The Incidental Development of L2 Proficiency in NS-NNS Email Interactions

Glenn Stockwell
Kumamoto Gakuen University

Michael Harrington
The University of Queensland

ABSTRACT

Recent research suggests that email can be a powerful motivator for authentic L2 interaction, but little is known about the efficacy of this medium in the development of target language proficiency. The present study addresses this issue by examining email exchanges between university learners of Japanese as a foreign language and native Japanese university students. Of interest is the effect of email interactions on the incidental development L2 syntax and vocabulary as reflected in both quantitative and qualitative measures. Messages sampled at regular intervals over a 5-week collection period indicated a reliable increase in syntactic development as reflected in several measures of structural mastery as well as in qualitative ratings supplied by native speaking raters. There was no evidence of quantitative development for vocabulary, but qualitative ratings did show a small improvement over the collection period. A sharp drop-off was noted between the first and the second samples across all measures, with learner performance then improving steadily till the end of the study. The findings are related to an interactionist account of L2 development that is embedded in the framework of computer-mediated communication (CMC). Issues in research methodology are also discussed.

KEYWORDS

Incidental Syntactic Development, Computer-mediated Communication, Authentic Interaction

INTRODUCTION

The introduction of email into the second language (L2) classroom opens up the world to learners who otherwise have limited access, or no access, to native speakers or other learners. The asynchronous and readily accessible nature of email means that individuals may send and receive messages at a time which is convenient to them, and they are not bound to a particular place (Warschauer, 1995b). Indeed, it has been noted that email, and more generally computer-mediated communication (CMC), fundamentally challenges the traditional view of the classroom as a purely same time-same place entity (Sussex & White, 1996, Paramskas, 1999).

The increased use of email has been accompanied by an increase in research examining how this medium affects the L2 learning and teaching process. Email has been embraced by many as a tool for teaching, and much has been written about its use in the L2 classroom (Soh & Soon, 1991; Beauvois, 1992, 1995, 1997, 1998; Barson, Frommer, & Schwartz, 1993; Cononelos & Oliva, 1993; Warschauer, 1995a; Chapman, 1997; Collentine & Collentine, 1997; Leh, 1997; Gray & Stockwell, 1998; Nelson & Oliver, 1999; Stockwell & Levy, 2001). The effect that email has on a range of L2 learning processes has been examined. These effects include the role of email in lowering learner anxiety (Warschauer, 1995a); its effect on interaction and participation structure (Kalaja & Leppanen, 1991); how email use interacts with individual differences in personality (Beauvois & Eledge, 1996) and attitudes (Beauvois, 1995); how the medium affects thematic and linguistic aspects of L2 communication (Tella, 1992; Holliday, 1996), as well as cognitive demands (Debski, Gassin, & Smith, 1997). Work has also appeared that compares email with other forms of communication in an instructional setting (Chun, 1994; Maynor, 1994; Deal, 1995; Kern, 1995; Condon & Cech, 1996; Chapman, 1997; Warschauer, 1996, 1997).

In addition to interest in how the use of email may affect learning processes, there is a small, but growing literature on effects of email usage on learning outcomes (Warschauer, 1997; Chapelle, 1998). For example, Flórez-Estrada (1995) attributed improved grammatical competence in Spanish, as measured by ACTFL proficiency guidelines, to the use of email in an L2 writing class. Similarly, Lee (1997) and Gray & Stockwell (1998) suggested that email use facilitates the acquisition of L2 cultural knowledge. Saita, Harrison, & Inman (1998) reported heightened learners' perceptions of their proficiency in Japanese as a result of interactions with native speakers. Although this research is suggestive, it remains to be seen how L2 proficiency develops in the context of email interactions. The crucial role that authentic interaction plays in L2 development is widely recognised (Long, 1996). Research on the role of L2 email interaction thus has potential implications for a fundamental issue in SLA theory. However, the computer-mediated nature of email interactions also addresses more fundamental issues concerning the nature of increasingly widespread electronic communication. The differences between authentic interaction in face-to-face and computer-mediated contexts, and the implications this issue has for L2 learning and teaching is emerging as a key issue in SLA research (Salaberry, 2000; Harrington & Levy, 2001).[1]

THE STUDY

The study examines L2 learner output over a series of exchanges with a native speaker partner for evidence for the incidental development of L2 syntax and lexis. The data were collected as part of a larger project that examined email exchanges as a pedagogic task in the Japanese L2 classroom (Stockwell, 2000).[2] The email exchanges were part of a unit in intercultural communication involving Japanese as a foreign language (JFL) students at an Australian uni-

versity and Japanese students at a Japanese university. Two types of evidence were sought for the development of proficiency over the course of the exchanges. We wanted to see whether L2 development would be reflected in linguistic text features, as measured by the average number of words per T-unit, percentage of error-free T-units, and average number of words per error-free T-units, and in qualitative measures consisting of native-speaker ratings of syntactic mastery. Potential gains in lexical mastery were assessed by calculating type-token ratios and native speaker ratings of lexical mastery. Performance on all measures was examined at specific points during the interaction period.

Subjects

Subjects were third year university JFL students at an Australian university. Advanced level students were chosen because they have sufficient linguistic resources necessary to sustain interactions with native speakers (Saita, Harrison, & Inman, 1998).

Data were collected over two 5-week periods. Each subject participated in only one data collection period, with 7 subjects observed in the first data collection period and 13 in the second. The subjects were native speakers of English, with the exception of two Taiwanese and one Korean speaker. Four of the students had visited Japan for approximately 12 months.

Data Collection

The email exchanges took place as part of a joint project between third year JFL learners and Japanese students in an intercultural communication course in a Japanese university. The goal of the email interactions was for the students to learn more about the everyday lifestyle of a university student in the other country. All exchanges were in Japanese, and the Japanese students were explicitly instructed to focus on the content of the exchange and not attempt to correct the Japanese produced by the Australian students.

The Japanese and Australian students were told that they should try to maintain at least 4 to 5 exchanges per week with their partners over a five week period, with a different topic being assigned each week. The topics were: *Self introductions*, *Perceptions of ourselves and others*, *Dining out*, *Relaxation and leisure*, and *Dating and socializing*. The topics were provided as a guide, and students were allowed to discuss other items of interest if they wanted to. Most of the dyads stayed on topic over the course of the data collection.

One hour of class time was allotted to the project, but exchanges were also strongly encouraged outside of class time. Messages were sent at a range of different times, from early morning through to very late at night, by both the native speakers and the JFL students. Before the data collection began, students were given practice in sending and receiving email to ensure that they possessed the technical skill needed to participate in the project (Jor & Mak, 1994). All the messages by the JFL students were sent from language laboratories at the respective schools using Japanese *Netscape 3.0*.

Data Analysis

After the interactions were concluded, the emails were sorted according to sender, date, and time, and the interactions between the native speakers and nonnative speakers were examined. The data presented in this study were obtained from subjects who completed 15 or more email exchanges in the course of the five-week period.[3] The text feature measures were calculated for all the message samples, and the ratings were elicited for samples from the 1st, 5th, 10th, and 15th message produced by each JFL subject. A subset of the NS data was also measured as a baseline and as a means to establish interrater reliability. The messages were spread out over the 5-week period in which all of the 15th messages were produced in the 5th week, with the exception of one student.[4] In order to control for the effect of length of message on the rating, the first five lines of each message were analyzed.[5]

Linguistic Text Features

As a measure of syntactic development, the number of words per minimal terminal units (T-units) was calculated for all of the messages produced by both the native speakers and nonnative speakers. The T-unit (or minimal terminal unit) is a unit of syntactic production that coincides (in English) with an independent clause and any attached or embedded material.[6] The length of the T-unit, as measured in the average number of words per unit, reflects the individual's ability to use longer, more complex structures.[7] The average number of words per unit is assumed to reflect the relative syntactic complexity of the unit and has been shown to be a stable index of the development of L1 (e.g., Nutter, 1981) and L2 (e.g., Larsen-Freeman, 1983) grammar. However, the average number of words per T-unit measure does not take into account the accuracy of the structure produced. As a result, the average number of error-free T-units has been used as an additional measure of syntactic mastery for L2 learners (Larsen-Freeman, 1978) and has been shown to be particularly sensitive to L2 development. Similarly, the amount of error-free T-units as percentage of total production has also been used as a measure of overall syntactic accuracy in production (Scott & Tucker, 1974).

T-unit measures have been used more recently by Casanave (1994) to examine the complexity and accuracy of writing of Japanese university level EFL students and also to compare the syntactic complexity of language produced by ESL learners engaged in synchronous and asynchronous CMC interactions (Salaberry, 2000). Although other units of syntactic production are available, the T-unit is used here because its application is straightforward and because it has been widely-used (for a review of the relevant issues in the use of the T-unit as an L2 syntactic measure, see Bardovi-Harlig, 1992). In this study, the measure of T-units in Japanese has followed the method first outlined by Harrington (1986), who defined the main clause in Japanese according to the classificatory framework developed by Martin (1975).

Type-token ratios (TTRs) are used as measures of lexical diversity and have been used to examine L1 development in children (e.g., Snow, Hassing, Jobse, Joosten, & Vorster, 1974; Garrard, 1988; Watkins, Kelly, Harbers, & Hollis, 1995) and in work on language variation (Evans & King, 1981).[8] A learner of low proficiency would be thought to use a smaller number of lexical items, recycling them more often, thereby providing a lower ratio. Research into second language acquisition has used TTRs in the measurement of corpus data of written and spoken speech (Halliday, 1992) and has already received some support in email (Warschauer, 1996; Holliday, 1996).

The T-unit counts and TTRs were calculated by the first author. In order to ensure coding reliability, a sample for 20 messages—representing about 7% of the total corpus analysed—was also coded by another individual. Interrater reliability for the two raters in this exercise was calculated; for the T-unit analysis a correlation of $r = .98$ ($p = .001$) was obtained, and for the TTR analysis, $r = .97$ ($p = .000$).

Proficiency Ratings

Ratings of syntactic and vocabulary mastery were obtained from two native-speaking JFL teachers. The raters rated the message samples using a six-point proficiency description scale for syntax and vocabulary. The description scale was adapted from the Foreign Service Institute interview procedure (Hughes, 1989) and the International Second Language Proficiency Ratings (ISLPR) which have been specially developed for email (Wylie & Ingram, 1999; Stockwell, 2000). The syntax and vocabulary rating scales are presented below.

Ratings for syntactic mastery.

1. Grammar almost entirely inaccurate phrases.
2. Constant errors showing control of very few major patterns.
3. Frequent errors showing some major patterns uncontrolled.
4. Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding.
5. Few errors, limited mostly to particle usage.
6. No more than one error during the message.

Ratings for lexical mastery.

1. Vocabulary inadequate for conveying even the most simple messages.
2. Vocabulary limited to basic personal and survival areas (time, food, etc.)
3. Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of more complex issues.
4. Vocabulary appears inadequate to discuss topics in depth. Sometimes the correct vocabulary is selected but incorrectly written.
5. Specialist vocabulary broad and precise; general vocabulary adequate to cope with complex discussions.

6. Vocabulary apparently as accurate and extensive as that of an educated native speaker.

In addition to the proficiency description, the raters were also asked to judge each sample for overall proficiency in syntax and vocabulary on a 10-point scale, as in

How do you rate this person's overall proficiency in syntax (vocabulary)?

Very Low									Very High
1	2	3	4	5	6	7	8	9	10

The raters were also asked to provide an assessment of overall proficiency, based on the 9-point ISLPR rating scale. Two raters, who were native speakers of Japanese and experienced JFL teachers in Australian universities, judged all the samples. The samples were rated in a random order with all identifying features of the messages removed. Mixed in with the nonnative speaker data were 10 samples of native speaker messages (constituting a figure of approximately 5% of the data set). The raters were not told that native speaker samples were included in the samples. After all of the messages had been rated, interrater reliability between the two raters was calculated. There was close agreement between the two raters on all the measures (r between .80 and 1.00, $p < .001$) obtained for the three ratings for syntax, vocabulary, and the overall ISLPR score. Average agreement between the two raters across all measures was $r = .96$ ($p < .001$). The native speaker samples were all rated consistently higher than the learner samples and showed much less variability. Both the JFL and native speaker results are reported in the tables below.

The effect of interaction over the 5-week period on the performance measures was tested using repeated measures one-way analyses of variance (ANOVA) on the individual text features and rating measures. Message number (1^{st} x 5^{th} x 10^{th} x 15^{th}) was the independent variable, and the textual feature counts and ratings were the dependent measures for the respective measurement categories. Planned contrasts were made between the 1^{st} and 5^{th} messages, the 5^{th} and 10^{th} messages, the 10^{th} and 15^{th} messages, and between the 1^{st} and the 15^{th} messages.

Results

The results for syntactic development are presented first, followed by an examination of the results for the vocabulary measures and overall proficiency ratings.

Syntactic Development: T-unit Measures

The means and standard deviations for the three T-unit measures across the

1st, 5th, 10th, and 15th messages for both JFL learners and their Japanese L1 email partners are presented in Table 1.

Table 1
Means and Standard Deviations for the Text Features Measures (T-unit and Type Token Ratio) across the 1st, 5th, 10th, and 15th Messages for JFL Learners and Japanese L1 Email Partners.

	JFL Learners (n = 20 for each message)							
	Syntactic						Vocabulary	
	Average # of words per T-unit		Average # of words per error-free T-unit		% of error-free T-unit		Type-token ratio	
Messages	M	SD	M	SD	M	SD	M	SD
1	10.9	.65	10.9	.69	86	4.5	.88	.04
5	9.2	.84	9.3	.96	72	10.6	.70	.04
10	10.3	.46	10.4	.62	79	13.0	.80	.03
15	11.0	1.11	11.6	1.12	81	8.7	.81	.02
Baseline Japanese L1 measures	10.8	.37	10.8	.37	100	.00	.81	.01

The means for the JFL learners reflect improvement over the fifteen messages. The average number of words per T-unit and average number of words per error-free T-unit both showed a small mean increase from the 1st to the 15th message, while the percentage of error-free T-units in the samples showed a decrease of 5% from the 1st to the 15th message. As expected, performance by the Japanese L1 partners was consistent across all 15 messages, ranging from .79 to .84, with a very small standard deviation.

One-way ANOVAs were performed on the three T-unit measures for the JFL subjects. All three measures yielded a statistically significant difference in performance across the four message levels. For average number of words per T-unit, $F(3,19) = 35.5$, $p = .0001$; for average number of words per error-free T-unit, $F(3,19) = 35.4$, $p = .0001$; and for the percentage of the error-free T-units $F(3,19) = 16.86$, $p = .0001$. Planned contrasts were also carried out to compare performance at four points in the 15-message interaction period. These results are presented in Table 2. Of the syntactic measures, only the percentage of error-free T-units in messages 10 versus 15 and average words per T-unit in messages 1 versus 15 were not significant at $p < .05$.

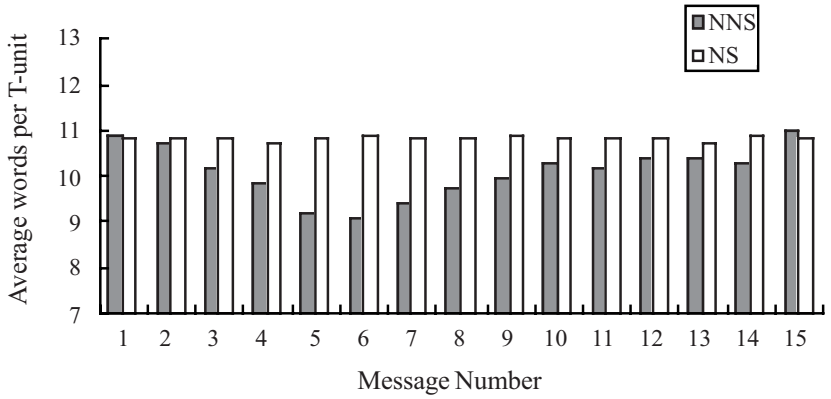
Table 2
Planned Contrasts for Textual Features Measures (T-unit and Type-Token Ratio)

Contrast	Difference	Measure	df	Sum of squares	F value	*p* value
Messages						
1 vs. 5	-1.70	Avg. words per T-Unit	1,19	26.08	55.07	.0001
	-1.60	Avg. words per error-free T-Unit	1,19	25.60	46.76	.0001
	-14.00	% of error-free T-units	1,19	2014.98	49.11	.0001
	-.18	Type-token ratio	1,19	26.08	55.07	.0001
5 vs. 10	1.10	Avg. words per T-Unit	1,19	10.61	22.40	.0001
	1.10	Avg. words per error-free T-Unit	1,19	12.21	22.30	.0001
	7.00	% of error-free T-units	1,19	511.95	12.48	.0008
	.10	Type-token ratio	1,19	10.61	22.40	.0001
10 vs. 15	.70	Avg. words per T-Unit	1,19	5.40	11.40	.0013
	1.20	Avg. words per error-free T-Unit	1,19	15.50	28.31	.0001
	2.00	% of error-free T-units	1,19	38.81	.95	.3348
	.01	Type-token ratio	1,19	5.40	11.40	.0013
1 vs. 15	.10	Avg. words per T-Unit	1,19	5.40	.48	.4935
	.70	Avg. words per error-free T-Unit	1,19	5.65	6.27	.0152
	-5.00	% of error-free T-units	1,19	257.05	10.27	.0022
	-.07	Type-token ratio	1,19	5.40	.48	.4935

Consistent across all the measures was a "first-message" effect, in which very high performance in the 1st message is followed by an immediate drop-off. This drop-off is steady from the 2nd messages to the 5th messages and is then followed by a steady increase in performance until the end. In all three measures the 1st message averages were among the highest produced, with performance only approaching or exceeding the level of the first message at the end of the interaction period.

Figure 1 plots this drop-off and subsequent recovery across the 15 messages for the average number of words per T-unit for both the L2 and L1 subjects.

Figure 1
Average Numbers of Words per T-unit Compared with Native Speaker Average for the High-interaction Category.



Like earlier research (Larsen-Freeman, 1978), the results on average number of words per T-unit and the average number of words per error-free T-unit differed, with the latter showing a greater increase over the last five messages. The mean differences between the 1st and the 15th messages were not significant for average number of words per T-unit (p = .493) and average number of words per error-free T-unit (p = .015), although the difference for the latter approached the .01 alpha level. The difference between the 1st and 15th messages for percentage of error-free T-units was significant (p = .002) although it was in the opposite direction: the first messages had a significantly higher percentage of error-free T-units. A possible explanation for the greater percentage of error-free T-units in the first message is that the learners were more careful in the initial exchanges. The "first-message" effect will be considered in the discussion section below.

Syntactic Development: Proficiency Ratings

Message samples were rated by two native speakers for proficiency in syntactic usage of Japanese. The syntactic proficiency descriptions refer to the 6-point proficiency scales adapted from the FSI interview procedure, while the rating of overall syntactic proficiency was on a 10-point Likert scale ranging from 1 ("Very Low") through to 10 ("Very High"). Because of the high correlation of scores between the two raters (r = .8 ~ .9 across the samples), the two rater scores were averaged (see Table 3).

Table 3
Means and Standard Deviations for the Syntax and Vocabulary Ratings across the 1st, 5th, 10th, and 15th Messages for JFL Learners and Baseline Japanese L1 Email Partners

	FSI-style proficiency level description rating (out of 6)				Overall proficiency rating (out of 10)			
	Syntax		Vocabulary		Syntax		Vocabulary	
Messages	M	SD	M	SD	M	SD	M	SD
1	4.35	.80	4.25	.75	7.93	1.04	7.93	.99
5	3.78	.80	3.70	.70	7.30	.94	7.33	.99
10	4.15	.75	4.18	.67	7.73	.95	7.78	.90
15	4.55	.83	4.53	.68	8.23	.95	8.23	.79
Baseline Japanese L1 measures (n = 10)	5.95	.16	5.85	.24	9.80	.50	9.70	.63

One-way ANOVAs were performed on the two rating measures. Both measures reached statistical significance for the overall analysis. For the FSI-style syntactic ratings, $F(3,19) = 2.24$, $p = .0028$; and for the overall syntactic proficiency rating, $F(3,19) = 1.76$, $p = .0001$. Planned contrasts were also carried out and the results are presented in Table 4. The decrease from the 1st to the 5th message was significant for both rating scales. The results show then a reliable increase across the successive contrasts at the 5th versus 10th and the 10th versus 15th messages. The only contrast that did not reach significance at the .05 level was the difference between the 1st versus 15th message for the overall rating ($p = .069$).

In summary, the results from the two syntactic rating measures were similar to those obtained in the text features analysis. Performance on the first message was relatively high and then dropped significantly until the 5th message, at which point it started to consistently increase through the 10th and 15th messages. The mean increase from the 1st to the 15th message for overall syntactic proficiency was the only measure that did not reach significance.

Lexical Development: Type-Token Ratio

The means and standard deviations of the type-token ratios across the 1st, 5th, 10th and 15th messages were presented in Table 1 above for both JFL learners and their Japanese L1 partners. As with the measures for syntactic development, there is a marked drop in the TTR from the 1st to the 5th message which shows some recovery after the 5th message until the end of the interaction period. Unlike syntax, however, the TTR does not exceed that of the 1st message, demonstrating only very small increases from the 10th to the 15th message. The language output of the L1 speakers was again very consistent with a minimal standard deviation across the interaction period. The one-way ANOVA results

overall showed a statistically significant difference across the message levels $F(3,19) = 1.41$, $p = .0032$. The individual planned contrasts in Table 2 show a significant drop from the 1st to the 5th message at *(p* = .0001) and then mean increases from both the 5th to the 10th and from the 10th to the 15th messages. These contrasts were all statistically significant.

TABLE 4
Planned Contrasts for Syntax and Vocabulary Ratings

Contrast	Rating	Domain	df	Sum of squares	F value	*p* value
Messages						
1 vs. 5	FSI-style description	Syntax	1,19	3.306	35.041	.0001
		Vocabulary	1,19	3.035	30.316	.0001
	Overall proficiency	Syntax	1,19	3.906	14.940	.0003
		Vocabulary	1,19	3.600	15.271	.0002
5 vs. 10	FSI-style description	Syntax	1,19	1.406	14.904	.0003
		Vocabulary	1,19	2.256	22.612	.0001
	Overall proficiency	Syntax	1,19	1.806	6.908	.0110
		Vocabulary	1,19	2.025	8.590	.0049
10 vs. 15	FSI-style description	Syntax	1,19	1.600	16.958	.0001
		Vocabulary	1,19	1.225	22.612	.0009
	Overall proficiency	Syntax	1,19	2.500	9.562	.0031
		Vocabulary	1,19	2.025	8.590	.0049
1 vs. 15	FSI-style description	Syntax	1,19	.400	4.239	.0441
		Vocabulary	1,19	.756	7.570	.0079
	Overall proficiency	Syntax	1,19	.900	3.442	.0687
		Vocabulary	1,19	.900	3.818	.0556

Lexical Development: Proficiency Ratings

The message samples were rated by two native speakers for degree of lexical mastery in the same manner as the syntactic ratings were carried out above. The ratings showed the same drop off from the 1st message to the 5th message observed in the other measures followed by a steady increase until the end of the

interaction period. One-way ANOVAs were performed on the two proficiency rating measures. Both measures reached statistical significance for the overall analysis. For the FSI-style vocabulary ratings, $F(3,19) = 2.06$, $p = .0002$; and for the overall proficiency rating, $F(3,19) = 1.59$, $p = .0001$. Planned contrasts are reported in Table 4 above. There was a reliable decrease from the 1st to the 5th message for both the FSI-style proficiency description scores and for the 10-point proficiency scales. The successive contrasts were also significant for all measures except for the 10-point overall vocabulary proficiency rating from the 1st to the 15th message ($p = .0556$).

The results for the two lexical measures are thus similar to those for syntax. The initial messages had a high score, a marked decrement in performance at message 5, and then a subsequent rise until the end of the interaction period. The mean difference between the 1st and the 15th messages for the FSI-style rating of vocabulary mastery was reliable ($p = .007$).

Overall Language Proficiency

Overall language proficiency was measured using International Second Language Proficiency Ratings (ISLPR). The ratings for overall proficiency closely mirrored those obtained for syntactic and lexical development. Table 5 shows the same pattern of the drop from the 1st message to the 5th message followed by an increase until the end of the interaction period. The overall ANOVA was significant, $F(3,19) = 2.39$, $p = .0021$. The drop in overall proficiency from the 1st to the 5th message was significant at $p = .0001$, as were subsequent increases from the 5th message to the 15th message.

Table 5
Planned Contrasts for ISLPR Overall Language Proficiency

Mean ISLPR Scores		Contrast	df	Sum of squares	F value	*p* value
Message 1						
M = 3.4	SD = .73	Message 1 vs. 5	1,19	1.406	20.023	.0001
Message 5						
M = 3.0	SD = .66	Message 5 vs. 10	1,19	1.225	17.443	.0001
Message 10						
M = 3.3	SD = .71	Message 10 vs. 15	1,19	.400	5.696	.0204
Message 15						
M = 3.5	SD = .66	Message 1 vs. 15	1,19	.306	4.361	.0413

The results can be briefly summarized as follows. Due to the presence of a pronounced "first-message" effect, there was a nonlinear increase in syntactic and lexical mastery over the entire interaction period (1st - 15th messages). There was a systematic decrease from the 1st to the 5th message followed by a systematic increase over the remainder of the interaction period. Given that the 5th - 15th message took place over a period of a month, the gains are impressive.

DISCUSSION

The email messages generated by advanced JSL learners over the course of the 5-week interaction period reveals a consistent improvement by the learners on most of the measures. Both the text features and the proficiency ratings showed steady improvement following the fall-off after the first message. This "first message" effect will be discussed below.

The JSL learners made gains of between .7 and 1.3 words per error free T-unit between the 1st and 15th and 5th and 15th messages, respectively, and a gain of 1.8 words per T-unit from the 5th to the 15th messages. There was also a gain of 9% for the percentage of error-free T-units. As noted, the largest gains were made between the 5th and 15th messages, the period in the interaction where the interlocutors were beyond the initial stage of getting acquainted and had achieved some degree of familiarity. The self-introductions had been completed by this time, and learners had exhausted their stock of formulaic expressions (e.g., "I am a student"). Improvements in the range of 1-2 words per T-unit and error-free T-units in this study compares with an improvement of two words between the third and fourth year Japanese university students in the small cross-sectional study reported in Harrington (1986). Although the magnitude of gains was similar for the two studies, the fourth year students in Harrington had lower T-unit scores overall (9.45 words per T-unit and 8.20 words per error-free T-units, compared to 11 and 11.6 in the students described here). They also had a lower percentage of error-free T-units (69% versus 81% for the present study). Methodological differences between the two studies limit the conclusions that can be drawn, but the similarity in the magnitude of gains across the five weeks in the present study, representing an additional year of study found in Harrington, is worthy of note.

First Message Effect

A recurring pattern across all the measures was the high performance on the first message. This level of performance was followed by a sharp drop-off in the next few messages and then a steady improvement until the end of the interaction period (15th message). The initial good performance may have been due to the fact that the JFL learners, who always initiated the first interaction, were attempting to convey a good first impression to their native-speaking interlocutors. Barson, Frommer, & Schwartz (1993) also observed a first message superiority effect in a study of intermediate learners of French involved in a collabo-

rative email project; contrary to the expectation of the authors, the learners produced surprisingly few mistakes. It could be argued that these learners (as with the students in the present study) wished to make a good "first impression" with their first message, hence the higher level of proficiency output in the message. (Others have also reported a drop off in motivation after the initial period [see Leh, 1997].)

One result of learners trying to create a good impression is that they pay closer attention to the language used. As the learners in the study became more familiar with their partners, and perhaps with the medium of electronic mail, they became less careful in writing the messages, and the language style used in the interactions seemed to be less of a focal point for the learners in the later messages (Stockwell, 2000). This "style-shifting" has been described by Tarone (1983), who suggested that learners can shift from their *superordinate* (or *careful*) style, which is more often found in formal situations, through to their *vernacular* style, which is more often associated with spontaneous language use. As the exchanges continued in the study reported here, the students no longer used their careful style but began producing unattended language forms based on their own developing interlanguage systems.

A second, possibly related explanation for the relatively higher performance on the first message may be due to the fact that in the first message the learners relied to a greater degree on well practiced formulaic utterances. Learners were most familiar with self introductions, which are essentially formulaic and commonly taught in the language classroom. It should be noted that the JFL learners always initiated the exchange. The learners made simple introductions of themselves and asked basic questions of their partners but did not have to respond to questions. As a result, learners were working with a repertoire of familiar structures that they were able to control in the initial messages. Of course, the shifts from more careful to more colloquial speech and from formulaic to more productive language are not necessarily independent. As the exchanges continued and the learners were asked to discuss an increasingly wider variety of topics, as well as respond to questions posed to them by the native speakers, they had to move beyond the use of their stock of formulaic phrases and structures.

The first message effect suggests that one-off studies in which learner production is sampled at one point, possibly when interacting with an unfamiliar interlocutor, may overestimate the learners underlying competence.

The length of the interaction period is also an issue. In this study, proficiency gains were evident until the final message. Research over a longer period would be beneficial in determining whether the gains continue or whether the learners cease to benefit from them after a certain point. It may be the short time-frame of this study which allowed such proficiency gains to occur. Tella (1991), for example, argued that motivational advantages of email are short-lived and that learners cease to continue with exchange projects after the initial excitement wears off. In addition, if the period of interaction is too short, learners do not have sufficient time to benefit from the interactions.

HOW DOES EMAIL INTERACTION PROMOTE L2 DEVELOPMENT

The computer-mediated (CM) email medium has both advantages and disadvantages vis-à-vis traditional face-to-face (FtF) interaction. The asynchronous nature of email interactions places less demand on learners' processing capacity, allowing them more time to focus on the linguistic cues. On the other hand, the lack of immediate physical cues such as gestures and physical referents makes interlocutors more dependent on linguistic cues to encode meaning. This absence obviously makes it more difficult for learners who are still developing their linguistic resources. Email messages are also static, allowing learners more time to comprehend and then respond. Because the interaction is mediated by the computer, learners may feel less anxiety than is common in FtF settings where considerations of saving face can heighten learner anxiety. Indeed, there is strong evidence that shy learners who may not typically produce much language in normal classroom discussions often produce significantly more in email discussion (Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996; Warschauer, 1996). This higher level of output means that these learners have increased opportunities to benefit from email interactions than they do in oral discourse.

Further, CM interaction appears to provide the salience for linguistic features that is normally associated with written text while maintaining the functional focus that purposeful FtF interaction entails. Thus, given these differences between email and traditional FtF interaction, we may ask how many of the learning outcomes observed here are due to interaction and how many are due to the specific computer-mediated nature of the interaction. Although the present study did not systematically compare FtF and CM interaction, other research has revealed differences between the two. Previous findings have shown the use of more complex structures in the target language in electronic versus FtF environments (e.g., Beauvois, 1992; Kelm, 1992; Kern, 1995; Warschauer, 1996). More recently, Salaberry (2000) has presented evidence revealing that morphosyntactic development (e.g., Spanish aspectual system) was more readily evident in a computer-mediated interaction task that it was in a matched FtF task. This area will become increasingly important in computer-mediated SLA research.

Email as Authentic Interaction

Interaction is generally recognized as an indispensable element in L2 learning.[9] Meaningful interaction in L2 requires learners to produce, comprehend, and revise the target language under the pressure of real-time discourse demands. Interaction facilitates learning because it provides a number of conditions that are important, if not indispensable, to the development of L2 proficiency. These conditions, which are referred to as a group as the Interaction Hypothesis (Long, 1996), can be classified into two categories. These are *task* conditions that facilitate the interaction process itself, including the opportunity to engage in meaningful interaction, and *psycholinguistic* conditions that facili-

tate the individual learner's uptake and processing of input. The email task used in the study meets these conditions in a compelling way.

Task Conditions Promoting Good Interaction

Learners must first of all engage in interaction tasks in which they are exposed to the target language in a meaningful setting. The email unit on intercultural communication required the JFL learners to actively engage in purposeful communication with their Japanese counterparts. In addition, the interaction task must be such that learners have the chance to modify their language use for the negotiation of meaning. This negotiation process helps learners comprehend the syntactic and semantic aspects of the linguistic input (Pica, 1991), and, as such, is a necessary precursor to internalizing knowledge in interlanguage. The tasks in the study presented here were information-gap activities that required the interlocutors to elicit and exchange information on a range of topics. Thus, the learners had to have, or find, the linguistic resources needed to achieve this end. Although the study did not examine the qualitative aspects of the interactions, it did uncover evidence that learners followed the examples written to them by the native speaker partners and that they altered their use of language from formal to informal or vice-versa dependent on the language used by their language partners (Stockwell, 2000).

Psycholinguistic Conditions Promoting Good Interaction

Good interaction provides several psycholinguistic conditions that facilitate learning. First of all, the linguistic characteristics of the target language input need to be made salient if learners are to incorporate these features into their interlanguage. There is convincing evidence that learners only acquire a very small proportion of the input to which they are exposed (e.g., Hulstijn, Hollander & Greidanus, 1996; Schmidt, 1999). The selective nature of learner uptake thus suggests that the salience of a specific linguistic feature is important to learning. Learners should have opportunities to produce target language output (e.g., Swain & Lapkin, 1995). The output allows learners to test hypotheses, receive feedback, develop automaticity and, in doing so, gain greater mastery of the L2 target (Gass, 1997). For learning to take place, learners need to notice errors and to correct their linguistic output (Swain, 1993). The correction of JFL syntactic errors by the native speakers was virtually nonexistent in the study, in part because they had been explicitly cautioned against providing such feedback. However, the Japanese email partners did provide implicit feedback in terms of recasts (Mackey, 1999). The following example shows a native speaker paraphrasing an incorrect utterance made by the nonnative speaker as a clarification request.

NNS: どんなアルバイトがしっているの。

NS: どんなアルバイトをしているかというと、バーテンダーをしています。

NNS: What kind of part-time job do you know?
NS: As for what kind of part-time job I do, I am a bartender.

The JFL learner has made an error in the use of the postpositional marker for the object (instead using the marker for indirect object) and has also made the wrong choice of word for "do" (instead using the word for "know"). In the response, the Japanese speaker repeats the incorrect part of the sentence in order to clarify (and perhaps implicitly correct) the meaning of the sentence before responding. It was also evident that learners were aware of their own errors, as was evident from the statements of apology or confirmation checks within their messages (see Stockwell, 2000). In these ways, the email task offers near optimal task conditions for interaction.

In conclusion, the findings reported here indicate that email interactions of the kind carried out in the study can lead to observable improvement in discrete measures of L2 knowledge. The nonlinear nature of the improvement, as well as the fact that it was obtained over a relatively short period of time (5 weeks) both indicate the need for further research.[10]

The present study makes a small contribution to the field of computer-mediated second language acquisition and further establishes the usefulness of email as part of a program of L2 instruction.

NOTES

[1] *ELL* and *TESOL Quarterly* have both devoted recent issues to this question.

[2] The study described here is a part of a larger study conducted by the first author as part of his Ph.D. dissertation. The larger study examined the use of email in a Japanese as a foreign language (JFL) class, with a focus on factors that might predict task success. These factors included the learners' computing experience, in-country experience, proficiency level, and academic specialization.

[3] This group is a subset of 48 JFL learners who participated in the original study. The remaining subjects produced less than 15 exchanges and are not considered here.

[4] Subject 42 of the study wrote more than any other student in the study, in excess of 30 messages, compared with the maximum of 15-16 messages of the other students in the category.

[5] The length of the message itself may also have bearing on development and the success of the interaction. Factors contributing to variation in length and number of turns are examined in Stockwell & Levy, 2001.

[6] Hunt (1977), who first developed the T-unit, defines it as "A single main clause … plus whatever other subordinate clauses or non-clauses are attached to, or embedded within, that one main clause."

[7] A unit which contains embeddings is more syntactically complex and is longer than a structure without the embeddings. Consider A and B:

A. *The boy is running. He is wearing a windbreaker. It is old. It is green.*

B. *The boy who is running is wearing an old green windbreaker.*

Both A and B express the same idea, but B expresses it in a single, longer T-unit, while A consists of four shorter T-units.

[8] TTRs are calculated by dividing the number of different words used by the total number of words (Arnaud, 1984). For example, if a learner produces the utterance, "The man is going to the shop," the number of tokens (the total number of words) is 7, and the number of types (the number of different words) is 6, giving a TTR of 0.85. In this way, we can calculate that the higher the number, the greater the lexical diversity used. When all of the lexical items used are different, the TTR is 1.

[9] Some researchers, Krashen most notably, have questioned the need for interaction, especially production, in L2 development (see Krashen, 1998).

[10] A possible explanation is that improvement in the measures reflects a more general effect of studying Japanese, that is, the development is the result of learning activities that were part of the class or possibly due to other out of class activities. The setting and nature of the original study precluded the use of a control group, and, thus, we cannot definitively rule this possibility out. However, since the gains were consistent across learners, we assume that the improvement does stem from the interaction in some way. A central issue, of course, is whether the gains are specific to the email task or whether it is reflected in more general proficiency gains.

REFERENCES

Arnaud, P. J. L. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. In *Practice and problems in language testing* (pp. 14-28). Colchester, England: University of Essex.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly, 26* (2), 390-395.

Barnwell, D. (1988). Some comments on T-unit research. *System, 16* (2), 187-192.

Barson, J., Frommer, J., & Schwartz, M. (1993). Foreign language learning using email in a task-orientated perspective: Interuniversity experiments in communication and collaboration. *Journal of Science and Technology, 2* (4), 565-584.

Beauvois, M. H. (1992). Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals, 25* (5), 455-464.

Beauvois, M. H. (1995). E-talk: Attitudes and motivation in computer-assisted classroom discussion. *Computers and the Humanities, 28,* 177-190.

Beauvois, M. H. (1997). Computer-mediated communication (CMC): Technology for improving speaking and writing. In M. D. Bush & R. M. Terry (Eds.), *Technology enhanced language learning* (pp. 165-184). Lincolnwood, IL: National Textbook Company.

Beauvois, M. H. (1998). Conversations in slow motion: Computer-mediated communication in the foreign language classroom. *The Canadian Modern Language Review, 54* (4), 198-217.

Beauvois, M. H., & Eledge, J. (1996). Personality types and megabytes: Student attitudes towards computer mediated communication (CMC) in the language classroom. *CALICO Journal, 13* (2 & 3), 27-45.

Casanave, C. (1994). Language development in students' journals. *Journal of Second Language Writing, 3* (3), 179-201.

Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology* [online journal], *2* (1). Available: polyglot.cal.msu.edu/llt/vol2num1/article1/index.html.

Chapman, D. (1997). A comparison of oral and email discourse in Japanese as a second language. *On-CALL, 11* (2), 31-39.

Chun, D. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System, 22,* 17-31.

Collentine, J., & Collentine, K. (1997). The compatibility of computer-mediated communication solutions with beginning level foreign language curricula. *Computer Assisted Language Learning, 17* (5), 411-425.

Condon, S. L., & Cech, C. G. (1996). Functional comparisons of face-to-face and computer-mediated decision making interactions. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 65-80). Amsterdam: John Benjamins.

Cononelos, T., & Oliva, M. (1993). Using computer networks to enhance foreign language/culture education. *Foreign Language Annals, 26* (4), 527-534.

Deal, N. (1995). Is the medium the message? Comparing student perceptions of teacher responses via written and email forms. *Proceedings of the National Educational Computing Conference, Baltimore, 1995,* 216-218.

Debski, R., Gassin, J., & Smith, M. (1997). (Eds.). *Language learning through social computing.* Melbourne: Horwood Language Centre & Applied Linguistics Association of Australia.

Evans, A. L., & King, T. R. (1981). *A stylistic content analysis of black college students* (Research Report). (ERIC Document Reproduction Service No. ED 236 731)

Flórez-Estrada, N. (1995). *Some effects of native-nonnative communication via computer email interaction on the development of foreign writing proficiency.* Unpublished doctoral dissertation, Graduate Faculty of Arts and Sciences, University of Pittsburgh.

Garrard, K. R. (1988). A mother's language to her twins: One child down syndrome and one child with normal development. *Journal of Childhood Communication Disorders, 12* (1), 11-23.

Gass, S. (1997). *Input, interaction and the second language learner.* Mahwah, NJ: Lawrence Erlbaum.

Gray, R., & Stockwell, G. R. (1998). Using computer-mediated communication for language and culture acquisition. *On-CALL, 12* (3), 2-9.

Halliday, M. A. K. (1992). *Spoken and written language* (3rd ed.). Oxford: Oxford University Press.

Harrington, M. W. (1986). The T-unit as a measure of JSL oral proficiency. *Descriptive and Applied Linguistics, 19,* 49-56.

Harrington, M. W., & Levy, M. (1999). CALL begins with "C": Interaction in computer-mediated language learning. *System, 29* (1) 15-26.

Holliday, L. (1996). The grammar of second language learners of English email messages. In S. Jager, J. Nerbonne, & A. Van Essen. (Eds.), *Language teaching and language technology* (pp. 136-145). Lisse: Swets & Zeitlinger.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: Cambridge University Press.

Hulstijn, J. H., Hollander, M., & Greidanus, R. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and recurrence of unknown words. *The Modern Language Journal, 80* (3), 327-339.

Hunt, K. W. (1977). Early blooming and late blooming syntactic structures. In C. R. Cooper & L. Odell (Eds.), *Evaluative writing: Describing, measuring, judging* (pp. 91-104). Urbana: National Council of Teachers of English.

Jor, G. C. K., & Mak, L. (1994). International email projects in ESL curriculum—what works, what doesn't? *Proceedings of the Tel-Ed '94 Conference.* Available: www.cuhk.edu.hk/ilc/Publi022.htm.

Kalaja, P., & Leppanen, S. (1991). Transaction and interaction in writing: The case of electronic mail. Communication and discourse across cultures. *AFinLa Yearbook (1991),* 2-10.

Kelm, O. R. (1992). The use of synchronous computer networks in second language instruction: A preliminary report. *Foreign Language Annals, 25* (5), 441-464.

Kern, R. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and characteristics of language production. *Modern Language Journal, 79* (4), 457-476.

Krashen, S. (1998). Comprehensible output? *System, 26* (2), 175- 182.

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly, 12* (4), 439-448.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger & M. H. Long (Eds.), *Classroom-oriented research in second language acquisition* (pp. 287-305). Rowley, MA: Newbury House.

Lee, L. (1997). Using Internet tools as an enhancement of C2 teaching and learning. *Foreign Language Annals, 30* (3), 410-427.

Leh, S. C. A. (1997). *Electronic mail in language learning.* Unpublished doctoral dissertation. Arizona State University, Phoenix, AZ.

Long, M. H. (1996). The role of linguistic environment. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21* (4), 557-87.

Martin, S. K. (1975). *A reference grammar of Japanese.* New Haven: Yale University Press.

Maynor, N. (1994). The language of electronic mail: Written speech? In G. D. Little & M. Montgomery (Eds.), *Centennial usage studies* (pp. 48-54). Tuscaloosa: University of Alabama Press.

Nelson, T., & Oliver, W. (1999). Murder on the Internet. *CALICO Journal, 17* (1), 101-114.

Nutter, N. (1981). Relative mean length of T-unit and sentence weight as indices of syntactic complexity in oral language. *English Education, 13* (1), 17-19.

Paramskas, D. M. (1999). The shape of computer-mediated communication. In K. Cameron (Ed.), *Computer assisted language learning: Media, design and applications* (pp. 14-31). Lisse: Swets & Zeitlinger.

Pica, T. (1991). Classroom interaction, negotiation, and comprehension: Redefining relationships. *System, 19* (4), 437-452.

Saita, I., Harrison, R., & Inman, D. (1998). Learner proficiency and learning tasks in email in Japanese. In K. Cameron (Ed.), *Multimedia CALL: Theory and practice* (pp. 221-228). Exeter: Elm Bank Publications.

Salaberry, M. R. (2000). L2 morphosyntactic development in text-based computer-mediated communication. *Computer Assisted Language Learning, 13* (1), 5-27.

Schmidt, R. (1999). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Scott, M., & Tucker, G. (1974). Error analysis and English language strategies of Arab Students. *Language Learning, 24,* 69-99.

Snow, C. A.-R., Hassing, Y., Jobse, J., Joosten, J., & Vorster, J. (1976). Mother's speech in three social classes. *Journal of Psycholinguistic Research, 5,* 1-20.

Soh, B. L., & Soon, Y. P. (1991). English by email: Creating a global classroom via the medium of computer technology. *ELT Journal, 45* (4), 287-292.

Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous interaction. *Language Learning & Technology* [online], *4* (1), 82-119. Available: llt.msu.edu

Stockwell, G. R. (2000). *An analysis of email based interactions between native speakers and advanced level nonnative speakers of Japanese.* Unpublished doctoral dissertation: University of Queensland, Brisbane, Australia.

Stockwell, G. R., & Levy, M. (2001). Sustainability of email interactions between native speakers and nonnative speakers. *Computer Assisted Language Learning, 14* (5), 419-442.

Sullivan, N., & Pratt, E. (1996). A comparative study of two ESL writing environments: A computer-assisted classroom and a traditional oral classroom. *System, 24,* 1-24.

Sussex, R., & White, P. (1996). Electronic networking. *Annual Review of Applied Linguistics, 16,* 200-225.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 50* (1), 158-164.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward. *Applied Linguistics, 16* (3), 371-391.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics, 4,* 143-163.

Tella, S. (1991). *Introducing international communications networks and electronic mail into foreign language classrooms: A case study into Finnish senior secondary schools* (Res. Rep. No. 95). Helsinki: Department of Teacher Education, University of Helsinki.

Tella, S. (1992). *Talking shop via e-mail: A thematic and linguistic analysis of electronic mail communication* (Res. Rep. No. 99). Helsinki: Department of Teacher Education, University of Helsinki.

Warschauer, M. (1995a). *Virtual connections.* Hawai'i: University of Hawai'i at Manoa, Second Language Teaching and Curriculum Centre.

Warschauer, M. (1995b). *Computer-mediated collaborative learning: Theory and practice.* Hawai'i: University of Hawai'i at Manoa, Second Language Teaching and Curriculum Centre.

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal, 13* (3), 7-26.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal, 81* (4), 470-481.

Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research, 38* (6), 1349-1355.

Wylie, E., & Ingram, D. (1999). *International second language proficiency ratings (ISLPR): Master general proficiency version.* Brisbane: Griffith University, Centre for Applied Linguistics and Languages.

AUTHORS' BIODATA

Glenn Stockwell is Associate Professor in the School of International Economics at Kumamoto Gakuen University, Japan. He has published in the areas of CALL, CMC, and second language education.

Michael Harrington is Senior Lecturer in Second Language Acquisition and Linguistics at the University of Queensland, Brisbane, Australia. He has published in the areas of second language processing, CALL, and applied psycholinguistics.

AUTHORS' ADDRESSES

Glenn Stockwell
Associate Professor, School of International Economics
Kumamoto Gakuen University
2-5-1 Oe, Kumamoto City
Kyushu
Japan 862-8680
Phone: +81 96 364 8294
Fax: +81 96 364 0702
Email: gstock@kumagaku.ac.jp

Michael Harrington
Senior Lecturer, Linguistics
School of English, Media Studies & Art History
The University of Queensland
Brisbane
4072 Australia
Phone: +61 7 3365 6719
Fax: +61 7 3365 2799
Email: mwharr@mailbox.uq.edu.au