learners' proficiency could then be calculated. The aim of this aspect of the research project was to compare the proficiency of two groups of language learners: first year writing class students and third year writing class students. In order to do this, the transcribed digitized samples of narrative writing were read and past tense verbs (or their absence) manually tagged, the tag typed immediately after the verb (or where one should have been in the case of

omission).

6.1 Protocol for Annotation/Tagging Sample One

All lexical items considered to be functioning as verbs were tagged on the basis of seven criteria: Verb type, Past tense inflection, Subject-verb agreement, Stylistic consistency, Existence of verb in English, Omission and Oversuppliance.

Tags therefore comprised two elements: an indication of the verb type (R, I, C, M or A) and an indication of whether the verb was correct (Y), incorrect (N) omitted (X) or oversupplied (O).

Table 1

Sample One Tag set

Category	Correct use	Incorrect use	Oversuppliance	Omission
Regular past	<RY>	<RN>	<RO>	<RX>
Irregular past	<IY>	<IN>	<IO>	<IX>
Copular past	<CY>	<CN>	<CO>	<CX>
Modal past	<MY>	<MN>	<MO>	<MX>
Auxiliary past (be, do, have)	<AY>	<AN>	<AO>	<AX>

Verb type: Five types of verbs were tagged: regular (R), irregular (I), copula (C), modal (M) and auxiliary (A), which included the verbs *be*, *do* and *have*.

Past tense inflection: Correctly inflected verbs were tagged Y; incorrectly inflected verbs, including misspellings, were tagged N.

Example: After they finish <RN> shopping and when they say <IN> good-bye, one person went <IY> back home happily, and the other was <CY> sadly.

Subject-verb agreement: Verbs were also tagged on the basis of whether they agreed with a given grammatically acceptable subject (Y) or not (N), regardless of whether the subject supplied by the writer was semantically appropriate or not.

Example: And Mike tried <RY> to wear the favourite stone washed jeans… it was <CY> last the stone washed jeans in there shop.

In the above example, the occasion for an error has been determined by an incorrect choice of lexical referent by the writer (the pronoun "it" instead of "they"), but as the verb agrees with the subject it has been tagged as correct.

Example: I was <CY> really tired last weekend with Mary. There were <CN> very crowded and I lost <IY> lots of money!

In the above case, the complement of the clause ("very crowded") indicates that the subject should be "it", and the "be" verb should function as a copula rather than auxiliary. The given subject "there" (confusing the pronoun and adverb forms) creates the grammatical problem, and as "were" agrees with neither subject nor complement but is required to function as a copula it is tagged CN.

In the case of a passive construction, if the auxiliary was correctly inflected to agree with the given (although erroneous) subject (as in the example below) the auxiliary was tagged as correct. This ensured that the research focused on verb inflection and suppliance rather than addressing clause construction.

Example: Tomoko was <AY> stolen her wallet. [Tomoko's wallet was stolen.]

Stylistic consistency: Writers may consciously choose to produce a

narrative about the past in the present tense in order to give it more immediacy. Determining whether writers of Sample One made this choice required evidence of the consistent use of the present tense throughout the text, and the appropriate use of past tenses to relate events that had occurred previously.

Example: First, they go to some shops to buy a bag. There are many types of bags. She can find what she want, but her friend says it doesn't suit you. Although her friend says that, she decides to buy it. It costs her 10,000 yen... On the way home, people around them sometime laugh at one people who bought <IY> purse because it doesn't match her.

Where use of present and past tenses were inconsistent with a stylistic choice, the inappropriate use of the present tense was tagged as incorrect. To some extent, evaluations depended on the subjective judgment of the researcher and relied on the answer to the question, "Given the context, should the past tense have been used here?"

Existence of verb in English: A basic criterion for tagging a verb as correct or not was whether the verb actually exists as a word in English. Neologisms were tagged as incorrect. In the example below "responsed" is approximate to "responded", a regular verb, so tagged R for regular and N for the use of a non-existent word.

Example: Hanako responsed <RN> "You had better make check list and you'll be able to bring what you want to bring".

With a verb that does exist in English, it was decided that if it was accurately supplied and correctly inflected, it would be tagged as correct regardless of whether it was lexically appropriate or not. This decision was perhaps not entirely satisfactory, and was, to some extent, a reflection on the limitations of the tag set, but it enabled the researchers to avoid making subjective judgments on the suitability of lexical choice. The range of inappropriate verbs supplied spanned from the reasonably acceptable to the unacceptable but understandable. For example, "to try to wear", a construction unintentionally suggesting an attempt to put on clothes in a shop, was used for "to try on".　No tags were available to differentiate between an incorrect "tried to wear" and a correct "tried on" as the two are distinguished by the use of an infinitive and a preposition, and "tried" in both cases is correctly inflected. At the other end of the spectrum, "to back home" was used for "to return home" or "to go (back) home", and while "to back" exists as a verb (as in "to reverse") its use here is entirely erroneous.

Example: They came <IY> a shop and watched <RY> clothes. [They came to a shop and looked at clothes.]

Similarly, the choice of copula, especially in the cases of "smell" and "taste", posed problems for writers. In the example below, the verb is tagged as incorrect due to it being present tense, not because it is in inappropriate.

Example: Leon's curry is <CN> good taste. [Leon's curry tasted good.]

Omission: A tag (X) was inserted where the writer had failed to supply a verb where obligatory.

Example: Both of two have <IN> an idea about getting some clothes, so they <IX> into the favorite clothing shop. [Both of them had an idea about getting clothes, so they went into the favourite clothing shop.]

Oversuppliance: A tag (O) was inserted after a verb where its use or inflection was unnecessary.

> *Example:* During the way back to home it was <AO> happened. [On the way back home it happened]

7. Annotation/Tagging of sample Two (Argumentative)

The prompt for Sample Two aimed to elicit an argumentative text in which students would present opinions with supporting evidence. The goal of this aspect of the research was to examine and quantify the type and frequency of devices used by students to create rhetorical cohesion in their writing, and to then compare the usage of such devices by first year and third year students. Using Hinkel's classification (2002), three text-rhetorical features (broadly termed sentence connectors here) were chosen for scrutiny and instances in the digitized texts manually tagged:

(1) Sentence-level coordinating conjunctions (or more commonly, transitions)　(e.g. *firstly, therefore, in fact*)

(2) Logical/semantic conjunctions and prepositions (e.g. *because of, despite, instead of*) often comprising a conjunction a preposition and a noun-phrase

(3) Exemplification markers (e.g. *for example*).

7.1 Protocol for Annotation/Tagging Sample Two

Each lexical item or string considered to be functioning as one of the three categories of rhetorical cohesion was tagged. Each word of the device was tagged according to category, with an additional tag at the end to indicate a complete unit.

Table 2
Sample Two Tag Set

Sentence connectors	Token/Word tag	Unit tag
Sentence-level conjunction	<ST>	<STU>
Exemplification	<ET>	<ETU>
Logical/semantic	<LT>	<LTU>

> *Example:* <s>On<ST>the <ST>other<ST> hand<ST><STU>, go abroad and study English is not the bad thing.</s>
>
> *Example:* <s>For <ET> example<ET><ETU>, you many not able to admit the other culture.</s>
>
> *Example:* <s>In <LT> my <LT> case<LT><LTU>, I didn't have enough knowledge of Japanese culture.</s>

Transitions and fragments: Determining what lexis could be considered as functioning as a sentence connector required the researchers to establish criteria for what could tolerably be considered a sentence.

> *Example:* <s>For <ET>example<ET><ETU>, school money and school money.</s>

With the absence of the necessary components of a clause, the above is not a full, correct sentence. However, it was decided that such text would be regarded as a sentence (and the conjunction, exemplification or logical/semantic transition tagged as such) on the basis that the student was intending the string of words to function as a separate sentence, however ill-formed.

Multiple transitions: While the use of a transition functions as a link between two pieces of text, the term might suggest only one link would be necessary. It was recognized, however, that more than one link is possible, for example, when the writer establishes two rhetorical connections. In the

instance below, the "So" is resultative, and the "first" enumerative. In this case, both units are tagged.

Example: <s>So<ST><STU>, first<ST><STU>we study Japanese. </s>

Noun phrases: Logical/semantic conjunctions consist of a conjunction and/or preposition plus a noun phrase that function adverbially.

Example: <s>Next <LT> day<LT><LTU>, she went to school.</s>

Determining whether to tag independent noun phrases that appeared to function as either logical/semantic or sentence-level conjunction was problematic. In Example 1 below, the noun phrase "Two point" clearly functions as the subject of the sentence, and so was not tagged. In Example 2, it could be argued that "Two point" is intended to have a sequencing function similar to that of "Second" or "Secondly". But since it could be equally argued that, by using a noun phrase, the writer is intending to establish an initial subject (e.g. A second point is (that) we make many new friends), it was decided that such phrases would not be tagged.

Example 1: <s>Two point is dangerous.</s></body>
Example 2: <s>Two point, we make new many friends.</s>

Adverbs: Adverbs and adverbial phrases that functioned as transitions, regardless of grammatical accuracy, were tagged. In the example below, "Exactly" has the same summative function as "To be exact" and was tagged accordingly.

Example: <s>I think merit is learn good pronunciation and English. </s><s>Exactly<ST><STU>,Japanese people pronunciation is not so good.</s>

8. Corpus Management and Statistical Analysis

All transcribed, digital files were preserved in folders (organized by sample, class and year) as the basic corpus for research. The raw and annotated data were statistically analyzed using WordSmith Tools 4.0 (Scott, 2004), RANGE (Nation, 2002) and SPSS 17.

Results

Fluency

Table 3

Token Counts in First and Third Year Students' Writing

Type of Writing	Narrative		Difference in Frequency	Proportional Increase	Argumentative		Difference in Frequency	Proportional Increase
Student Year	1st	3rd			1st	3rd		
GSL 1st	7,992	12,001	4,009	50.17%	9,437	13,319	3,882	41.14%
GSL 2nd	873	1,246	373	42.75%	488	808	320	65.65%
AWL	28	80	52	190.08%	184	349	165	89.72%
Others	887	1,211	324	36.58%	364	474	110	30.07%
Total	9,779	14,538	4,759	48.67%	10,473	14,950	4,477	42.75%

Note 1: Token counts of the first year students (190 essays for narrative and 188 essays for argumentative) were equalized according to the number of essays written by the third year students (131 essays for both narrative and argumentative).

Note 2: The total increase was 9,236 tokens (29,488-20,252=9,236) which is a 45.61 percent increase from the baseline of the first year.

Fluency of writing is operationally defined in this study as the number of tokens (words) that students could produce in the time allotted for the writing task. Twenty minutes were allotted for each task. Table 3 shows the total token counts for first and third year essays after equalization since the number of first year samples were greater than those from the third year (see note 1). Overall, third year students produced 45.61% more tokens in both writing tasks combined, which comprise 48.67% more tokens in the narrative essay and 42.75% more tokens in the argumentative essay. These gains are further broken down into gains in three lexical categories (the General Service Word list first and second thousand most frequent words and the Academic Word List) as measured

by Nation's Range software. The gains are discussed in more detail in the following section on lexical profile.

Lexical profile

The lexical profiles of first and third year writers are represented in Table 4 and 5 as type-token ratios (TTR) in the narrative writing task and the argumentative writing task. For the narrative writing task, the type token ratio of the first year writing was compared with those of their third year counterparts. The type-token ratios calculated by Nation's RANGE (2002) software are shown for tokens belonging to the General Service Word List 1, 2, the Academic Word List and other (words not belonging to any of the 3 groups) in table 4. A chi-square test of independence was performed to compare these ratios. For General Service Wordlist 1, first year students produced 11,591 tokens with 686 types compared with third year students writing, which contained 12,001 tokens and 774 types, yielding type token ratios of 0.059 and 0.064 for first and third year writers respectively. The chi-square test of independence showed no significance in the difference between the first and third year $[X^2(1)=2.531$, $p=.112$, $n.s.]$. Similarly, chi-square tests of independence showed no significant difference between the type-token ratios of first and third year students writing in Academic Word List items $[X^2(1)=0.109$, $p=.742$, $n.s.]$. Significant differences did emerge in the comparison of type-token ratios in the General Service Wordlist level 2 $[X^2(1)=5.829$, $p<.05]$ and also for the combined type-token ratio of all the categories $[X^2(1)=6.626$, $p<.01]$ suggesting an overall improvement in lexical variety from first to third year in the narrative writing task.

Table 4

Type-and-Token Ratio of First and Third Year Students and Differences in Their Ratios in Narrative Writing

Type of Writing	First Year Students			Third Year Students			Chi-square test of Independence
Student Year	Token	Type	TT-Ratio	Token	Type	TT-Ratio	
GS 1st	11,591	686	0.059	12,001	774	0.064	$X^2(1)=2.531, p=.112, n.s.$
GS 2nd	1,266	163	0.129	1,246	210	0.169	$X^2(1)=5.829, p<.05$
Academic	40	16	0.400	80	36	0.450	$X^2(1)=0.109, p=.742, n.s.$
Others	1,286	290	0.226	1,211	299	0.247	$X^2(1)=0.979, p=.322, n.s.$
Total	14,183	1,155	0.081	14,538	1,319	0.091	$X^2(1)=6.626, p<.01$

Note 1: Actual token frequencies of the first year students were used for this table.
Note 2: The Chi-square test of independence was performed on token and type frequencies between students .of the first and the third year students.

For the argumentative writing task, the type-token ratio of the first year writing was compared with those of their third year counterparts. As for the narrative writing task, the type-token ratios calculated by Nation's Range software are shown for tokens belonging to the General Service Word List 1, 2, The Academic Word List and other (words not belonging to any of the 3 groups) in table 5. A chi-square test of independence was performed to compare these ratios. For General Service Wordlist 1, first year students produced 13,543 tokens with 674 types compared with third year students writing, which contained 13,319 tokens and 768 types, yielding type token ratios of 0.050 and 0.058 for first and third year writers respectively. The chi-square test of independence showed significance in the difference between the first and third year $[X^2(1)=7.399$, $p<.01]$. In contrast, chi-square tests of independence showed no significant difference between the type-token ratios of first and third year students writing in General Service Wordlist level 2 $[X^2(1)=0.719$, $p=.396$, $n.s..]$. Significant differences did emerge in the comparison of type-token ratios in the Academic Word list items $[X^2(1)=5.831$, $p<.05]$ and also for the combined type-token ratio of all the categories $[X^2(1)=16.797$, $p<.001]$ suggesting an overall improvement in lexical variety from first to third year in the argumentative writing task.

Table 5

Type-and-Token Ratios of First and Third Year Students and Differences in Their Ratios in Argumentative Writings

Type of Writing	First Year Students			Third Year Students			Chi-square test of Independence
Student Year	Token	Type	TT-Ratio	Token	Type	TT-Ratio	
GS 1st	13,543	674	0.050	13,319	768	0.058	X2(1)=7.399, p<.01
GS 2nd	700	129	0.184	808	166	0.205	X2(1)=0.719, p=.396, n.s.
Academic	264	36	0.136	349	80	0.229	X2(1)=5.831, p<.05
Others	523	120	0.229	474	134	0.283	X2(1)=2.202, p=.138, n.s.
Total	15,030	959	0.064	14,950	1,148	0.077	X2(1)=16.797, p<.001

Note 1: Actual token frequencies of the first year students were used for this table.

Note 2: The Chi-square test of independence was performed on token and type frequencies between students of the first and the third year students.

Grammatical Accuracy

For the narrative writing task, an obligatory occasion analysis was used to determine the accuracy level at which students in the first and third year produced past tense verbs. The cases of correct use are presented in contrast to cases of incorrect use, which included errors of incorrectness, oversuppliance and omission. Using Pica's (1984) formula, accuracy can be expressed as a percentage. The actual number of correct/incorrect uses and accuracy level for the past tense categories of regular past, irregular past, copula past and auxiliary past for first and third year students are presented in table 6. A chi-square test of independence was used to compare the ratio of correct/incorrect uses of these tenses to determine whether they differed significantly between first and third year students' writing. For regular past, first year essays contained 326 cases of correct usage to 110 cases of incorrect usage yielding a ratio of 326:110 in comparison to third year students' writing which yielded a ratio of 418:93. The chi-square test of independence showed that these ratios differed significantly [$X^2(1)=6.903$, $p<.01$] with a higher accuracy rate for third years (81.80%) compared to the first year (74.77%) . For use of the

irregular past, the ratios of correct/incorrect usage in first year and third year writing were compared (736:141 and 665:80 respectively) using the chi-square test of independence, and again showed a significant difference [$X^2(1)=9.758$, $p<.01$] with an increase in accuracy between first and third year writing (83.92% and 89.26% respectively). For use of the copula past, the ratios of correct/incorrect usage in first year and third year writing were compared (297:151 and 308:105 respectively) using the chi-square test of independence, and also showed a significant difference [$X^2(1)=7.055$, $p<.01$] with an increase in accuracy between first and third year writing (66.29% and 74.58% respectively).   Finally, for use of the auxiliary past, the ratios of correct/incorrect usage in first year and third year writing were compared (131:140 and 122:58 respectively) using the chi-square test of independence, and showed a significant difference [$X^2(1)=16.594$, $p<.001$] with an increase in accuracy between first and third year writing (48.34% and 67.78% respectively).

Table 6

Past Tense Use by First and Third Year Students and Differences in Ratios of Correct/Incorrect Usage

Past Tense Type	Regular Past		Irregular Past		Copula Past		Auxiliary Past	
Student Year	1st	3rd	1st	3rd	1st	3rd	1st	3rd
Correct	326	418	736	665	297	308	131	122
Incorrect	101	89	111	75	124	88	50	23
Oversuppliance	5	4	25	4	4	3	63	17
Omission	4	0	5	1	23	14	27	18
Total	436	511	877	745	448	413	271	180
% Correct	74.77%	81.80%	83.92%	89.26%	66.29%	74.58%	48.34%	67.78%
Chi-square test of Independence	$X^2(1)=6.903$, $p<.01$		$X^2(1)=9.758$, $p<.01$		$X^2(1)=7.055$, $p<.01$		$X^2(1)=16.594$, $p<.001$	

Note: The Chi-square test of independence was performed on the ratio of correct and incorrect frequencies between students of the first and third year.

Rhetorical/cohesive devices

In the total token counts of argumentative essay texts, the actual frequencies of tokens contributing to sentence connector units of three types (sentence-level conjunction, exemplification and logical/semantic) are shown in Table 7. The proportion of tokens contributing to each of these types of conjunction units is expressed as a percentage of total tokens. The Chi-square test of independence was performed on the ratio of frequencies of tokens contributing to each conjunction type to the remaining token count (non-appearance) between 1st year and 3rd year student essays.   In the case of the sentence-level conjunctions, the ratio of 513 token frequency of appearance to 14,927 token frequency of non-appearance (calculation of 15,440 – 513) in first year students' samples compared to the ratio of 550 token frequency of the same connector type to 14,422 token frequency (14,972 – 550) were examined to determine whether there was a significant difference after two years of writing instruction. The result showed no significance [$\chi^2(1)$=2.776, p=.096, $n.s.$], indicating no change in the number of tokens contributing to sentence-level conjunctions. The same analysis of the Chi-square test of independence was performed for exemplification (116 to 15,324 for 1st year, and 109 to 14,863 for 3rd year). Again, the result showed no significance [$\chi^2(1)$=0.056, p=.813, $n.s.$], indicating no change in the number of tokens contributing to exemplification conjunctions. The same Chi-square test of independence was performed for logical/semantic conjunctions   (25 to 15,415 for 1st year, and 78 to 14,894 for 3rd year). The result showed significance [$\chi^2(1)$=29.033, p<.001], indicating that the third years students (78 tokens or 0.52%) used logical/semantic conjunctions more frequently than the first year students (25 tokens or 0.16%).

Table 7

Sentence-level Connectors Used by 1st and 3rd Year Students and Differences in Proportions of Lexical Counts

Sentence Connector Type	Sentence-level conjunction		Exemplification		Logical/Semantic Conjunction	
Student Year	1st[t]	3rd	1st	3rd	1st	3rd
Token Frequency	513	550	116	109	25	78
Total tokens	15,440	14,972	15,440	14,972	15,440	14,972
Proportion (%)	3.32%	3.67%	0.75%	0.73%	0.16%	0.52%
Chi-square test of Independence	$\chi^2(1)$=2.776,p=.096, $n.s.$		$\chi^2(1)$=0.056, p=.813, $n.s.$		$\chi^2(1)$=29.033, p<.001	

Note 1: Total token frequencies were 15,440 times for 1st year and 14,972 times for 3rd year.

Note 2: Proportions in the table were calculated by dividing the token frequency of each conjunction type by the total token frequency.

Note 3: The Chi-square test of independence was performed by a token frequency of each conjunction type and non-appearance in the total token frequency between 1st year and 3rd year samples.

The total numbers of first and third year argumentative writing samples in which sentence connector units (sentence-level conjunction, exemplification and logical/semantic) appeared at least once or did not appear at all are shown in Table 8. The proportion of the essays in which these conjunctions appeared is expressed as a percentage. A Chi-square test of independence was performed comparing the ratio of appearance to non-appearance between the first and third year to determine whether the difference was significant. Of the 188 first year essays and 131 third year essays, the proportion appears to have increased in each case. Inter-sentence-level conjunctions appeared at least once in 152 of the 188 first year student essays yielding an appearance/non-appearance ratio of 152:36 (a proportion of 80.85%).

The third year student writing samples showed an appearance/non-

appearance ratio of 124:7 (a proportion of 94.66%). The chi-square analysis of these ratios showed a the difference was significant [$\chi^2(1)$=12.617, p<.001]. The same analysis was conducted for ratios of appearance/non-appearance of exemplification conjunctions (First year 52:136 and third year 46:85). Although there was an increase in appearance, the difference was not significant [$\chi^2(1)$=2.016, p=.156, *n.s.*]. For logical/semantic conjunctions, the same analysis was conducted for ratios of appearance/non-appearance between first and third year writing samples (first years 8:180 and third year 21:110). The chi square test of independence showed a significant difference [$\chi^2(1)$=12.953, p<.001] suggest

ing that after two years of writing instruction, students employ both inter-sentence level and logical/semantic conjunction types more frequently in argumentative essay writing.

Table 8
Sentence Connector Units Used by 1st and 3rd Year Students and Differences in Text Coverage

Conjunction Unit Type	Sentence-level conjunction		Exemplification		Logical/Semantic	
Student Year	1st	3rd	1st	3rd	1st	3rd
Appearance	152	124	52	46	8	21
Non-appearance	36	7	136	85	180	110
Proportion (%)	80.85%	94.66%	27.66%	35.11%	4.26%	16.03%
Chi-square test of Independence	$\chi^2(1)$=12.617, p<.001		$\chi^2(1)$=2.016, p=.156, *n.s.*		$\chi^2(1)$=12.953, p<.001	

Note 1: Total numbers of argumentative writing samples were 188 for 1st year and 131 for 3rd year.

Note 2: Proportions in the table were calculated by dividing the number of texts used each conjunction type by the total text frequency.

Note 3: The Chi-square test of independence was performed on the number of texts in which each connector type appears and does not appear in the total texts between samples of 1st year and 3rd year.

Discussion

The present study has attempted, perhaps too ambitiously, to derive a

profile of writers before and after 2 years of a three-year writing program in a number of areas including fluency, lexical richness, grammatical accuracy and textual-cohesion. The preceding results will be discussed with attention to the limitations and weaknesses of the present study.

With respect to writing fluency, a rudimentary count of tokens was the basis of comparison and showed that third year students were able to produce more (upwards of 40%) in both types of writing tasks. Fluency in writing is arguably less important than accuracy, but for those whose writing may be assessed by standardized tests with a writing component, such as IELTS or TOEFL, being able to organize one's thoughts and to articulate and develop them in clear written form within time limits is important. These tests may not be seen as authentic writing tasks but are the means by which English skills are judged often for selection for employment and promotion in professions requiring English language skills in Japan. The increase in number of words produced is probably not the result of in-class timed writing tasks, but general familiarity with, and confidence in the writing process and having skills available to produce text.

Making comments on lexical richness and variation is a more problematic issue. In the present study, type-token ratios were employed to measure change in lexical richness. In and of itself, type-token ratios as a measure of lexical richness has garnered criticism for being unreliable as a measure of lexical richness in comparison with standardized tests (e.g., Vermeer, 2000; Daller et al, 2003) and unsatisfactory for yielding accurate and informative measures in comparison with other measures such as the Guiraud index, Advanced TTR and the Guiraud Advanced, the last two of which examine lower frequency and academic vocabulary, and are therefore more suitable for assessing advanced learners. In particular, text length has a considerable effect on the measure of lexical richness in the TTR. This draws into question the gains observed at the GSL 1 and AWL categories of vocabulary since the third year writers produced over 45% more tokens than their first year counterparts. As Daller et al (2003) and

Daller and Phelan (2007) point out, longer texts yield lower TTRs; therefore, given the third year students produced more tokens, the observed gains may be considered a conservative estimate. In any event, further measures with more sophisticated tools may either support the observed gains or diminish them so that they may not be significant at all. Classes in the English writing program do not have an explicit vocabulary component and whatever gains there are could be attributed to incidental learning or learning within other parts of the English program as a whole. The preliminary findings using the TTR in different lexical frequency groups in the present study do seem encouraging, but warrant more rigorous investigation.

Gains in grammatical accuracy are observed in all four past tense categories under investigation (regular past, irregular past, copula past and auxiliary past) and appeared to differ significantly. While this does demonstrate improvement in grammatical accuracy, it should be emphasized that the analysis was limited to past tense and during annotation, the researchers observed other grammatical categories in which improvements may be less marked or altogether absent. Of specific interest in future investigations, may be the use of passive constructions. In using the obligatory occasion analysis, it should be noted that in none of the categories did the group of first or third year writers (as a whole) exceed 90%, the higher threshold criteria for determining acquisition of these grammatical morphemes adopted in previous studies employing the technique. A survey of writing teachers instructional objectives revealed that a focus on grammatical accuracy was certainly not shared by all and so gains should not be attributed exclusively to the writing program itself but more realistically to the overall program and the exposure to English it entails. Future investigations of this corpus should examine more diverse grammatical categories and also include an examination of grammatical/lexical related issues.

While the analysis of token frequency ratios only appeared to show significant differences between first and third year students for logical

semantic conjunctions, the comparison of distribution of appearance of intra-sentence level (transitions) and logical semantic conjunctions showed greater gains between first and third year. This contrast may be explained by the overreliance on more basic connectors (and, so, but) by first years which has been often observed in NNS academic writing, whereas third years perhaps use more sophisticated conjunctions but more sparingly. This hypothesis requires revisiting the data in order for it to be substantiated. Also, since it has been shown in Hinkel's (2002) research, Japanese and other L2 English writers typically overuse sentence-level conjunctions, logical/semantic conjunctions and exemplification than do NSs, so seeing no increase in the frequency is perhaps not troubling. What is apparent is that third years do demonstrate more application of conjunctions in their essays while a large proportion of first years failed to employ these conjunctions at all in their writing.

Having established the corpus and concluded the tentative investigation and attempts to derive profiles of first and third year students' writing, a number of short-comings have emerged which lay the foundation for, and establish direction for more thorough research. The corpus itself is valuable as a body of data collected under uniform conditions from which confident comparisons between groups may be drawn. The methodology for dealing with lexis, in particular, is not satisfactory, and the data should be revisited with more sophisticated tools. In the process of dealing with the raw data, more problematic areas of grammar and lexis have been observed and may drive future studies. Finally, at the time of writing, the corpus is being further developed with the inclusion of longitudinal data from the first year student group of the present study as they enter their second year of instruction. The issues and directions for research raised by the present study have provided invaluable insights and experience from which to move forward.

References

Abe, M. and Tono, Y. (2005). Variations in L2 spoken and written English: Investigating patterns of grammatical errors across proficiency levels. *Proceedings from the Corpus Linguistics Conference Series*, 1(1)

Andersen, R.W. (1978). An implicational model for second language research. *Language Learning, 28*, 221-282.

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge MA: Harvard University Press

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

Daller, H. and Phelan, D. (2007) What is in Teacher's Mind? In H. Daller, J. Milton & J. Treffers-Daller, (Eds.). *Modelling and Assessing Vocabulary Knowledge* (pp. 234-244). Cambridge, UK: Cambridge University Press

Daller, van Hout, R., and Treffers-Daller, J. (2003). Lexical Richness in the Spontaneous Speech of Bilinguals. *Applied Linguistics* 24, 197-222.

Granger S. (2004) Computer learner corpus research: current status and future prospects. In Connor U. and Upton T. (eds) Applied Corpus Linguistics: A Multidimensional Perspective. Amsterdam & Atlanta: Rodopi. 123-145.

Granger, S., Hung, J. and Petch-Tyson, S. (Eds.)(2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. John Benjamins: Amsterdam

Hinkel, E. (2002). *Second Language Writers' Text: Linguistic and Rhetorical Features*. New Jersey: Lawrence Erlbaum Associates.

Hyland, K. and Tse, P. (2007). Is There an Academic Vocabulary? *TESOL Quarterly*, 41(2), 235-253

Laufer, B. (1994).The Lexical Profile of Second Language Writing; Does it change over time? *RELC Journal, 25*, 21-33.

Millar, N., & Lehtinen, B. (2008). DIY local learner corpora: Bridging gaps between theory and practice. *JALT CALL Journal*. 4(2): 61-72

Muehleisen, V. (2006). Introducing the SILS Learners' Corpus: A Tool for Writing Curriculum Development. *Waseda Global Forums, 3,* 119-125.

Pica, T. (1984). Methods of morpheme quantification: their effects on the interpretation of second language data. *Studies in Second Language Acquisition 6, 69*-78.

McEnery, T., Xiao, R. and Tono, Y. (2006). *Corpus Based Language Studies: an advanced resource book*. Routledge: London

Nation, P. and Kyongho, H. (1995). Where would General Service Vocabulary Stop and Special Purposes Vocabulary Begin? *System*, 23(1), 35-41

Nation, I.S.P. (2002). RANGE [computer software]. Available from http://www.vuw.ac.nz/lals/publications/software.aspx

Scott, M., (2004). WordSmith Tools version 4, Oxford: Oxford University Press.

Tono, Y. (2000). A computer learner corpus based analysis of the acquisition order of English grammatical morphemes. In L.Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective (*pp.123-132). Frankfurt-am-Main: Peter Lang.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17, 65 - 83.

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