

A corpus-based analysis of Japanese university-level learners' L2 writing development over a one-year period

Nicolai Struc and Nicholas Wood

Abstract:

This paper describes the development and analysis of a longitudinal learner corpus comprised of Japanese university students' English writing over the period of one year. Students completed two writing tasks, a narrative and an argumentative essay, in response to the same prompts at two points in time one year apart. The resulting subcorpora are analyzed and compared with respect to fluency, lexical diversity, grammatical accuracy and use of rhetorical/cohesive devices. Gains were observed in these areas, most notably in fluency and lexical diversity. Methodological issues in analyzing grammatical accuracy and use of rhetorical/cohesive devices render interpretation of these results less conclusive. Patterns of observed developments in these four areas are discussed, followed by an acknowledgement of the limitations of the study and considerations for directions in further research.

Acknowledgement: This research was made possible through a research grant provided by the Reitaku Linguistic Research Center (Reitaku University)

1. Overview and Purpose

The study described in this paper focuses on learner writing. The analysis of second language learner writing (and speaking) has been helped in recent years by the application of tools and methodology from the field of corpus linguistics and methodology associated with electronic text analysis, for example, the use of concordances and word frequency lists (see Adolphs, 2006). The technological developments, which have grown from this field of inquiry, are, likewise, effectively applied in interrogating learner corpora. In practice, this means we can investigate specific phenomena in large amounts of text with programs, such as Wordsmith Tools (Scott, 2004).

This study carries forward the initial work done in constructing and analyzing a cross-sectional corpus (Struc & Wood, 2009). In the original corpus, writing samples were collected from students in the first, second and third years of a university English writing program in Japan. The areas investigated included lexical diversity, fluency, grammatical accuracy and use of rhetorical/cohesive devices. Comparisons were only made between first and third year students because of limited time and resources and with the expectation that greater gains would be observed. First and second year writing samples were not compared. Instead, a new opportunity presented itself, namely, the opportunity to collect data from the first year group as they entered the second year of the writing program. This provides the valuable opportunity to examine data from the same

students and thereby make stronger claims about their developmental changes.

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 one year of writing instruction change? If so, by how much?
2. Lexical diversity: Does the lexical diversity brought to bear on the writing tasks before and after one year of writing instruction change?
3. Grammatical accuracy: Does the grammatical accuracy in past tense/aspect in writing produced before and after one year of writing instruction change?
4. Textual cohesion: Does the frequency, distribution and characteristics in the use of sentence connectors in writing produced before and after one year of writing instruction change?

2. Background

The use of learner corpora in investigating patterns of language acquisition is by now well established with numerous studies investigating many aspects of learner language development (e.g., Granger, Hung, & Petch-Tyson, 2002; Granger, 2004; McEnery et al., 2006). However, a corpus in itself is no more useful than stacks of learner writing and while software tools can help to efficiently access what lies within the texts, consideration of the theoretical approaches to the respective areas under investigation are of greater relevance to an informative study. Rather than a review of learner corpus related literature,

it is more appropriate to provide a background to the specific areas dealt with in this study. With a corpus of learner writing at our disposal, we hoped to capture profiles of the writing produced at two points. The profiles include aspects of writing in which development is associated with quality of writing and overall language proficiency. These areas include: fluency, lexical diversity, grammatical accuracy and use of rhetorical cohesive devices. Each one will be briefly explained in terms of relevant studies and developments in the respective areas, as well as a rationale for the selection of specific procedures and measures.

2.1 Fluency

Fluency, as defined in the present study, is a straightforward production metric, represented by the amount of writing that a student can produce in a fixed period of time, as measured by the number of tokens. Wolfe-Quintero et al. (1998) describe fluency as meaning “that more words and structures are accessed in a limited time”. Certainly other measures, such as mean sentence length, mean length of t-unit (Larsen-Freeman, 2006), mean length of clauses or mean length of error-free clauses (see Wolfe-Quintero, 1998), are viable options. However, these may not always be suitable for application to low-proficiency level writing. T-unit and error-based analyses are problematic for analyzing texts produced by low-proficiency learners in which error-free clauses are infrequent, t-units frequently cross sentence boundaries and sentence fragments may show no clear connection with surrounding sentences. Fluency is certainly related to cognitive processes in writing, in which learners make decisions about approaches to the task and consider grammatical constructions and spelling. These are without

doubt all related to fluency, as the decisions and hesitations made by a writer will affect the rate of production in a limited time. Wolfe-Quintero et al (1988) cite 18 studies which have used number of words as developmental measures, 10 of which show high correlations with proficiency or show an overall effect for proficiency. For these reasons, we adopt this somewhat unsophisticated metric as an indication of fluency in the present study. On a final note, measures involving t-units and clause length may overlap with, and perhaps better serve analysis of syntactic complexity or accuracy and will be employed in this type of analysis in forthcoming studies.

2.2 Lexical diversity

Lexical diversity (LD) is a construct, which is contentious for a variety of reasons. It has been noted, for example, that many terms including lexical diversity, lexical richness, lexical variation, etc., have been used in describing this construct. At its most basic, lexical diversity is an expression referring to the variety of lexical items exhibited in a given text. McCarthy and Jarvis (2007) succinctly describe it as “the range and variety of vocabulary deployed in a text by either a speaker or a writer.” Historically, LD has been measured using the type-token ratio (TTR) which, simply put, is a ratio of the total number of different words in a given text divided by the total number of words. What had become apparent very early is that this method of calculating lexical diversity is extremely sensitive to text length. Texts will necessarily exhibit a declining TTR as the length increases. Even if new words are consistently introduced, it is impossible to avoid repetition of many function words such as articles, pronouns, conjunctions as well as lexical items related to themes under discussion.

Having recognized this problem, a number of competing solutions to the problem have been proposed. One, the Guiraud index G or Root TTR (T/\sqrt{N}), is also based on the TTR but it seeks to compensate for the effect of text length by using the square root of the tokens. This formula helps to diminish the effect of text length but not totally eliminate it (Vermeer, 2000). Still, it is a more effective measure than the TTR for comparing texts of different length and is considered as very stable for use with learner data (Van Hout & Vermeer, 1988). Another advantage is its simplicity of calculation; if the type token ratio is known, it is simply one more step to arrive at a more equalized representation. Indeed, many researchers, even in light of new more sophisticated and complex measures, have elected to use the Guiraud index or one of its variations (Daller et al., 2003). It has been criticized among other TTR-based measures for simply being a rescaled type-token ratio measure (Malvern & Richards, 2002). That is, the differences observed in the TTR are preserved but the range may become narrower.

Malvern & Richards (1997) proposed another competing measure called a D value. This is a highly complex measure seeking to overcome the text length effect by employing segmental type/token ratios derived from repeated random samples of between 34 to 50 words drawn from a text, and compares the resulting mean type-token ratios at 16 points to a theoretical curve of the expected decline. This measure has seen wide acceptance and has been adopted in many studies (e.g., Yu, 2007; Johansson, 2008; Duran et al., 2004) but has also been severely criticized for not being as resistant to text-length effects as the developers had postulated. It appears that the measure becomes unstable at longer text lengths (McCarthy and Jarvis, 2007). Despite these criticisms, the measure is still used extensively by researchers and is considered to be

stable and useful for dealing with shorter texts, especially within the 100-400 word range lengths (McCarthy and Jarvis, 2007). Yu (2007) has found measures of lexical diversity D positively correlated with overall ratings of speaking and writing and general language proficiency in the Michigan English Language Assessment Battery (MELAB).

For the purposes of the present study, both the Guiraud Root TTR and D have been adopted for different reasons and purposes. While the D may be seen as more effective in overcoming text-length effects, it has been cautiously employed here for the following reason. In order to carry out the VocD analysis using Meara and Miralpeix's D_Tools software, a minimum text size of 50 tokens is necessary. Since approximately 1/3 of all the texts in the corpus fail to reach this criterion, alternative solutions were considered. It was decided that combining the two texts (the narrative and argumentative writing samples) produced by the same writers at the same point of data collection could serve as a workaround. This gives the advantage of being able to include 98% of the writing samples. The disadvantage is that it would likely yield D values which are artificially inflated because of this combination. Indeed, Yu (2007) observed that different topics and topic types had effects on lexical diversity. The implication is that the results should be treated with caution and used more for internal comparison rather than attempting to draw comparisons with data obtained in other studies.

In light of these drawbacks, a second measure, the Guiraud G (root TTR) has been employed to measure overall lexical diversity in the respective tasks as well as within different lexical subgroups used by Laufer and Nation (1995) in their Lexical Frequency Profile. The Lexical Frequency Profile can be calculated by means of Nation's Range software providing type and token counts for words, which belong to the General

Service Lists (West, 1953) of the first and second thousand most frequent words, as well as words belonging to Coxhead's (2000) Academic Word List. While the content of the AWL has been the subject of discussion in regard to its constituent items, a coverage rate of 10.6% was reported when applied to a 3.3 million-word corpus of academic texts (Hyland & Tse, 2007). The GSL have proven to be fairly robust as well, with reports of consistently high coverage in non-fiction (75%) and fiction (90%) texts (Nation & Kyongho, 1995). The lexical frequency profile has been primarily adopted for use with the texts of more advanced learners and is often used to identify the lexical profile of items beyond the GSL word lists (e.g., Laufer, 1994; Daller et al., 2003). For the present study, using all three lexical subgroups with the addition of a fourth (see section 3.2) provides a way to give more detailed definition of lexical diversity by determining where the gains occur, even within the GSL.

2.3 Grammatical Accuracy

The construct of grammatical accuracy has been approached in corpus-based studies with measures such as counts of error-free t-units (see Larsen-Freeman, 2006) and examining patterns of errors related to nouns and verbs across proficiency levels (Abe & Tono, 2005). These measures offer certain advantages but may be inappropriate for the present study. First of all, a count of error-free t-units may not be a sensitive enough measure for this study since the general proficiency of learners' whose writing comprises this corpus is quite low and we would expect to find very few if any error-free t-units. An approach such as Abe and Tono's (2005) captures errors in a wide variety of grammatical areas and also distinguishes between types of errors but for the present study was ultimately deemed impractical given limited time and resources.

The obligatory occasion analysis, pioneered by Brown (1973) and later adopted in various studies (e.g., Andersen, 1978; Pica, 1984; Mochizuki et al., 2008, etc.), is an approach, which distinguishes cases of correct and incorrect usage and later came to include categories of omission and oversuppliance (Pica, 1984). It has the advantage of being able to produce a calculation reflecting accuracy as a percentage of correct usage as well as provide some information about the nature of errors. Typically, accuracy rates of over 80 or 90% (depending on the study) are considered to indicate that a given grammatical morpheme has been successfully acquired (Brown, 1973, Andersen, 1978). This approach was adopted in the previous study (Struc & Wood, 2009) as well as in the present study, so that comparisons may be drawn.

2.4 Use of rhetorical cohesive devices

As cited in Hinkel (2002), the deployment of conjunctions is reported to be correlated with assessments of discourse cohesion in learner writing (Davidson, 1991) and is the most lexically simple (and thus easily measurable) way of improving text cohesion (Halliday & Hasan, 1976). An exhaustive study by Hinkel (2002) of learners from various L1 backgrounds examined a range of rhetorical cohesive devices including hedges, conjunctions, demonstratives and emphatics. For the present study, adopting a wide focus of this kind was not practical and instead, a small subset of these areas was selected for analysis. Specifically, the use of sentence level conjunctions (i.e., sentence initial conjunctions which indicate a semantic connection with previous sentences while maintaining the new sentence as an independent clause), exemplification markers, and logical semantic conjunctions/prepositions are examined. As suggested above, these are easily identifiable lexical exponents of text cohesion, and

therefore, the overall distribution and frequency of deployment and characteristics of these devices may be indicative of development in rhetorical writing skills.

3. Method

3.1 Longitudinal Learner corpus data

The present study has sought to further build on the corpus initially constructed by Struc & Wood (2009), The methods of data collection have thus, remained constant. The data is comprised of two writing samples from each writer at two points in time: the beginning of their first year in the program and the beginning of their second year in the program. With the first year data having been collected, the first task was to collect samples from the students again as they began the second year. Students were given two writing tasks, one a narrative and an argumentative essay, each to be written for no more than 20 minutes, by hand and without use of reference materials, such as a dictionary. The writing prompts were presented in their native language (Japanese) but are presented here in an English translation:

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.”

Students were provided with an explanation of the general goals of the research and a request to participate. In addition, students were asked to complete a questionnaire and write their student ID numbers on each

writing sample as well as their questionnaire for matching with first year data. Each file was initially labeled according to their ID number but these were later recoded to ensure anonymity.

3.2 Data processing

After all samples were collected, files were transcribed to electronic txt files, proofread, and any incorrect transcriptions were edited. Next, all files were matched between the first and second year data and any files whose counterpart in either the first or second year were absent were discarded, thus creating the longitudinal corpus on which the analysis is based.

The resulting number of usable files for task 1 was 340 (170 X 2), and 344 (172 X 2) for task 2. The total number of writing samples is 688 texts, and the total number of words in the longitudinal corpus is 57,793. This means that there were 170 comparable files for the first task, and 172 comparable files for the second task.

For the lexical analysis, all files were submitted to an initial processing with Nation's Range software. Any items which were not found in the 3 baseword lists provided were assembled as a separate list, and from this, a fourth baseword list was compiled comprising any words which were not misspellings, neologisms, morphological misapplications or proper nouns (with the exception of country names or nationalities). Any word families were identified and organized into subgroups as lemmas.

The narrative writing samples were tagged for grammatical accuracy of past tenses. 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	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	<AY>	<AN>	<AO>	<AX>

For the argumentative writing samples, Hinkel's (2002) classification scheme for three categories of textual-cohesive devices discussed earlier was used. 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.

- (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*).

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>

Given the level of student writing, a number of decisions were made in the application of both sets of tags in ambiguous cases. For a detailed discussion of the protocols employed in the tagging process, the reader is directed to Struc and Wood (2009).

4 Results and Discussion

4.1 Fluency

Fluency is defined in this study simply as the number of tokens that students were able to produce in the 20 minutes allotted for each writing task. A paired-samples t-test was conducted to compare tokens produced by writers in both tasks in their first year and their second year. There was a significant difference in the mean token counts for the narrative writing task (task 1) between the first year ($M=74.11$, $SD=39.5$) and second year ($M=88.52$, $SD=38.57$) samples; $t(169)=5.85$, two-tail $p=.000$. A 95% C.I. for gains in the narrative writing task is (9.55, 19.27). There was also a significant difference in the mean token counts for the argumentative writing task (task 2) between the first year ($M=79.37$, $SD=43.53$) and second year ($M=95.24$, $SD=43.38$) samples; $t(171)=5.85$, two-tail $p=.000$. A 95% C.I. for gains in the narrative writing task is (15.88, 21.23). There appear to be significant gains in the number of tokens produced in both tasks between the first and second year measures.

4.2 Lexical diversity

Lexical diversity in the writing samples produced by the subjects were analyzed using two measures: the Guiraud index of root type-token ratios and the D.

4.2.1 D (VocD) measures of lexical diversity

Since the VocD calculation requires minimum samples of at least 50 tokens, the text from both tasks were combined and the total resulting D values were compared between first and second year samples. A total of seven student samples had to be discarded because the writers had failed to produce a total of 50 tokens in the combined writing tasks in the first or second year or because data from one of the tasks was missing leaving 165 samples to be compared. A paired-samples t-test was conducted to compare D values obtained in the combined texts from both tasks in the first year and second year. There was a significant difference in the mean D values between the first year ($M=65.50$, $SD=18.02$) and second year ($M=72.57$, $SD=20.24$) samples; $t(164)=4.329$, two-tail $p=.000$. A 95% C.I. for gains in the combined writing tasks is (3.85, 10.30). These results show an overall gain in lexical diversity between the first and second year writing samples.

4.2.2 Guiraud index G measures of lexical diversity

In addition to the VocD, Guiraud's Root type-token ratio was used to analyze the lexical diversity in the respective writing tasks between the two years' data. A paired-samples t-test was conducted to compare the mean Guiraud index scores of lexical diversity produced by writers in both tasks in their first year and their second year. There was a significant

difference in the mean Guiraud index scores for the narrative writing task (task 1) between the first year (M=5.03, SD= .952) and second year (M=5.46, SD= .862) samples; $t(169)=6.68$, two-tail $p=.000$. A 95% C.I. for gains in the narrative writing task is (.306, .562). There was also a significant difference in the mean Guiraud index scores for the argumentative writing task (task 1) between the first year (M=5.20, SD= .96) and second year (M=5.59, SD= .921) samples; $t(171)=5.762$, two-tail $p=.000$. A 95% C.I. for gains in the argumentative writing task is (.255, .521). Similar to the VocD values, these results show gains in lexical diversity in both tasks between the first and second year writing samples.

Table 3
Guiraud index G means for both tasks in lexical subgroups and paired t-test results

	Yr	TASK 1					TASK 2				
		M	N	SD	t	p	M	N	SD	t	p
Total	1	5.0303	170	.95178	6.677	.000	5.1983	172	.96151	5.762	.000
	2	5.4642	170	.86238			5.5866	172	.92055		
GSL 1	1	4.6653	170	.91857	4.942	.000	4.8632	172	.89024	5.041	.000
	2	4.9854	170	.81930			5.1821	172	.85288		
GSL 2	1	1.6318	162	.45693	4.362	.000	1.2768	143	.48302	3.262	.001
	2	1.8517	162	.52393			1.4436	143	.50814		
AWL	1	1.1183	7	.20212	2.035	.088	1.0186	90	.27161	2.863	.005
	2	.9077	7	.24433			1.1566	90	.39262		
BSW D4	1	1.1292	96	.41817	2.414	.018	1.3295	116	.43528	1.360	.176
	2	1.2827	96	.44035			1.4086	116	.45782		

Nation's (2002) Range software has permitted a description of the lexical diversity within lexical subgroups (General Service Lists 1, 2, The

Academic Word List and the Baseword 4 list comprising other acceptable lexical items) as expressed in the table below. A paired-samples t-test was conducted to compare the mean Guiraud index scores of lexical diversity produced by writers for these lexical subgroups in both tasks in their first year and their second year.

The table above (Table 3) shows significant gains in the Giraud index between years for the GSL first and second thousand most frequent words but only for the AWL in the argumentative writing task and also only for the BSWD 4 group in the narrative writing task.

4.3 Grammatical accuracy

Measures of grammatical accuracy in the past tense types under investigation are represented as percentages which are derived from a ratio of correct instances of usage to erroneous usage, and these are broken down into the categories of incorrect, oversuppliance and omission in the table below. Only the narrative writing task samples were examined for grammatical accuracy.

The obligatory occasion analysis was only applied in cases in which the writers chose to deploy a particular grammatical structure, meaning that data for each of the tenses was not available for each writer, which precluded a paired comparison. Instead what is presented is simply a descriptive account of the group as a whole in terms of accuracy in instances of usage or non-usage as the case may be. Still, patterns of gains are observed for the regular past (72.94% to 82.26%), the irregular past (83.75% to 85.68%), the copula past (66.33% to 77.24%) and the auxiliary past (48.19% to 65.63%). The accuracy in the modal past

however, shows a decline (83.11% to 81.71%). According to Brown's (greater than 90%) criteria for a grammatical morpheme having been successfully acquired, none of the reported values would qualify although, in general, gains in accuracy are observed.

Table 4

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		Modal Past	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Student Year										
Correct	283	357	665	790	260	319	120	126	123	143
Incorrect	97	75	100	106	106	82	47	34	23	32
Oversuppliance	5	2	24	27	4	1	59	14	1	0
Omission	3	0	5	6	22	11	23	8	1	0
Total	388	434	794	922	392	413	249	182	148	175
% Correct	72.94	82.26	83.75	85.68	66.33	77.24	48.19	65.63	83.11	81.71

4.4 Use of rhetorical/cohesive devices

The use of rhetorical/cohesive devices was measured by the instances of use of sentence-level conjunctions, exemplification and logical semantic conjunctions in the argumentative writing task only. The first analysis examines the distribution of these conjunctions among the texts to determine whether they are deployed by more writers after one year of writing instruction. The table below provides the actual frequencies of appearance or non-appearance in the individual texts and these values are

summarized as a proportion of the total texts in which they appear. The total numbers of units are reported as actual frequencies as well as the mean number of units in the texts in which they appear. Given the level of text coverage shown in the table, only a descriptive account can be provided for the group and no opportunities for paired comparison are immediately apparent.

Table 5

Sentence Connector Units Used in 1st and 2nd Year Writing and Differences in Text Coverage

Conjunction Unit Type	Sentence-level conjunction		Exemplification		Logical/Semantic	
	1st	2nd	1st	2nd	1st	2nd
Appearance	139	148	45	53	6	3
Non-appearance	33	24	127	121	166	169
Proportion (%)	80.81%	86.63%	26.16%	30.81%	3.49%	1.74%
TOTAL UNITS	423	491	52	63	6	3
Mean occurrences/text	3.04	3.33	1.16	1.18	1	1.74

A second step of analysis involves the breakdown of the sentence-level conjunctions and the sophistication of the lexical item chosen by the writer to achieve the desired rhetorical/cohesive function. Broadly speaking, the most basic inter-sentence level conjunctions available to writers are AND, SO and BUT. It is expected that these will decrease in the second year and be replaced by more sophisticated lexical items/units

and in some cases to achieve more specialized functions, such as enumeration. Use of a concordancer allowed the researchers to compile a list of final lexical items in all sentence transition units. The table on the following page (Table 6) shows a list in descending frequency of the final words of sentence-transition units in first and second year argumentative writing samples (e.g. (In) ADDITION = 2). Concordances give a clearer picture of the characteristics of the sentence transition units.

While the majority of sentence transitions in first year and second year texts are comprised of AND, SO and BUT (67% and 49% respectively) the remainder of the units show some difference in their characteristics. For first year writing, the remainder appears to be comprised of predominately enumerative type transitions, the second year writing samples are characterized by instances of more sophisticated transitions functioning to show contrast, cause-effect and result.

Overall, second year writers appear to have both a greater number of transitions at their disposal, and a greater range of particular types of transitions. First year writers, for example, make use of basic enumerative forms (e.g. FIRST and SECOND), but a year later, there is a noticeable employment of both basic and *-ly* forms (e.g. FIRST and FIRSTLY, SECOND and SECONDLY).

Table 6

Frequency of final words comprising sentence-transition units in First and second year argumentative writing samples

First year		Second Year	
word	frequency	word	frequency
BUT	118	SO	104
SO	100	BUT	80
AND	64	AND	59
FIRST	38	HOWEVER	52
SECOND	26	FIRST	49
HOWEVER	11	SECOND	29
THEN	10	HAND	18
ALSO	9	ALSO	12
LAST	5	NEXT	12
NEXT	4	MOREOVER	7
THING	4	THEREFORE	7
FACT	4	FACT	6
HAND	4	FIRSTLY	6
THEREFORE	4	ALL	5
THIRD	4	SECONDLY	5
NOW	3	THEN	5
ALL	2	THIRD	5
ACTUALLY	2	FINALLY	4
ADDITION	2	ANYWAY	3
FINALLY	2	OTHERWISE	3
MOREOVER	2	FURTHERMORE	2
BESIDES	1	LASTLY	2
FURTHERMORE	1	NOW	2
RESULT	1	REASON	2
WITH	1	RESULT	2

		THINGS	1
		TWO	1
		ADDITION	1
		CERTAINLY	1
		CONCLUSION	1
		CONTRAST	1
		ONE	1
		PARTICULAR	1
		SEE	1
TOTAL	423	TOTAL	491

5. Discussion

The scope of the present study was quite broad and has also suffered from some methodological limitations. The corpus itself is without doubt a useful resource for inquiry into the area of language acquisition and writing development, particularly with respect to the population dealt with in the present study.

McNeill (2007) reports that word count (up until about 400 words) correlates among other factors including lexical diversity, most strongly with holistic ratings. This clearly supports the importance of development of fluency in writing. Fluency gains measured in number of tokens produced were observed in both tasks but there are two considerations, which suggest caution in interpreting these results. First, while the gains appear to be statistically significant, we must ask what length of text might represent real progress in fluency. While presenting similar data at an international conference, scholars and educators from European contexts were surprised at the extremely low number of tokens reported.

Does a gain of between 10 and 20 tokens in 20 minutes of writing really represent an increase in fluency? Of course, second language learners in other contexts and from different L1 backgrounds may likely demonstrate different patterns of gains.

A comparable native speaker corpus may provide a useful baseline from which objective measures of native-like fluency can be derived. The second consideration is the effect of practice. Research has shown that task repetition can have positive effects on measures of fluency (Bygate, 2001 cited in Larsen-Freeman, 2006) and since these two writing tasks were an exact repetition of the tasks performed one year earlier, these effects may well confound the gains in fluency reported here. Finally, the definition of fluency must be scrutinized. Other measures of fluency including MLTU may also be representative of fluency and should not be ignored. In the continuing work on this corpus we are at present in the beginning stages of constructing a parallel native speaker corpus and beginning analysis of syntactic complexity, the data from which may provide a fuller spectrum of measures to describe fluency.

For the present study, two measures of lexical diversity were employed and both consistently showed gains in lexical diversity in general and particularly within the first two thousand most frequent words. As outlined earlier, measures such as the Guiraud index have been criticized for their lack of resistance to text length effects. This has been mathematically proven and is not seriously contended but minimizing the effect of text length is certainly desirable. Nevertheless, this criticism has no serious implications for this study because as shown in the significant gains by writers in their second year in both tasks, the texts are longer.

Therefore, we should expect a declining TTR and, if anything, this would work against finding significant gains between a shorter and longer text. Significant gains were found and may even be reported with greater confidence even if the text-length has an effect. The second measure which was employed was the D value as calculated by VocD. Because of its robustness and resistance to text-length effects, it was chosen as an additional supporting measure. As described earlier, the somewhat unorthodox step of combining texts from two genres (i.e., the narrative and argumentative writing tasks) was adopted. Some may argue, with good cause, that this artificially inflates the D values obtained, especially considering the short lengths of the texts examined. Indeed, the reported means appear to be quite high when comparing these to those cited for different populations by the developers of the measure. The high values obtained in some cases may be a result of the combination but may also be influenced by short text lengths in which there is less chance of repetition. Whether these results can be used in comparing the lexical diversity of the population represented in this corpus with others is a problematic question. However, the segmental TTR curves observed for both groups appeared normal with no error rates giving any cause for alarm. The lexical diversity measure D as calculated by the VocD is reported to be very stable for short texts but the texts which comprise the present corpus are extremely short and have presented new problems for dealing with the construct of lexical diversity in low-proficiency learner texts. Perhaps other avenues of lexical description based on word lists, such as lexical sophistication (or the proportion of rare words in text), as measured by Meara & Bells's (2001) P_Lex tool, may be a more practical method for describing low-proficiency learner writing.

Grammatical accuracy as measured by the obligatory occasion analysis showed patterns of gain between the first and second year but were ultimately not usable for making any paired comparison for students between tasks. The principal problem is that even though the task was to write a narrative and past tenses were expected to occur, certainly not all of the past tense types were present in all students writing and furthermore they may be 'optional' as a matter of writing style. For example, a writer may choose to write in the present as a means of evoking more immediacy in the text. The obligatory occasion analysis may be more suited to translation-type tasks which specifically elicit certain tenses whose presence and patterns of use and misuse can be more confidently interpreted. Furthermore, the narrow spectrum of grammar considered does not provide a comprehensive profile of grammatical accuracy. A wider approach, such as that employed by Abe and Tono (2005) may be more preferable in future analysis.

Finally, the use of rhetorical cohesive devices as measured by use of a variety of sentence-level transitions shows very moderate gains by the writers in the argumentative writing task. As Hinkel's (2002) native speaker baseline data shows, patterns of use of these types of connectors by native speakers is often at odds with what might be expected. More sophisticated writing may in fact exhibit less use of these devices. Furthermore, certain developmental stages may be characterized by the overuse of these devices. To provide more depth in the description of the instances of use, a concordance was employed to examine the variety of single and multi-word transition units. This data is, however, only descriptive and interpretations are difficult to make. Perhaps what is necessary is a broader investigation of cohesive devices, such as use of

hedges and demonstratives.

At the conclusion of the present study a number of directions for methodological revision and future analysis have become clear. If investigation into grammatical accuracy and rhetorical cohesive devices is to continue, the data must be revisited with measures which are more suited to corpus data. The construct of fluency must take syntax into account. Currently, methodology is being developed to analyze the longitudinal corpus in terms of syntactic complexity. Finally, a native-speaker corpus is currently being constructed which comprises samples of writing of native speakers from a similar age group. Once a reasonable quantity of data has been collected, it will provide an invaluable resource from which to draw comparisons from learner data.

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)
- Adolphs, S. (2006). *Introducing Electronic Text Analysis: A practical guide for language and literary studies*. New York: Routledge
- 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
- Bygate, M. 2001. 'Effects of task repetition on the structure and control of oral language' in M. Bygate, P. Skehan and M. Swain (eds): *Researching Pedagogic Tasks*. Harlow: Longman.
- 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.
- Davidson, F. (1991). Statistical support for training in ESL composition rating. In L. Hamp-Lyons (Ed.), *Assessing Second Language Writing* (pp. 155-165). Norwood, NJ: Ablex.

- Durán. P; Malvern D., Richard, B. and Chipere, N.(2004). Developmental Trends in Lexical Diversity. *Applied Linguistics*. 25, 2; Research Library pg. 220
- 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
- Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. London: Longman
- 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
- Jonansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. Lund University, Dept. of Linguistics and Phonetics. *Working Papers* 53, 61-79
- Larsen-Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. *Applied Linguistics*. 27/4: 590–619
- Laufer, B. (1994). The Lexical Profile of Second Language Writing; Does it change over time? *RELC Journal*, 25, 21-33.
- Malvern, D. and Richards, B. (1997). 'A new measure of lexical diversity' in A. Ryan and A. Wray (eds.): *Evolving Models of Language*.

Papers from the annual Meeting of the BAAL held at the University of Swansea, September, 1996. Clevedon, U.K: Multilingual Matters, pp. 58-71

- Malvern, D. and Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19 (1) 85–104.
- MacEnery, T., Xiao, R. and Tono, Y. (2006). *Corpus Based Language Studies: an advanced resource book*. Routledge: London
- McCarthy, P. and Jarvis, S. (2007). Vocab:A theoretical and empirical evaluation. *Language Testing*. 24, 459-488.
- McNeill, B. (2007). “A summary of ‘A comparative statistical assessment of different types of writing by Japanese EFL college students’”. *Journal of School of Foreign Languages* 33: 141-155
- Meara, P., & Bell, H. (2001). P_lex : A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5–19.
- Mochizuki, Struc, N., Iso, T. & Maeno, F. (2008). 日本人英語学習者の文法項目の習得 (Japanese Learner Grammatical Morpheme Acquisition) 麗澤大学言語研究センター・プロジェクト・リポート
- 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>
- Pica, T. (1984). Methods of morpheme quantification: their effects on the interpretation of second language data. *Studies in Second*

Language Acquisition 6, 69-78.

- Scott, M., (2004), WordSmith Tools version 4, Oxford: Oxford University Press.
- Struc. N, and Wood, N. (2009) Using a Learner Corpus to Identify Language Learner Profiles. *Reitaku Review*. 15 (pp.53-82)
- 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
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Second Language Teaching & Curriculum Center, University of Hawaii
- Yu, G. (2007). Lexical Diversity in MELAB Writing and Speaking Task Performances. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, Volume 5, English Language Institute, University of Michigan

