Examining the Lexical Profile and Speaking Skills of English L2 Learners

G. Clint Denison

英語 L2 学習者の語彙プロファイルと スピーキングスキルの調査

デニソン・G. クリント

Abstract

This paper describes a pilot study that examined the lexical profile and formulaic phrase usage of EFL university students on speaking tests administered as part of a discussion course. The course aimed to develop productive vocabulary and overall speaking skills of learners through the instruction, study, and focused practice of academic vocabulary and formulaic sequences suitable for discussions. Results suggest that the course was effective, even for learners who were already proficient speakers, with increased usage of NGSL 3 and NAWL words, resulting in a more sophisticated lexical profile overall. In addition, the explicit instruction and practice of formulaic sequences based on the suggestions of Dörnyei and Thurrell (1994) resulted in increased utilization of conversational strategies such as performing confirmation checks. Implications and future directions for research are discussed.

Keywords: discussion, vocabulary, conversation strategies, formulaic speech

（Received September 25, 2018）

抄　　　録

　本研究ではディスカッションコースの一環として実施されたスピーキングテストでの EFL 大学生の語彙プロファイルと定型的な文章の使用についてのパイロット調査を行った。このコースではディスカッションに適したアカデミックな語彙と定型的な文章の明示的な指導および練習を通じて学習者の生産的な語彙や全体的なスピーキングスキルを養成することを目的としていた。結果は英語能力がすでに高い学習者であっても NGSL 3 と NAWL の語彙の使用率が増加し、全体的に洗練された語彙プロファイルになった。加えて、Dörnyei と Thurrell（1994）の提案に基づく定型的な文章の明示的な指導と練習が確認チェックの実行などの会話ストラテジーの増加につながったと考える。

キーワード： ディスカッション、語彙、会話ストラテジー、定型的な文章

（2018 年 9 月 25 日受理）

The development of speaking skills for second language (L2) learners is no easy task. Speaking occurs on-line (Yuan & Ellis, 2003), that is, L2 speakers must conceptualize a message, select contextually appropriate words, formulate utterances, articulate, and monitor speech in real time while maintaining fluent and accurate production (Kormos, 2006). Although speakers can cope with these difficulties by reducing speech rate or increasing the number and length of pauses to provide additional time for cognitive processing, the resulting disfluencies are frequently cited as indicators of low speaking proficiency (Baker-Smemoe, Dewey, Bown, & Martinsen, 2014; Revesz, Ekiert, & Torgersen, 2016), limiting the effectiveness of such compensatory strategies. In addition, these problems are exacerbated in contexts such as discussions where learners must engage with multiple interlocutors and actively direct their attention to follow the ensuing conversation.

Despite these difficulties, when questioned which language skill they would most like to develop, many learners answer speaking. This desire is not lost on instructors, many of whom search for effective ways to help learners overcome the barriers to acquiring speaking skills, especially in complex speaking contexts. Although there are many ways that L2 speaking can be instructed, one popular approach is the use of discussion-based activities, which allow for learners to engage with language in ways that foster development. Discussion facilitates interaction between learners and creates a window of opportunity for learning to occur (Mackey, 2007). In particular, interaction pushes learners to produce output, which is crucial for development (Muranoi, 2007), and provides opportunities for focused-practice in contextually appropriate situations (DeKeyser, 2007). Thus, discussions have become ubiquitous in communicative language classrooms, although the optimal way to teach the related skills remains unclear.

In this study, I investigated two instructional components of an English as a foreign language (EFL) discussion course that were hypothesized to develop speaking and discussion skills in learners. Both components were language-focused learning (Nation, 2007) approaches to instruction. The first targeted formulaic expressions and conversational strategies that could help learners have smoother and more successful discussions. The second targeted productive vocabulary knowledge because, as Schmitt (2010) noted "learning vocabulary is an essential part of mastering a second language" (p. 4), and it was hoped that an improved and expanded lexicon would help learners to have more meaningful discussions.

Literature Review

Vocabulary is at the heart of all language learning—an essential part of L2 acquisition—and in order to maximize vocabulary learning, the basic principle is to increase engagement with lexical items (Schmitt, 2008). Although there are many ways to facilitate engagement, one

of the most effective is intentional learning through a combination of explicit instruction and productive practice (Nation, 2013). The benefits of explicit vocabulary study on developing L2 proficiency are most apparent when the most frequent and useful words are targeted (Nation, 2013). As such, frequency-based lists such as the New General Service List (NGSL; Browne, Culligan, & Phillips, 2013b) and the New Academic Word List (NAWL; Browne, Culligan, & Phillips, 2013a) have been utilized to make intentional learning more efficient. Although word cards are the traditional means of study for such lists, computer-based flashcard programs, which utilize spaced repetition to maximize learning, are becoming prevalent (Nakata, 2011). Nakata (2015) showed that gradually increasing spacing was superior to equal spacing in vocabulary learning, suggesting that computer programs that can implement spacing could be an effective instructional tool.

Although there has been debate over the role of output in L2 acquisition (e.g., Krashen, 1982; Swain, 1985), a growing body of research suggests that output is not only beneficial for learning (De Bot, 1996; Muranoi, 2007), but is necessary to develop productive skills due to the role of transfer appropriate processing in language acquisition (DeKeyser, 2007). Simply put, learners must practice speaking in the L2 if they are to become proficient L2 speakers. As mentioned above, one method for integrating this type of spoken practice into instruction is the use of discussion activities. However, there are several issues that must first be considered. In discussions, it is necessary for learners to utilize language functions, such as disagreeing politely, asking and giving opinions, and making suggestions, in addition to conversational strategies, such as checking and asking for clarification, to help them speak more naturally (Dörnyei & Thurrell, 1994). These strategies must be taught explicitly and frequently practiced if learners are to acquire them and, therefore, any discussion component of a language course should also include explicit instruction on the use of phrases and strategies while requiring learners to use them frequently (Dörnyei & Thurrell, 1994).

Dörnyei and Thurrell's (1994) suggestion for developing these strategies through the teaching of formulaic sequences, that is, sequences of words or other elements which are prefabricated and retrieved from memory as a whole at the time of use (Wray, 2002), is well-founded. Not only are formulaic sequences important for comprehension because they are a core component of language (Schmitt, 2010; Wray, 2002), but they are also used extensively by proficient L2 learners in several languages (Wray, 2002). Thus, targeting conversation phrases and strategies for instruction appears to be a reasonable course of action for improving discussion skills.

Based on the previously discussed studies, which suggest that language-focused learning of vocabulary and formulaic phrases could assist in L2 development, the current study was designed to answer two overarching research questions about the vocabulary use and speaking skills of L2 learners:

1. Does explicit study in a vocabulary program change the lexical profile of L2 learners on speaking tests? If yes, how does the lexical profile change?

2. Does focused instruction and practice on the use of conversation phrases increase their usage on speaking tests? If yes, which ones and to what degree?

Methods

Participants

The study was conducted at a small women's university in western Japan. The university prides itself on language education and producing graduates who can speak English to a high degree of proficiency. The graduates often continue to work in positions that require English, such as cabin attendants and airline ground staff. To prepare students for such positions, the university has a structured language curriculum that targets a variety of skills. Even within a general skill such as speaking, there are several required classes such as conversation, presentation, and discussion. The current study was conducted within the university's discussion course.

The course is for second-year students and is focused on techniques and skills that allow learners to have effective, meaningful discussions in English on a variety of topics of varying difficulties. Coursework progresses through four major discussion structures: sharing opinions, making suggestions, making decisions, and synthesis. Sharing opinions involved discussions centered on a statement, such as "All university students should be required to take P.E. or play a sport" or "English is the world's hardest language." Making suggestions involved topics such as "Meal ideas for students living alone" and "Possible mascots for the university". Making decisions often involved constructing lists, "What are the top 5 tips for successful job hunting?" Synthesis topics combined the previous three stages by requiring groups to think through a task and produce something concrete. For example, "Create a proposal for improving English education at this school." Several weeks are spent at each stage to encourage repetition before moving to the next.

The learners ($N = 21$) included in the study came from a variety of backgrounds and were in a single section of the course. While some learners had lived in Japan their whole lives and spoke Japanese as a first language, there were several participants from abroad who were studying both English and Japanese as foreign languages. There were also participants who lived in Japan but came from mixed-heritage families. Countries represented included Indonesia, China, Thailand, the Philippines, and Japan. However, despite their diverse linguistic backgrounds, no student in the study spoke English as a first language. Although all participants were in their second-year of university, there was a wide range of ages from 19 to mid-30s. In particular, some participants from abroad were older. Due to the timing of

the study, it was not possible to administer a background questionnaire, so details about the language learning histories of participants beyond what has been mentioned are not known.

While details of histories are not well-known, the learners were all at a moderately high level of proficiency. There are several sections of the course and the section in the current study was the top level, meaning that some of the most proficient students in the school were included. A minimum TOEIC score of 500 is required to register for the section; the participants' scores ranged from 500-900. On the Common European Framework of Reference (CEFR; Council of Europe, 2011), these students would be ranked at B2, with some at the B1 or C1 levels. Relative to typical second-year students at Japanese universities, the level of these students could be considered quite advanced.

Instruments

Vocabulary program and quizzes. As part of the course requirements, all participants were required to study in the university vocabulary program, which runs for all students every year throughout their stay at the university. The program is based on empirical findings from L2 vocabulary research and integrates vocabulary learning into the school curriculum (Cornwell, 2017; McLean, 2018). In the program, students explicitly study words from the NGSL and the NAWL in order of frequency. In other words, students in their first year study the most frequent words of the NGSL, moving through the NAWL by the end of the program.

Explicit vocabulary study is done through the use of an application called Memrise (Memrise, 2018). Memrise is a digital flash card program that utilizes spaced repetition to facilitate efficient learning as students are presented with words to study just before they are forgotten (Memrise, 2018). As students successfully answer questions, the time between subsequent meetings with the word increases until they are acquired. Vocabulary questions created by the university for Memrise assess both productive and receptive word knowledge. There are tests of passive-recognition, where learners are presented with the L2 English word and must choose the corresponding L1 Japanese word, and active-recognition, where learners are presented with the L1 Japanese word and must choose the L2 English. There are also productive questions where learners must type a response without the assistance of multiple-choice options. Contextualized example sentences and corresponding audio files are included.

NAWL words were chosen for study due to the advanced level of the students. Over the course, students studied 481 lemmas from the NAWL at a pace of 37 per week for 13 weeks (see Appendix A). A key component of the vocabulary program is that students are assessed weekly through quizzes in addition to using Memrise. Because speaking was the focus of the course, productive spoken quizzes that assessed explicit productive knowledge were used. Quizzes were administered in pairs with each member quizzing the other to produce English words

from Japanese translations (see Appendix B for an example quiz). It is also important to note that a key feature of the program its cumulative nature, that is, the test range is not limited to words from the current week. Each week, ten words are selected for assessment, five from the current week and five from previous ones. Each member in a pair received different versions of the quiz. Although versions were constructed from the same pool of words, there were no duplicates between versions, which was necessary due to the interactive spoken nature of the test.

Conversation phrase materials. Discussion materials designed by Denison (2016) were utilized in the explicit instruction and practice of conversation phrases. This included a list of formulaic phrases and multi-word units that built upon the most important conversation strategies as suggested by Dörnyei and Thurrell (1994). The list, shown in Appendix C, provides learners with a range of language to use in a variety of functions such as appealing for help, giving and asking opinions, agreeing and disagreeing, checking understanding, and asking for clarification. Students were given the list on the first day of class and added to it throughout the course. The list was used frequently in the beginning classes to scaffold participants during discussions and, as the course progressed, it was used less as participants became more proficient with the phrases.

Speaking tests. Speaking tests were administered twice during the course during weeks 6 and 15 and were used as the primary means of data collection. The tests were administered in a group format with 3-4 participants per group. Students were presented with a topic and given one minute to think and plan. During the tests, participants engaged in discussions on the assigned topic, which elicited one of the target discourse models of sharing opinions, making suggestions, or making decisions. Tests lasted six minutes for groups of three and eight minutes for groups of four.

In this particular group speaking test, modified with permission from a test format first proposed by M. Grogan (personal communication, March 29, 2017), students are given two scores. The first is a holistic individual score based on language use, active participation, and the use of conversation phrases studied in class. The second is a group score based on the lowest number of sentences spoken by a member, and every member gets that score. That is, all members of the group must pass a threshold of utterances to achieve a certain score, encouraging a balanced conversation. In other words, each member must contribute to the group in a balanced way to ensure that all members can achieve the required minimum number of utterances during the test time.

Procedure

In the current study, I examined whether the explicit instruction and study of conversation phrases and vocabulary from the NAWL would change the lexical profiles of learners or result

in increased usage of conversation phrases on speaking tests. Conversation phrase materials were distributed, and the meanings of phrases were confirmed in Week 1. This step was immediately followed by a practice discussion on the topic of Japanese foods to provide an opportunity to become acquainted with the discussion format and phrases. Students also received a brief orientation about the vocabulary program in Week 1. Students had already studied in the program during the previous year, so they were familiar with the requirements and procedures. The orientation outlined the quiz schedule with quizzes held at the start of every class in Weeks 2-14. Week 1 words were assessed in Week 2 and so on. I also ensured that participants had access to the word decks in Memrise and reminded them that the test range was cumulative.

In addition to studying vocabulary using Memrise and taking vocabulary quizzes, participants engaged in discussions on a variety of topics within target discourse models of the course (see Table 1). Participants were reminded before each discussion to use discussion phrases from the course materials as much as possible. Discussions were held in groups of four

Table 1 *Overall Course Plan and Topics*

Week	Discourse Model	Topics
1	Sharing opinions	What is the best Japanese food?
2	Sharing opinions	All university students should be required to take P.E. or play a sport. English is the world's hardest language.
3	Sharing opinions	What are the most important aspects of a job? What are the most important qualities in a partner?
4	Making suggestions	Meal ideas for students living alone. Possible mascots for the university.
5	Making suggestions	Ways to improve English ability. Ways to find job opportunities.
6	**Test 1**	*Randomly chosen from Weeks 2-5.*
7	Making decisions	Make a list of the top 5 tips for successful job hunting. Make a list of the top 5 ways to improve fitness in university students.
8	Making decisions	Make a list of the top 5 most important things to do in a disaster.
9	Synthesis	Creating a campaign to attract new students to the university.
10	Synthesis	Designing a program to help 1st-year students adjust to the university.
11	Synthesis	Creating a program for improving English education in Japan.
12	Synthesis	Planning a local/global public service project.
13	Synthesis	*Repeat synthesis topics in different groups.*
14	Synthesis	*Repeat synthesis topics in different groups.*
15	**Test 2**	*Randomly chosen from the following:* What is the most important quality in an employee? What is the most important quality in a company? Ways to make new friends in a new location. Ways to improve your daily life. Decide on a list of the top 5 restaurants in Kansai. Decide on a list of the top 5 companies in Japan.

or five with the instructor monitoring and encouraging participants to use conversation phrases throughout. After each topic was discussed, a language-focused learning (Nation, 2007) session was conducted where language problems and solutions were addressed. Participants then changed groups and repeated the topic. In principle, this basic pattern was repeated throughout the course.

Speaking tests were held on Week 6 and 15 using the group format. Groups were randomly assigned the week before the test and although participants knew which types of topics would appear, actual topics were not confirmed until the day of the test. Topics were chosen randomly just before the test for each group to limit the extent that groups who finished early could share information with groups testing later. All tests were video recorded for later analysis.

Analyses

Transcription and data preparation. The study is based on data gathered from two administrations of the group speaking test (Weeks 6 and 15). Video recordings of the speaking tests were transcribed and compiled to create corpora of speech from each test that could be analyzed and compared. Video data were transcribed using a three-stage process that utilized YouTube's auto-captioning service (YouTube, 2018) and the Project Psych Transcriber (Embleton, accessed 2018).

First, video files were uploaded to YouTube to take advantage of the auto-caption service that analyses the audio stream of a video for speech and automatically adds subtitles to videos. Although the auto-caption service was not designed to be used for transcription, it is quite accurate when the audio recording is of sufficient quality. The auto-caption process takes approximately three hours from the time the video is uploaded, and all videos can be kept private during the entire process to protect the privacy of participants. However, the captions produced by YouTube cannot be utilized immediately as they are embedded in the video on the website. Thus, the Project Psych Transcriber was used for additional processing. In this second stage, the Transcriber was used to strip captions from the YouTube video files, restore punctuation, and compile the captions in a single block of text. All videos were deleted from YouTube once caption-stripping was completed. Finally, transcripts were edited manually to add speech that was either missing or incorrectly recorded as a result of the auto-captioning. Although the combination of YouTube's auto-caption and the Transcriber was able to correctly transcribe large portions of the speaking tests, manual transcription was necessary for situations where the audio quality or overlapping speech prohibited automatic transcription. Because the focus of the current study was on vocabulary and phrase use, hesitations, false starts, fillers, and similar speech phenomena not constituting full words were not transcribed for use in the analysis.

In order to compare overall differences between tests, transcripts for different groups were compiled into a single corpus for each test. In other words, all transcripts from Test 1 were combined and all transcripts from Test 2 were combined resulting in two corpora to be compared. The Test 1 corpus contained 4,352 words while the Test 2 corpus contained 4,416 words as measured by Word (Microsoft, 2016). The total length of audio was approximately 40 minutes for each test.

Lexical analysis. To examine the lexical profile of participants, AntWordProfiler (Anthony, 2014) was used to compare the Test 1 and Test 2 corpora individually against NGSL and NAWL lemma lists (available from http://www.laurenceanthony.net/software/ antwordprofiler/). This allowed for the total number of lemma (dictionary forms), types (unique occurrences), and tokens (total occurrences) of NGSL and NAWL words that were used on the speaking tests to be calculated and compared.

Although the NAWL was referenced as a whole (963 lemmas), the NGSL was compared at each of three frequency bands, that is, the first 1,000 most frequent lemma (NGSL 1), the second 1,000 most frequent (NGSL 2), and the next 801 most frequent (NGSL 3). Due to the design of the study, the use of statistical tests such as a Chi-Square test of independence to analyze changes in the lexical profile could not be used without violating the assumption of independence of data. Instead, comparisons of percentages were utilized.

In addition to examining the lexical profile of the two tests from a general perspective as discussed above, AntWordProfiler was again utilized, this time to compare the test corpora against the specific portion of the NAWL that was included in the testing range for the vocabulary program throughout the course (see Appendix A). This helped to determine if any changes in vocabulary use could be accounted for by explicit study in the vocabulary program. First, the NAWL words used on each test were examined. These words were then compared against the class list to calculate the ratio of NAWL words studied in the course to total NAWL words used.

The lexical diversity of each test corpus was also calculated using both the type-token ratio and the vocabulary D statistic. Although type-token ratio is the most straightforward method of calculating lexical diversity, it suffers from a sensitivity to text length, with longer texts being rated as less complex because there are fewer chances for unique words to appear as length increases (Schmitt, 2010). On the other hand, the D statistic is calculated through a sophisticated process that accounts for differences in text length. First, 100 samples of 35 randomly selected words from the text are used to generate individual type-token ratios which are then averaged. The process is repeated for samples of 36-50 randomly selected words resulting in 16 type-token means. The D algorithm then compares these means to a series of theoretical curves and assigns a D value based on the best fitting curve, with typical scores (Schmitt, 2010). The D value for each test was calculated using D_Tools (Meara & Miralpeix,

2015).

Phrase analysis. To determine if conversation phrase use increased between Test 1 and Test 2, AntConc (Anthony, 2018) and Word (Microsoft, 2016), were used to search for instances of the phrases targeted for instruction and practice (see Appendix C). The corpora of both tests were examined and totals for both phrase type (e.g., agreeing, disagreeing, checking understanding) and individual phrase (e.g., Do you mean ~ ?; What's your opinion?; Could you say that again?) were tabulated for comparison. Slight variations of the same phrase were collapsed into the same total. For example, "How can I say ~ ?" and "How do I say ~ ?" were considered to be drawing on knowledge of the same formulaic speech pattern and deemed functionally equivalent for the purposes of analysis.

Results

Lexical Profile

The results of the lemma, type, and token analysis comparing NGSL and NAWL lists with the Test 1 and Test 2 corpora are shown in Table 2. Percentage differences between Tests 1 and 2 for each category are also displayed. For NGSL 1 words, usage increased between tests for all measures. Conversely, for NGSL 2 words, usage dropped for all measures between tests. For NGSL 3 words, usage increases slightly between tests for lemma and type measures but dropped slightly for tokens. For NAWL words, usage increased slightly on all measures. Finally, usage of off-list words which do not appear on any of the lists (e.g., numbers, days of the week, proper nouns) decreased for all measures.

Table 2 *Lemma, Type, and Token Analysis for Speaking Tests 1 and 2*

Frequency Band	Lemma (%)			Type (%)			Token (%)		
	Test 1	Test 2	% Diff.	Test 1	Test 2	% Diff.	Test 1	Test 2	% Diff.
NGSL 1	334 (61.06)	354 (65.19)	4.13	430 (65.75)	466 (70.61)	4.86	4039 (88.81)	4242 (92.34)	3.53
NGSL 2	78 (14.26)	64 (11.79)	−2.47	88 (13.46)	69 (10.45)	−3.01	185 (4.07)	115 (2.5)	−1.57
NGSL 3	25 (4.57)	28 (5.16)	0.59	26 (3.98)	28 (4.24)	0.26	61 (1.34)	44 (0.96)	−0.38
NAWL	12 (2.19)	18 (3.31)	1.12	12 (1.83)	18 (2.73)	0.9	20 (0.44)	29 (0.63)	0.19
Off-list	98 (17.92)	79 (14.55)	−3.37	98 (14.98)	79 (11.97)	−3.01	243 (5.34)	164 (3.57)	−1.77
Total	547	543		654	660		4548	4594	

Note. NGSL 1 contains the first 1000 most frequent lemma; NGSL 2 contains the second 1000 most frequent lemma; NGSL 3 the next most frequent 801 lemma; NAWL contains 963 academic lemmas not contained in the NGSL.

The frequencies and breakdown of NAWL words that were used on the speaking tests are shown in Table 3. With regard to the ratio of NAWL words studied in the vocabulary program to total NAWL words spoken on the test, both lemma and token measures were examined. In this case, all lemma were unique and it was not necessary to also analyze type. For lemma, on Test

1, 50% of lemma (6 of 12) were words studied during the course while for Test 2, this number increased to 55.56% (10 of 18). When examining total tokens, on Test 1, 60% of tokens (12 of 20) were studied words, while on Test 2, this number dropped to 48% (14 of 29). There were three NAWL words that appeared on both tests (*actively*, *homework*, and *multi*). Of these, only one word, *actively*, was included in the study program and was used only once on each test.

Table 3　*NAWL Words Used on Speaking Tests 1 and 2 by Frequency*

Test 1	Frequency	Test 2	Frequency
artificial*	4	homework [†]	7
aspect*	2	orientation*	5
assignment	2	similarity	2
hormone	2	actively* [†]	1
radiation*	2	bonus*	1
separately*	2	criteria*	1
actively* [†]	1	impact*	1
anti	1	indirect	1
discrimination*	1	junior*	1
homework [†]	1	lecturer*	1
multi [†]	1	leisure	1
vitamin	1	mall	1
		multi [†]	1
		non	1
		planner*	1
		randomly*	1
		realistic*	1
		subjective	1

Note. * denotes a NAWL word that was included in the vocabulary program of the course.
[†] denotes a word that was used on both tests.

Finally, for lexical diversity, the vocabulary diversity statistic *D* and type-token ratio were calculated for each test. The lexical diversity of Test 1 (*D* = 71.83) was slightly lower than that of Test 2 (*D* = 72.12) when measured by *D*. However, the type-token ratios of the two tests were nearly identical (Test 1 = 14.5%; Test 2 = 14.4%).

Conversation Phrase Use

The breakdown of 42 conversation phrases of seven types is shown in Table 4. Although overall phrase use decreased slightly from a total of 95 phrases used on Test 1 to 91 on Test 2, there was not an even pattern of change across phrases, with the usage of some phrases increasing, some decreasing, and others remaining unchanged or never used. For types, giving

an opinion (−1), agreeing (−6), and disagreeing (−7) all experienced decreased usage on Test 2. On the other hand, usage of clarification requests (+1) and confirmation checks (+9) increased. There was no change in total usage for asking for help or asking an opinion. The most extreme changes in total usage were for confirmation checks, which increased by nine, and for disagreeing, which decreased by seven. Of note is that giving an opinion was the most used phrase type on both tests with "I think (that) ~" being the most used individual phrase.

Discussion

The first research question addressed if explicit study in a vocabulary program could change the lexical profile of learners on speaking tests. The results suggest that although improvements are meager, the vocabulary program had a positive impact on the lexical profile of learners. Usage of NAWL and NGSL 3 vocabulary increased overall, suggesting a slightly more sophisticated lexical profile on the second test. There was a decrease in the use of NGSL 2 words, which is an expected result of an increase in the use of NGSL 3 and NAWL words. Curiously, the percentage of words from NGSL 1, which represent the easiest and most frequent vocabulary, also increased. One possible explanation for this is that the increase of NGSL 3 and NAWL vocabulary was accompanied by an increase in function words, which was necessary to facilitate their use. Because NGSL 1 contains a large number of function words, an increase in function words would correspond to increased NGSL 1 coverage. The relationship between NGSL 1 function words and NAWL vocabulary is a potential area of research which should be explored.

It could be argued that the increase in NAWL usage on the second test was not due to the vocabulary program, which did not exhaustively cover the NAWL during the course in the current study. It is possible that participants simply used more NAWL words that they knew from elsewhere. However, this does not appear to be the case. Not only were participants using more NAWL words in general, but they used more studied NAWL word types than non-studied on the second test. In addition, the higher percentage of studied NAWL words used on Test 2 for the lemma measure suggests that participants were using a larger proportion of studied words on the second test than they had on the first. Although studied words did account for a lower percentage of NAWL tokens on the second test, when the lemma measure is considered this means that NAWL words not in the vocabulary program were simply repeated more often. Even so, this is a positive outcome as it suggests that participants not only used words they had learned but also used more academic vocabulary overall.

For lexical diversity, it seems that vocabulary study did not result in much change between the two tests. In fact, the difference in lexical diversity between the two tests on both measures was so small that it is arguably negligible. It appears that although usage of academic

Table 4 *Conversation Phrases Used on Test 1 and Test 2 by Type*

Type	Phrase		Test 1	Test 2	Diff.
			Frequency		
Ask for help	What's the word for ~ ?		0	0	0
	What do you call ~ ?		1	0	−1
	How can/do I/you say ~ ?		2	3	1
	What should I say here?		0	0	0
		Type total	3	3	0
Give an opinion	I think (that) ~		39	43	4
	I don't think that ~		1	0	−1
	I believe (that) ~		1	0	−1
	I'm sure (that) ~		1	0	−1
	In my opinion ~		3	1	−2
	Personally, I feel that ~		1	1	0
		Type total	46	45	−1
Ask an opinion	What do you think?		4	4	0
	Do you have any ideas?		1	4	3
	What's your opinion?		1	1	0
	How/What about you?		12	9	−3
		Type total	18	18	0
Agreeing	You're right.		0	0	0
	I think ~ is right, because ~		0	0	0
	I agree (with) ~		13	8	−5
	Exactly.		0	0	0
	Of course.		0	1	1
	That's right.		2	0	−2
		Type total	15	9	−6
Disagreeing	I don't think so.		0	0	0
	I don't agree with ~		0	0	0
	You said ~ , but ~		0	0	0
	I see your point, but ~		0	0	0
	That might be true, but ~		1	0	−1
	But don't you think that ~		0	0	0
	I disagree		6	0	−6
		Type total	7	0	−7
Clarification request	Pardon?		0	0	0
	Could you say that again?		0	1	1
	Sorry, what was the last word?		0	0	0
	I'm sorry, but I couldn't understand what you said.		0	0	0
	What do you mean?		0	0	0
	What are you trying to say?		0	0	0
	What is ~ ?		0	0	0
		Type total	0	1	1
Confirmation check	Do you mean ~ ?		0	1	1
	Are you saying ~ ?		0	0	0
	Did you say ~ ?		0	0	0
	So, you mean ~		0	4	4
	Okay?		6	10	4
	Is that clear?		0	0	0
	Are you with me?		0	0	0
	Do you understand what I said?		0	0	0
		Type total	6	15	9
Overall total			95	91	−4

vocabulary did increase, it was not enough to improve overall lexical diversity even when accounting for text length. One possible explanation is that the academic vocabulary targeted for study in this course occur infrequently in these types of speaking situations and, thus, do not have a substantial impact on overall lexical diversity. Another explanation is that for the high-proficiency learners in the current study, lexical diversity was already near the realistic maximum. This is supported by Schmitt (2010) who suggested that a lexical diversity D value of 50 was average for most texts. The corpora from speaking tests in this study had D values over 70, far beyond what might be expected.

The second research question addressed if focused instruction and practice on the use of conversation phrases could increase their usage on speaking tests. Here, however, the results are less positive as it appears that not much of a change occurred between tests, despite the focused instruction participants received on the phrases. In fact, the total number of phrases used slightly decreased on the second test, so it appears that for these learners focused instruction did not result in increased usage. One explanation relates to proficiency levels of participants. It is possible that learners were already sufficiently proficient with other phrases and did not need the additional ones targeted in the study to satisfactorily complete the discussions. It is also possible that participants avoided the use of the target phrases through the use of speech that was easier for them to control. For example, although it was not targeted for instruction, there were 86 instances of the word "yeah" on Test 1, used for both expressing agreement or performing a confirmation check depending on the context. Similarly, "yeah" was used 76 times on Test 2. It is possible that the versatility and easy-to-master nature of the word resulted in reduced usage of more complex phrases. On these timed speaking tests which required a minimum number of utterances, saving time and increasing the number of responses by using simpler phrases might have been a testing strategy that was implemented.

Although overall usage decreased on Test 2, there were still several interesting changes in the use of different phrase types. For example, agreeing and disagreeing saw a large decrease in use, suggesting that learners did not debate their opinions to the extent that occurred on the first test. This is interesting considering that the use of phrases to express an opinion still remained quite high. In fact, "I think (that) ~" was the most used phrase on both tests. In other words, although participants were expressing their opinions frequently, it appears that there was a decreased desire or necessity to express agreement or disagreement. Further analysis will be necessary to determine if this was indeed the case, however, anecdotal evidence from observations during instruction suggest that as the course progressed, and participants became better acquainted with each other, the required time to reach a consensus of opinion decreased.

The only phrase type that showed noteworthy improvement was confirmation checks. This is perhaps the strongest evidence in the current study in support of conversation phrase

instruction because it suggests that confirmation checks, which are quite difficult for learners to master, can be improved through instruction. One possible explanation is that making decisions, a discussion style that was both practiced in class and used on Test 2, necessitated that information be confirmed before the discussion could proceed. This would have facilitated the use of more confirmation checks, and instruction allowed learners to utilize those phrases effectively. However, more focused research will be necessary to determine if this theory sufficiently explains the increase.

Conclusion

There are several limitations of findings that must be addressed. First, due to the timing of the study, it was not possible to conduct pre- and post-tests of the NAWL words that were studied in the vocabulary program during the course. Although mean quiz scores were normally high, and it can probably be assumed that these particular learners were studying the words diligently, it is not possible to know how vocabulary knowledge changed between Test 1 and Test 2 with the current data. Second, for the current study it was not possible to separate out the corpora and conduct analyses at an individual level. This would have provided a more nuanced perspective and allowed for the use of inferential statistics. However, it was deemed unfeasible to prepare the data in the available timeframe for additional analyses. Although the methodology used in the current study did allow for examinations of overall change to be conducted, an individual level analysis of vocabulary and phrase use would be more powerful and is a future direction to pursue.

The findings must also be contextualized considering the overall proficiency level of the participant group. Because the participants were already at a moderately high level on average, they probably knew many of the points that were targeted for instruction, leaving little room to identify vocabulary development. The fact that several NAWL words which were not targeted for instruction appeared on the first test is evidence that some participants at least had partial knowledge of NAWL words prior to studying in the course. A similar situation can be assumed for the development of speaking skills. Although the participants have not stabilized and could continue to develop aspects of their English proficiency, it might require a longer period of time to observe development than was available in the context of this study.

Because the participants had already mastered many of the most frequent words of English, (e.g., NGSL 1, NGSL 2) in previous courses, the NAWL was targeted for instruction. Although this is a logical and reasonable progression from a pedagogical perspective, evaluating whether learners have acquired productive knowledge of NAWL words appears difficult due to limited opportunities to actually use these words in speaking contexts. Even though spoken usage of academic vocabulary did increase, the relatively small percentages for

NAWL word use overall does seem to suggest that it is difficult to incorporate these words into speech due to their difficulty or rarity. One possibility that must be considered is that even if learners had mastered the words, there was simply not a chance to use them in an appropriate way on the test. Thus, groups of learners with a broader proficiency range should be included in future studies in order to determine how development occurs with more frequent words such as NGSL 2 and NGSL 3.

However, despite the limitations of the study, it appears that the instructional approach used was at least partially successful. Learners were able to increase their usage of NGSL 3 and NAWL words, resulting in a more sophisticated lexical profile. In addition, the explicit instruction of formulaic speech as conversation phrases does appear to have resulted in increased use for some strategies such as confirmation checks. However, in future research, different proficiencies of learners should be included and different combinations of instructional methods for vocabulary and phrases should be compared. These types of comparisons will contribute to finding the best combination of instructional approaches for developing the skills necessary to have successful L2 discussions for learners.

References

Anthony, L. (2018). AntConc (Version 3.5.6) [Computer software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Anthony, L. (2014). AntWordProfiler (Version 1.4.0) [Computer software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*(4), 707-728. https://doi.org/10.1111/flan.12110

Browne, C., Culligan, B. & Phillips, J. (2013a). The New Academic Word List. Retrieved from http://www.newgeneralservicelist.org.

Browne, C., Culligan, B. & Phillips, J. (2013b). The New General Service List. Retrieved from http://www.newgeneralservicelist.org.Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*(4), 707-728. https://doi.org/10.1111/flan.12110

Chandler, J. (2009). Response to Truscott. *Journal of Second Language Writing*, *18*(1), 57-58. https://doi.org/10.1016/j.jslw.2008.09.002

De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, *46*(3), 529-555. https://doi.org/10.1111/j.1467-1770.1996.tb01246.x

DeKeyser, R. M. (2007). Introduction: Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). Cambridge, UK: Cambridge University Press.

Denison, G. C. (2016). Course materials for Communication English III. *Temple University Japan Studies in Applied Linguistics*, *104*, 165-187.

Dörnyei, Z., & Thurrell, S. (1994). Teaching conversational skills intensively: Course content and rationale. *ELT Journal*, *48*(1), 40-49.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon Press.

Mackey, A. (2007). Interaction as practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 85-110). Cambridge, UK: Cambridge University Press.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 51-84). Cambridge, UK: Cambridge University Press.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17-38. https://doi.org/10.1080/09588221.2010.520675

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, *37*(4), 677-711. https://doi.org/10.1017/S0272263114000825

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd Ed.). Cambridge, UK: Cambridge University Press.

Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*(1), 2-13. https://doi.org/10.2167/illt039.0

Revesz, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, *37*(6), 828-848. https://doi.org/10.1093/applin/amu069

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329-363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York, NY: Palgrave Macmillan.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1-27. https://doi.org/10.1093/applin/24.1.1

Appendix A:

Vocabulary Study List for the Course by Week

Week 1

afterward
amongst
artistic
backward
bodily
carrier
collective
computation
continuity
definite
freely
generalization
generalize
goodness
historically
importantly
interestingly
locally
machinery
marker
meaningful
namely
nationalism
neat
nicely
noisy
outer
painful
pardon
parental
partial
partially

photographic
planner
problematic
readily
realism

Week 2

realistic
rewrite
ruler
sexuality
simplify
sometime
specialty
subset
supposedly
systematic
terribly
thickness
underneath
unemployed
whichever
whoever
accent
actively
adaptation
adaptive
apple
authority
availability
bang ; loud bang
bat ; a bat
blank

bleed
bound
broadly
bucket
calculation
calculator
characterization
cheat
cheers
clever
clip

Week 3

clue
commentary
commonly
comparable
complication
conditional
connector
conscious
consciousness
container
correction
correctly
cure
detection
developmental
directive
disturbance
economically
economist
enormously

equality

flip

ghost

hedge

identification

incredible

incredibly

indicator

individually

industrialization

industrialize

inequality

influential

instability

intensity

intensive

interrupt

Week 4

interviewer

junior

kilometer

lab

likelihood

lump

mathematical

memorize

minus

monkey

nasty

nest

observer

occurrence

periodic

physically

plug

politically

positively

presume

processor

productive

productivity

progression

progressive

projection

pronounce

punch

punish

punishment

purely

quotation

ray

reactive

reactor

recipe

reliability

Week 5

replacement

resistant

ridiculous

rope

rub (_____ on / out)

selective

separately

separation

similarity

slavery

snake

socially

specification

spray

stabilize

standardize

sword

tech

tempt

tense

traditionally

tricky

unstable

variability

variance

variant

wisdom

absorb

absorption

accelerate

acceleration

accumulate

accumulation

accuracy

accurately

acid

acidic

Week 6

admission

adolescent

affirm

agriculture

alien

alliance

allocate

allocation

approximate

approximation

archaeology

architect

aspect

assembly

assert

assignment

athletic

atom

atomic

audit

bacteria

bacterial

bargain

beam

behavioral

candidate

cattle

circulate

circulation

civilization

clarify

client

clinic

colonial

colony

communicative

communist

Week 7

compensate

competent

composer

conceive

conception

conceptual

conduction

conference

confine

consent

conservation

conserve

constitution

constrain

consultation

consumption

continent

contradict

contradiction

contradictory

controversy

coordinate

coordination

correlate (with)

correlation (between)

correspondence

corruption

criteria

critically

crystal

curriculum

cyclic

damp

deliberately

demonstrator

dense

depict

Week 8

derivative

dictate

dimensional

disability

discharge

discourse

discrimination

distribution

diverse

dominant

domination

dose

drain

drift

effectiveness

elaborate

elevate

elevation

elimination

elite

emergence

emission

emit

enforcement

essentially

estimation

evident

evolutionary

execute

execution

explicit

explicitly

exploit

fabric

facilitate

faculty

fertility

Week 9

fiber

flesh

flexibility

formally

formulation

found (on)

fundamentally

genetically

genetics

globalization

goods

grasp

gravity

gross (about / nearly);

gross (domestic product = GDP)

hip

ideology

immune

impact (on / upon)

independently

inevitably

infect

infectious

initiate

initiation

inject

injection

insert

instinct

integration

interact (with)

interfere

intervene

invasion

irrelevant

justification

lecturer

legend

Week 10

legitimate

liable

logical

magnetic

manipulate

manipulation

manual

marginal

mechanic

mechanical

media; (social) media

merge

methodology;

(research) methodology

migrate

migration

missile

(social) mobility

modification

molecular

molecule

morality

mortality

motive

myth

naked

neutral

objection

obtain

occupation

oral

organ

orientation

peasant

philosopher

philosophical

powder

practitioner

Week 11

precede

prediction

probe

profound

prominent

psychiatric

psychologist

psychology

publish

puzzle; solve a puzzle

quantitative;

quantitative (research)

radiation

randomize

randomly

rational

rationality

reconstruct

regime

reinforce

rejection

render; (services) rendered

reproduce

reproduction

republic

resemble

revolutionary

rhythm

ritual

scatter

sensible

sensitivity

shortly

simultaneously

sin

sophisticate

sponsorship

statistical

Week 12

statistically

statistics

strategic

strictly

substitution

subtle (difference)

sufficiently

(to commit) suicide, (a) suicide

superior

sustainable

swell

symbolic

technically

theorist

thereby

ton

transaction

transformation

translation

transmission

transmit

treaty

tremendous

tribe

ultimate

unity

utility

vague

valid

validity

virtue

weave

activate

acute

adjacent

adverse

aesthetic

Week 13

aluminum

ancestor

anthropology

array

arrow

articulate

artificial

auction

audio

autonomy

barrel

basin

biologist

biology

bizarre

bonus

bubble

bulk

bullet

bundle

calcium

campus

capitalism

capitalist

censor

chemistry

chronic

chunk

cinema

classification

classify

clay

click

clone

closure

cognitive

coherent

Appendix B:

Example of a Productive Vocabulary Test

How do you say _____ in English, please?

The first letter is _____.

Do you know a different word which means _____?

Question   Answer

1) 逆行の；後ろへ   backward

2) 集団的な；集合的な   collective

3) 明確な；限定的な   definite

4) 一般的に話す；～を一般化する   generalize

5) 重要なことには   importantly

6) 機械   machinery

7) 意味のある   meaningful

8) すなわち；つまり   namely

9) 国家主義   nationalism

10) きちんとした   neat

_____ out of 10

Appendix C:

Discussion Phrases

These are phrases that you can use to have good discussions. Try to use different phrases to make the discussion more interesting.

Ask for help
- What's the word for ~ ?
- What do you call ~ ?
- How can I say ~ ?
- What should I say here?
-
-

Give an opinion
When you give an opinion you should give a reason.
- I think that ~ , because ~
- I don't think that ~ , because ~
- I believe that ~
- I'm sure that ~
- In my opinion ~
- Personally, I feel that ~
-
-

Ask an opinion
- What do you think?
- Do you have any ideas?
- What's your opinion?
- How about you?
-
-

Agree
- You're right.
- I think ~ is right, because ~
- I agree with ~ , because ~
- Exactly.
- Of course.
- That's right.
-
-

Disagree
If you disagree you should explain why.
- I don't think so.
- I don't agree with ~
- You said ~ , but ~
- I see your point, but ~
- That might be true, but ~
- But don't you think that ~
-
-

I don't understand
- Pardon?
- Could you say that again?
- Sorry, what was the last word?
- I'm sorry, but I couldn't understand what you said.
- What do you mean?
- What are you trying to say?
- What is ~ ?
-
-

Check understanding
- Do you mean ~ ?
- Are you saying ~ ?
- Did you say ~ ?
- So, you mean ~ , right?
- Okay?
- Is that clear?
- Are you with me?
- Do you understand what I said?
-
-