The Contribution of High-frequency Multi-word Sequences to Speech Rate and Listening Perception Among EFL Learners

Michael McGuire[a] and Jenifer Larson-Hall[b]

[a]Doshisha University; [b]University of Kitakyushu

Abstract

This experiment tested gains in spoken fluency and ability to complete a dictation listening task accurately among 33 Japanese L1 English language users. Both a control group ($N = 17$) and an experimental group ($N = 16$) studied Anki vocabulary cards each week for 10 weeks and described three picture stories that contained the vocabulary words every week. Both groups studied 10 common bigrams (such as *take advantage*) each week while the experimental group additionally studied sets of 10 reduced trigrams (*how do you*) and did narrow listening homework each week. The results for spoken storytelling fluency found a large advantage for the experimental group while fluency for the free speaking task showed a medium advantage for the experimental group that was not statistically significant. For the listening dictation task, both groups reduced their errors from pretest to posttest but neither group was statistically different from each other.

1 Introduction

Rather than building every sentence one word at a time, fluent speakers of a language assemble much of their speech from prefabricated, or formulaic, multi-word chunks of language (Wood, 2010). There are many terms used to describe this type of language and we have chosen the term *multi-word sequences* (MWS) for this article. Through repetition, the spoken production of high-frequency MWSs becomes automated, which allows for fluent speech with less cognitive load (Bybee, 2002a), and the proceduralization of these strings "could allow expression to occur fluently under the constraints of time which real-life speech entails" (Wood, 2010, p. 2).

It follows that language learners might benefit from learning and using more MWSs as well. We examine two types of frequent MWSs in our experiment: (1) Bigrams such as *high probability* or *take advantage*; and (2) The most frequent trigrams in spoken English, phrases such as *and then I* or *how do you*, which we call "reduced MWSs." These phrases do not have much independent lexical content, and we label them as "reduced" since considerable phonological changes occur when they are uttered (Bybee, 2002b; Bybee & Hopper, 2001) (consider, for example, the reduction of *how do you* in fluent speech to [hawd͡ ʒə]). We are targeting these two types of MWSs because we believe their sheer frequency makes them valuable.

Previous research on the bigrams used here (Nguyen & Webb, 2017) found that English majors at a Vietnamese university did not have strong knowledge of the collocations in spite of the fact that all of them were composed from the 3,000 most frequent words of English (mean score 45% on average). The study did not try to teach participants the collocations; it simply tested their knowledge of them.

Although previous studies in the SLA literature have demonstrated that a focus on identifying and heightening awareness of the importance of formulaic sequences can lead to improved fluency measures in English users (McGuire & Larson-Hall, 2017; Wood, 2009), we only know of one study which targeted specific MWSs and then measured fluency on a speaking task. Thomson (2017) used a "fluency workshop" approach similar to Wood (2009) over 6 weeks and targeted 4-gram combinations like *what would you like* and *I think I will*. For productive knowledge of the 30 targeted MWSs in this study, Thomson found a marked difference for the experimental group over the control group (Cohen's d = 1.41), which was perhaps to be expected given the control group had not seen the MWSs that the experimental group had seen in their materials. However, for productive fluency, Thomson used a restaurant role play of ordering and in speech rate the groups did not differ statistically, and effect size was negligible (eta-squared = 0.02). Examining the number of MWSs used the groups also did not differ but there was a small effect size (Cohen's d = 0.41). Although this study was very carefully planned and carried out, the use of a very basic type of ritualized scenario for the pretest/posttest may have muted the effects of the experiment. The current study aims to improve on Thomson (2017) by using a more challenging task for testing (storytelling), adding a spontaneous task, and using a longer time frame for acquisition.

Considering our "reduced MWSs" category, previous studies which have looked at such collections of frequently occurring words have found that phonological reduction in these sequences is difficult for language learners. Henrichsen (1984) had 65 L2 students listen to recordings of English speech both with the presence and absence of connected speech processes (CSPs) and transcribe what they heard. Compared to native speakers, L2 listeners saw a large drop in accuracy when transcribing recordings that featured CSPs regardless of learners' proficiency level.

To the best of our knowledge no major studies have examined the effect of reduced MWSs on spoken fluency. The present study combines previous work on teaching formulaic language along with a frequency-based corpus approach. We asked students to use flashcards to study 88 of the highest-frequency MWSs (which contain CSPs) and 100 bigrams involving high-frequency words. Given the basic finding in SLA that output encourages the acquisition of language (Swain, 1993), we designed a treatment that incorporated the production of the same MWSs weekly by describing picture stories. We hoped that this would encourage the participants to use the phrases in guided production and also produce MWSs spontaneously in unplanned speech, leading to greater overall fluency in their speech rate. This focus led us to the two main research questions:

1. Do students who learn and practice using common bigrams and high frequency reduced trigrams make improvements in overall speech rate when asked to describe picture stories incorporating the MWSs and when doing free speech?
2. Does this also improve listening perception of casual fluent speech that uses these formulaic sequences?

2 Methods

2.1 Participants

The study took place over 10 weeks in the spring semester with Japanese university students majoring in English at two universities. Because of attrition or failing to complete study assignments, a number of students either dropped out of the study or were dropped from the study. Two more students who completed all requirements were dropped from the study because they did not make any errors on the MWS portion of the dictation test, leaving no room for improvement. The number of participants in each group ended up being Control = 17 and Experimental = 16.

All study participants studied 10 collocations every week using the *Anki* flashcard program and described or listened to their classmates describing three picture stories every week that incorporated the collocations. The experimental group received practice with the reduced MWSs by (1) receiving 20 Anki flashcards per week with the 10 reduced MWSs on them; (2) watching a 5-minute introduction of the week's MWSs and their CSPs; and (3) doing at least 15 minutes of narrow listening homework.

2.2 Selection of target language

- Collocations

The 100 target collocations used in this study are found in Nguyen and Webb (2017) and contain randomly selected collocations of two-word adjective-noun and verb-noun collocations using only words from the first 3,000 most frequent words in English.

- Reduced MWSs

The list of reduced MWSs used in this study (McGuire, 2020) was created for classroom use over the course of a semester in 10 weekly lessons after examining and categorizing the most frequent 3- and 4-gram MWSs in the spoken section of the Open American National Corpus (Fillmore et al., 1998). The MWSs that appear on the list are predominantly made up of function words and are not as psycholinguistically salient as fixed expressions or idioms, so students are not expected to memorize them like vocabulary. Rather, the list is intended to give students extensive practice with the most common MWS variations that feature CSPs.

2.3 Procedures

All participants loaded their list of MWSs every week into their Anki flash-card system (see Appendix A). Both groups had 10 collocations while the experimental group also had 10 reduced MWSs. All students were asked to study the cards at least three times a week. For the collocations, the Anki cards showed an English collocation, sentence, and audio on one side with a Japanese translation of the phrase and sentence on the other side. For the reduced MWSs, the cards showed the text on one side and an audio recording of the reduced MWS on the back. The participants were encouraged to repeat and shadow the reduced MWSs aloud when they turned over the card to the audio version in order to become familiar with the reduced pronunciation.

For the reduced MWS activities, the experimental group completed "Narrow listening" (Krashen, 1996) homework by choosing 15 minutes of YouTube videos of a person engaged in conversation and counted how many times they heard the weekly "reduced MWSs" in their listening. This activity allows students to become familiar with content and their selected person's voice but left them free to revisit recordings in order to notice different linguistic features (Chang, 2017)

Both groups saw three pictures per week that collectively incorporated all 10 weekly collocations. All stories contained a written description with the collocations included. The participants were given a few minutes to read the description previous to then looking at a version of the picture story without the written description. The students were asked to describe the picture in their own words but to not forget the collocations. Below is the story description that students read, while Figure 1 shows the page without the description but with the prompt words on it that participants used while telling the story.

A man is starting out building a new business. It is a fast-food restaurant called "Buddy's". He wants to do business in a part of town that used to have a lot of serious crime. However, in the past decade this area has been getting nicer. He reduces his risk of crime by installing an alarm system.

The participants took five tests before and after the experiment. Since we were concerned with fluency and listening perception, the two main tests we report here are two speaking tasks and a listening dictation. The pretest/posttest storytelling speaking task involved three multi-panel picture stories different from any used during the experiment. Similar to the practice conditions each week, the participants had 2 minutes to read a description of the story and then had to tell the story in their own words. Different from the weekly practice stories, there were no collocations listed on the story-less pictures, but there were some prompt words helpful for describing the story, possibly including one word of a collocation (for example, in one story which described a bank robbery, the collocation *serious crime* was not listed but *crime* was). A second speaking task asked the participants to freely describe what they would do in their ideal day. The participants had 1 minute to speak spontaneously. Audio recordings of these tasks were analyzed using the *Syllable Nuclei v2* script (de Jong & Wempe, 2009) using PRAAT software (Boersma & Weenink, 2021) to determine speech rate (number of syllables divided by duration) as a quantitative measure of fluency.
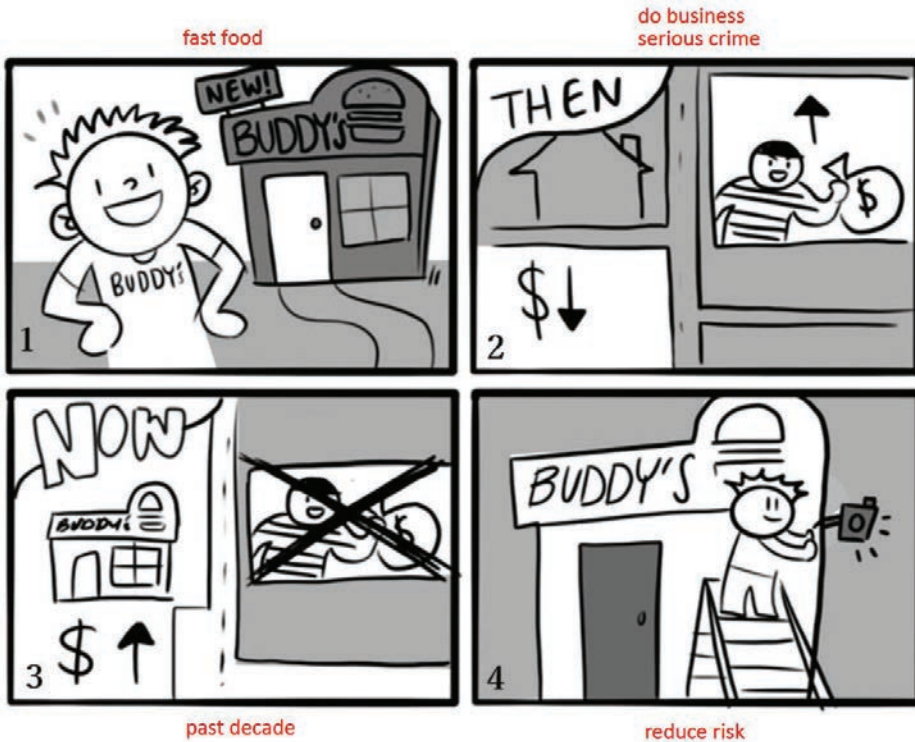
Figure 1. Example Picture for the Experiment.

In preparing these speech samples for analysis, the stories and free speaking were cut out of the recording stream when the participant began speaking and ended after they finished speaking. In some cases, there were interruptions in the storytelling and these interruptions were removed so that the monologue would sound natural.

The listening perception task was a dictation task consisting of a naturalistic conversation that participants wrote down word for word. It contained 10 collocations and 27 reduced MWSs. The participants heard the taped conversation twice at a normal native-speaker speed.

3 Results

For statistical reporting unless otherwise noted all of the data was found to be normally distributed using the Shapiro-Wilk normality test where the alternative hypothesis is that the distribution is not normal ($p \geq 0.05$; normality accepted) (see Larson-Hall, 2016, pp. 109–110). We declare an alpha level of 0.05 but focus discussion on the corresponding 95% confidence intervals (CIs) and effect sizes.

3.1 Storytelling task

Table 1 gives statistics for the storytelling speaking task. The number for each participant is a composite average of their speech rate from all three stories.

The speech rate measure divides the number of syllables by the duration (in seconds) of the speaking activity, so it includes the speed of speech as well as pauses and breakdowns; this measure is most similar to what one would think of as general fluency (Tavakoli et al., 2020). Both groups started with nearly identical speech rates, but the control group's speech rate declined slightly from pretest to posttest while the experimental group's speech rate increased. The statistics in the "Gain Score" column compare one group's pre-test and post-test scores using the effect size for a paired-samples *t*-test. The statistics in the "Comparison" column compare the gain scores between the control and experimental group using an independent samples *t*-test.

Figure 2 shows a parallel coordinate plot for the speech rate of the control group and the experimental group separately. This plot charts the path from pretest to posttest with a separate line for each participant while the larger black line shows the average for the group.

For speech rate, the control group's data was not normally distributed (Shapiro-Wilks *p* = 0.046) so Welch's *t*-test was used. The control group's decline was noticeable and the 95% confidence interval [0.10, 0.41] shows that the decline in the

Table 1. Statistics for the Storytelling Task

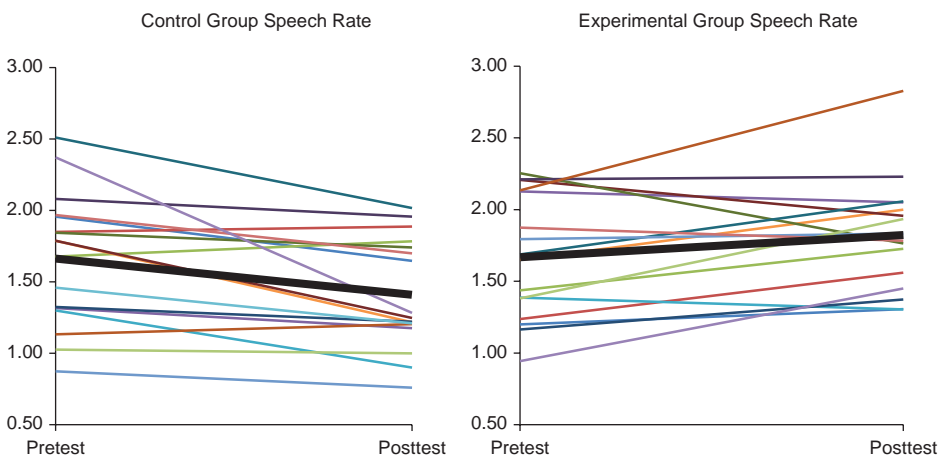|	Pretest	Posttest	Gain score	Comparison
Control (*N* = 17)	1.66 (0.46)	1.41 (0.39)	0.25 (0.3) 95% CI: [0.10, 0.41] *d* = 0.85 95%CI: [0.28, 1.40]	Mean difference: 0.41 95% CI: [0.19, 0.63] *d* = 1.3 95% CI: [0.48, 2.14] Welch's *t*-test: *t* = −3.80, df = 30.6, *p* = 0.0006
Experiment (*N* = 16)	1.67 (0.43)	1.82 (0.39)	0.15 (0.31) 95% CI: [0.32, −0.01] *d* = 0.49, 95%CI: [−0.04, 1.00]	

Figure 2. Figures for the Storytelling Task Speech Rate Separated by Group.

population, which was -0.25 points in this sample, might be as small as 0.10 points or as large as 0.41 points. A Cohen's *d* effect size for this decline is medium-to-large sized at *d* = 0.85 (but could be as small as 0.28). For the experimental group, the increase in speech rate is not statistical as the 95% CI passes through zero [0.32, -0.01]; the effect size is also small-to-medium (*d* = 0.49 on average but could be as little as zero). Comparing the gains of the two groups to each other using an independent samples *t*-test, the groups are statistically different with a possible difference in speech rate, with 95% confidence, from as little as 0.19 to as much as 0.63 faster. The effect size can be labelled large at *d* = 1.3, although the true effect size in the population might be as small as *d* = 0.48, which would only be a small-to-medium effect.

In summary, there were solid gains in speech rate in the experimental group, meaning that the time spent on learning collocations and reduced MWSs seemed to promote speaking fluency.

3.2 Free Speech Task

Table 2 gives statistics for the free speech task in which participants talked about what they would do on their ideal day (parallel coordinate plot in Figure 3).

Again, the control group slightly declined in their speech rate from pretest to posttest and the experimental group increased. But in both cases, the 95% CI goes through zero so neither change is statistical. However, the experimental group has a larger effect size than the control group. Statistically there is no difference between the control and experimental group, but the effect size shows that the difference between groups could be nothing or as large as *d* = 1.4. A larger number of participants could help narrow this confidence interval.

3.3 Dictation Listening Task

Table 3 shows descriptive statistics for errors on the dictation task. An error was any word that was incorrect or the absence of a word when one as needed. There were 177 words in the test, but some participants wrote in extra incorrect words, so the total number of errors is theoretically larger than 177. We examined the total number of errors (Figure 4) as well as errors due to reduced MWSs only (Figure 5).

Table 2. Statistics for the Free Speech Task.

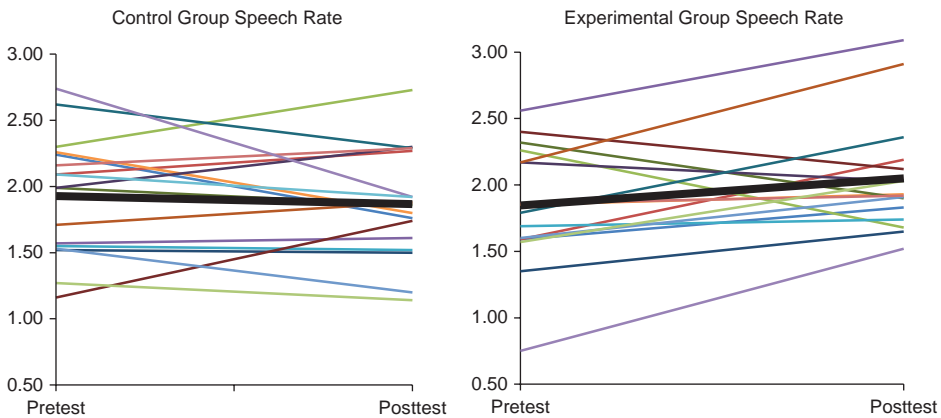|	Pretest	Posttest	Gain score	Comparison
Control (*N* = 17)	1.93 (0.45)	1.87 (0.42)	−0.06 (0.36) 95% CI: [−0.25, 0.12] *d* = −0.18 95%CI: [−0.65, 0.13]	Mean difference: 0.27 95% CI: [−0.01, 0.54] *d* = 0.70 95%CI:[−0.04, 1.42]
Experiment (*N* = 16)	1.84 (0.46)	2.05 (0.43)	0.21 (0.41) 95% CI: [−0.01, 0.42] *d* = 0.50 95%CI:[−0.03, 1.02] (data skewed)	*t* = −2.00, df = 29.8 *p* = 0.054

Figure 3. Figures for the Free Speech Task Speech Rate Separated by Group.

Table 3. Descriptive Statistics for the Dictation Task

	Pretest		Posttest		Gain score	
	Total number of errors	Errors in reduced MWSs only	Total number of errors	Errors in reduced MWSs only	Total number of errors	Errors in reduced MWSs only
Ctrl	71.88	25.0	56.41	17.59	−15.47	−7.41
(N = 17)	(16.47)	(7.65)	(17.93)	(7.14)	(9.3)	(5.14)
Exp	65.62	26.62	47.12	15.12	−18.5	−11.5
(N = 16)	(26.42)	(13.07)	(25.59)	(10.53)	(11.97)	(7.45)

The descriptive statistics show that both groups improved by reducing their number of errors from the pretest to the posttest. Paired t-tests for each type of reduction in errors within each group (tests 1–4 in Table 4) were statistically significant and effect sizes were similar, as can be seen in Table 4. The experimental group declined in errors in reduced MWSs by a larger numerical amount than the control group, but the difference was not large enough to find statistical differences between the two groups when it came to gains in reducing errors in either the entire test or specifically just in MWSs (see independent sample t-test results 5–6 in Table 4). The average effect size for differences in the reduction of MWS errors was medium, and larger than the small average effect size for differences in the reduction of total errors.

It appears that both the control and experimental group were able to do much better on the dictation task by the end of the semester, thus reducing their errors both overall and specifically on MWSs by a large amount (looking at the average Cohen's d for pretest versus posttest). This area does not show that the extra work done by the experimental group provided any important results.
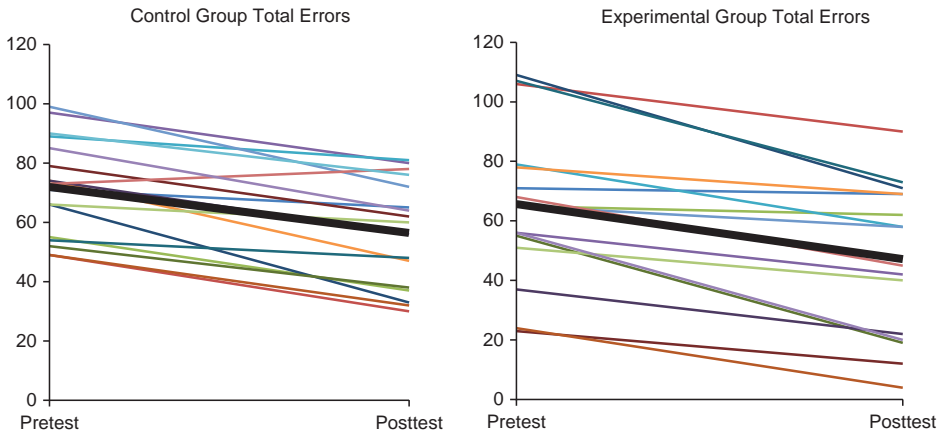
Figure 4. Figures for the Listening Dictation Task (All Errors) Separated by Group.
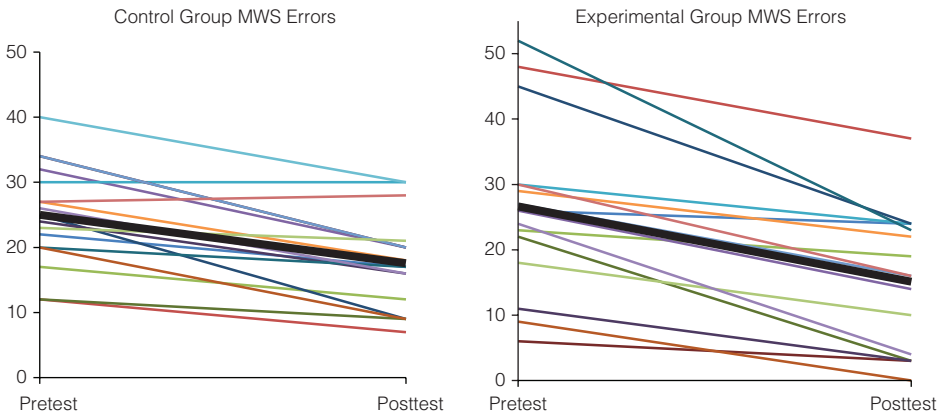

Figure 5. Figures for the Listening Dictation Task (Only MWS Errors) Separated by Group.

Table 4. Inferential Tests for the Dictation Task

Pretest versus posttest	*t*	Df	*p*	95% CI	Cohen's *d*
(1) Total errors Control	6.86	16	*p* < 0.0001	[10.69, 20.25]	1.66 [0.91, 2.40]
(2) Total errors Experimental	6.18	15	*p* < 0.0001	[12.12, 24.88]	1.55 [0.80, 2.27]
(3) MWS errors Control	5.95	16	*p* < 0.0001	[4.77, 10.05]	1.44 [0.75, 2.12]
(4) MWS errors Experimental	6.18	15	*p* < 0.0001	[7.53, 15.47]	1.54 [0.80, 2.27]
Control versus Experimental					
(5) Total errors	0.82	31	*p* = 0.42	[−4.55, 10.61]	0.28 [−0.41, 0.97]
(6) MWS errors	1.85	31	*p* = 0.08	[−0.43, 8.61]	0.64 [−0.08, 1.35]

4 Discussion

We explored the question of whether targeted attention to reduced MWSs would improve our Japanese students' speaking fluency so that their speech rate would increase when telling picture stories and speaking spontaneously. All participants memorized bigrams and practiced using them while telling stories in class, so it was expected that all participants might improve in their speaking fluency, but with special attention paid to very frequent 3-gram we thought the experimental group might improve even more. For speaking rate, we found that the control group declined in speech rate while the experimental group increased. Although the difference between groups was only statistical for the story task and not the free speaking task, effect sizes seemed to indicate that the experimental group did benefit in their speech rate from the additional work with reduced MWSs. We also expected the approximately 15 minutes of homework per week for the experimental class would have considerably improved the experimental group's performance on a listening dictation test, especially when looking at errors due to reduced MWSs only. However, we did not find increased improvement for the experimental group. We plan to continue gathering more data next year using the same experiment in order to gain more power to examine the question of whether speaking fluency and listening ability can be enhanced through a study of reduced MWSs.

References

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.1.44) [Computer software]. http://www.praat.org/

Bybee, J. (2002a). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109–134). John Benjamins Publishing Company.

Bybee, J. (2002b). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*(2), 215–221. https://doi.org/10.1017/S0272263102002061

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Benjamins.

Chang, A. C.-S. (2017). Narrow listening: A subset of extensive listening. *Journal of Teaching & Learning English in Multicultural Contexts, 1*(1), 9. https://doi.org/10.1002/9781118784235.eelt0564

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Fillmore, C. J., Ide, N., & Macleod, C. (1998, May 28–30). An American national corpus: A proposal. *Proceedings of the First Annual Conference on Language Resources and Evaluation* (pp. 965–969). Granada, Spain.

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning, 34*(3), 103–123. https://doi.org/10.1111/j.1467-1770.1984.tb00343.x

Krashen, S. D. (1996). The case for narrow listening. *System, 24*(1), 97–100. https://doi.org/10.1016/0346-251X(95)00054-N

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge.

McGuire, M. (2020). *Developing a categorized list of high-frequency spoken multi-word sequences for improving English fluency* [Unpublished].

McGuire, M., & Larson-Hall, J. (2017). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal, 34*(3), 1–25. https://doi.org/10.18806/tesl.v34i3.1271

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298–320. https://doi.org/10.1177/1362168816639619

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 50*(1), 158–164. https://doi.org/10.3138/cmlr.50.1.158

Tavakoli, P., Nakatsuhara, F., & Hunter, A. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal, 104*(1), 169–191. https://doi.org/10.1111/modl.12620

Thomson, H. (2017). Building speaking fluency with multiword expressions. *TESL Canada Journal*, *34*(3), 26–53. https://doi.org/10.18806/tesl.v34i3.1272

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics, 12*(1), 20. https://journals.lib.unb.ca/index.php/CJAL/article/view/19898

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications.* Bloomsbury Publishing.

Appendix

Appendix A. Targeted Formulaic Sequences Used in Experiment.

Week 1	Week 2	Week 3	Week 4	Week 5
Collocations				
high drama	low income	fast food	save energy	difficult situation
large number	improve quality	possible exception	inner city	catch fish
work hard	main road	past decade	major concern	quick glance
long tradition	special occasion	public confidence	big surprise	change direction
strong desire	central figure	serious crime	take pleasure	free speech
make money	spend time	take pride	feel pain	take place
lose hope	give birth	express concern	lose weight	give voice
enter college	real life	take comfort	take notice	watch television
take advantage	pay attention	do business	make music	make progress
whole family	do research	reduce risk	build muscle	good reason
Reduced MWSs				
a lot of	I don't know	you know I	I think that*	I think that*
one of the	I don't think	you know it	I think it	you know that*
some of the	I don't have*	you know and	and I think*	that you know
a couple of	I don't like	you know the	I think it's	I think that's*
that kind of*	I don't really	you know you	I think that's*	things like that
a lot of people	I don't want	you know they	I think they	something like that
kind of a	I don't even	you know it's	I think the	that kind of*
what kind of	I don't see	you know that*	yeah I think	to do that*
out of the	I don't remember	you know we	but I think	I know that
lot of the	I don't get	you know if	well I think	that would be

Week 6	Week 7	Week 8	Week 9	Week 10
Collocations				
become law	smell sweet	general population	important component	natural disaster
favorite subject	human nature	important feature	significant decrease	declare victory
small talk	open space	dark cloud	red cross	provide relief
natural beauty	wide range	top secret	private sector	have access
senior citizen	perfect match	gain knowledge	send word	lose faith
direct traffic	lose sight of	make noise	prime minister	take refuge
heavy rain	take note	commit murder	good fortune	gain admission
take issue	make sense	long silence	pay cash	feel sympathy
take flight	hard copy	high degree	basic rule	collect data
feel pressure	mass media	use technology	long term	give priority
Reduced MWSs				
and you know	you have to	and it was*	what do you	to do it
and I think*	I have a	when I was	do you have*	I used to
and it was*	do you have*	it was a	do you think	be able to
and I don't	we have a	it was just	do you do	to be a
and then I	I don't have*	and I was*	how do you	to go to
and things like that	you have a	I was in	do you like	to do that*
and I was*	to have a*	and that was	do you know	to have a*
and all that	have a lot	there was a	do you feel	I like to
and so I	I have to*	was kind of	or do you	I have to*
and then you	they have a	that was a	well do you	talking to you

*Reduced MWSs that are repeated in multiple weeks.