



Malaysian Journal Of ELT Research

ISSN: 1511-8002

Vol. 7 (1), 2011

A New Approach To Vocabulary Assessment: A Matrix Model For Predicting Future Vocabulary Performance

MAKOTO YOSHII

*Prefectural University of Kumamoto
Japan*

Abstract

This paper introduces a new approach to vocabulary assessment. We normally test learners' performance to show what they have learned, but what if we could test learners to make predictions of their future performance? Teachers might be able to provide early intervention or affirmation to the learners to aid their learning. First, the paper introduces a model (Meara, 1990) which makes such predictions possible and explains how it works. Second, the paper reports on two studies conducted to validate the model and the accuracy of predictions. One study deals with retaining vocabulary knowledge and the other with learning new vocabulary. The first study was conducted with 43 EFL learners in Japan. They had already learned 200 new words. They were tested to see how well they could retain the words. The learners' performance was compared with the predictions. The second study was conducted with 27 EFL learners in Japan. They learned 200 new words over 2 months, 20 words per week for 10 weeks. Predictions were made on the process of learning the new words. The results of the studies indicated that

the model made reasonably accurate predictions of learners' future performance both for retention and for learning.

KEYWORDS: Vocabulary, Long-term change, Prediction, Matrix model, Validation.

Introduction

We know vocabulary is an essential component for teaching and learning a language. L2 vocabulary knowledge has an impact on L2 reading (Coady, Magoto, Hubbard, Graney, & Mokhtari, 1993, Haynes, 1993; James, 1996; Laufer, 1997), L2 writing (Laufer & Nation, 1995), L2 listening (Stæhr, 2009), and L2 speaking (Joe, 1995; Joe, Nation, & Newton, 1996; Newton, 1995). We also know about some aspects of L2 learners' vocabulary development (de Groot & Keijzer, 2000; Jones, 1995; Laufer & Paribakht, 1998; Parry, 1993; Schmitt, 1998). However we do not know whether it is possible to make meaningful predictions about vocabulary learning. This paper introduces a model which has potential to make such predictions possible and to afford us a new approach to vocabulary assessment. We normally test learners' performance to show what they have learned. Teachers give advice to learners based mainly on what has been done. This type of assessment is reactive. If we could test learners to give predictions of their future performance, teachers might be able to provide early intervention or affirmation to the learners to aid their learning. This would allow assessment to be proactive as well as reactive.

In this paper, the model will be explained first, followed by a review of previous studies of relevance to the model. Then, the paper introduces two empirical studies conducted in order to verify the validity of predictions made through the model. For each study, the purpose and the research questions will be introduced first, and the research design and the results will be reported. Finally, the implications of the results will be discussed and questions for future studies will be presented at the end.

About matrix models

This paper introduces a particular model called the matrix model. It was initially introduced in areas such as mathematics and sociology (Bradley & Meek, 1986) and it has been adopted for other academic areas. Meara (1990) saw in the model potential as a new tool for vocabulary assessment. The model predicts long-term vocabulary change through a transitional probability matrix (Meara, 1990), which will be explained later. The basic framework of the model needs to be explained first. It is a discrete model in the sense that vocabulary knowledge is classified into discrete states. Vocabulary knowledge does not have to change in a developmental, hierarchical order, but knowledge can move out of one state and move into any state at any moment. In this paper, the most basic two-state model is used to illustrate the working of the model. A two-state model has the following two states: State 0 (S0) which is the state where a learner has no knowledge of

a word, and State 1 (S1) which represents the presence of knowledge of the word. Each state can be translated as “I don’t know this word,” for S0, and “I know this word,” for S1, respectively. Word knowledge can move from S0 to S1 or vice versa at any time, and it can also remain as S0 or S1.

In order to make a prediction, this model examines two tests given at two different times and looks at the rate of change between the tests. The model is based on the assumption that the change rates calculated through data from two tests are stable and will not change. Self-rating tests are frequently used to examine changes in word knowledge. This enables learners to rate hundreds of words in a comparatively short period of time. A learner simply rates his or her knowledge of each word as either S0 (“I don’t know this word.”) or S1 (“I know this word.”). Suppose that a learner scores 70 words as S0 and 30 words as S1 at a test at Time 1 as seen in Figure 1. In the next test, Time 2, the learner scores 66 for S0 and 34 for S1. Of 70 S0 words, 90% remained at S0 in the second test. 10% of the S0 words moved up to S1. This can be described in the matrix seen in the middle of Figure 1 with the change rates: $S0=0.9$, $S1=0.1$.

When the change rates for S0 and S1 are combined, this produces the matrix table which reflects the overall change rates for both states as seen in the middle table in Figure 3.

Using this transitional probability, one can make predictions for future data. Figure 3 shows that 90% of the 70 S0 words will stay in S0, which amounts to 63 (70×0.9) words. Ten percent of the 30 S1 words, that is 3 (30×0.1) words, will move back to S0.

Therefore, the next test will have 66 (63 plus 3) S0 words. That is what the S0 section of the table for Time 2 displays. Similarly, 10% of the 70 S0 words will move up to S1, that is 7 (70×0.1) words. Within the 30 words of S1, 90% will stay as S1, that is 27 (30×0.9) words. Therefore, the model predicts that the sum of the S1 words for the next test will be 34 (7 plus 27) words. The S1 section of the table for Time 2 indicates exactly that number.

Figure 3: Example of transitional probability matrix

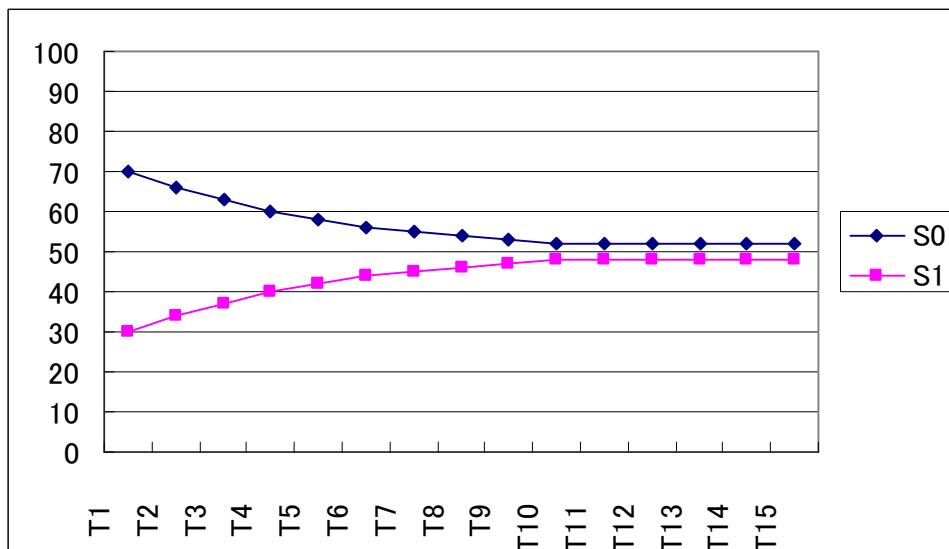
	Time1	S0	S1	Time2
S0	70	0.9	0.1	66
S1	30	0.1	0.9	34

In order to predict Time 3, the same calculation can be administered using Time 2 data. When the process is repeated, one can produce the future prediction as Table 1 and Figure 4 display. As mentioned before, this model is based on the assumption that the transitional probability stays the same throughout the period of inspection.

Table 1: Example of long-term prediction

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
S0	70	66	63	60	58	56	55	54	53	52	52	52	52	52	52
S1	30	34	37	40	42	44	45	46	47	48	48	48	48	48	48

Figure 4: Example of the matrix prediction

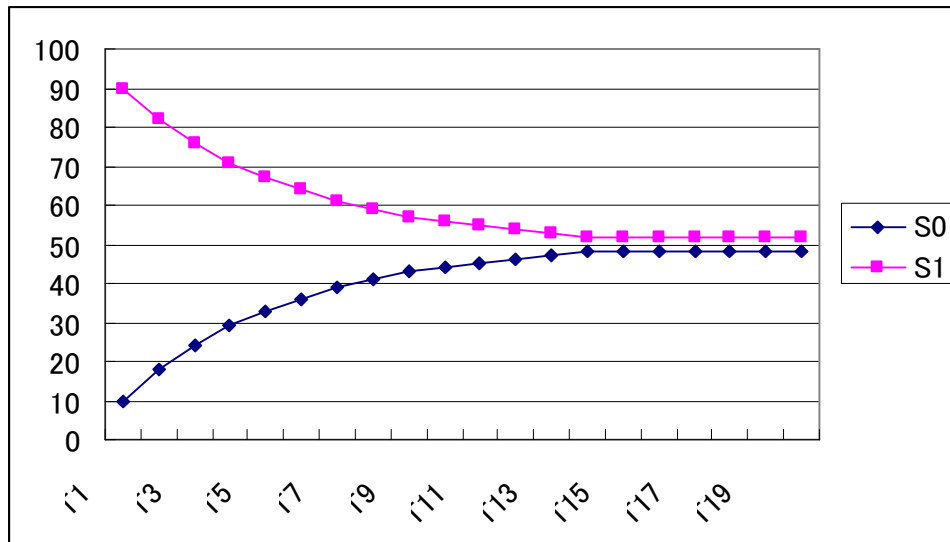


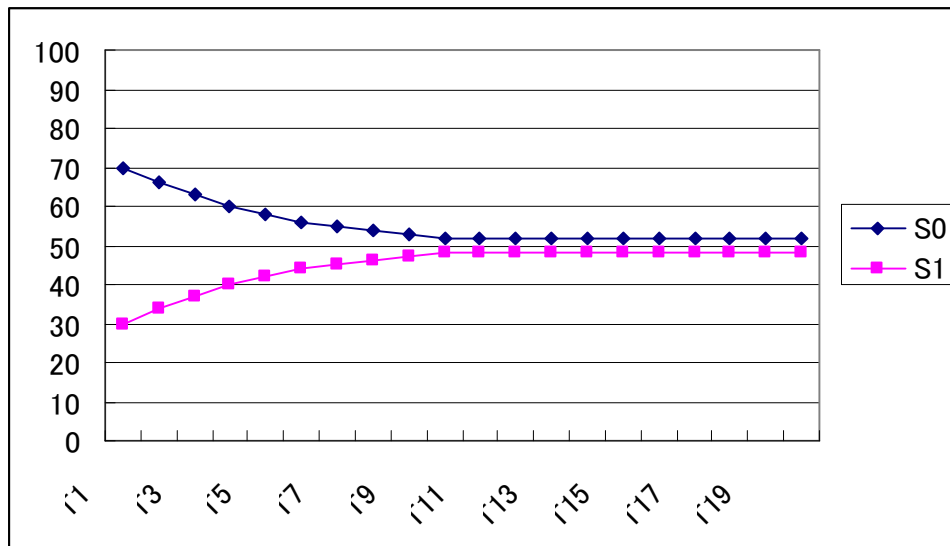
As Figure 4 displays, the studies have shown that long-term changes tend to settle into a plateau, an equilibrium state (Meara, 1990; Horst & Meara, 1999; Yoshii, 2009b). This makes us cautious about vocabulary assessment. When we give a test to measure the number of S0 or S1 for a set of target words at a time before the equilibrium, it may give us erroneous long-term estimates of known and unknown words since the scores tend to change more. When the scores reach equilibrium, it is safer to refer to the vocabulary size of the known words for the target words which will likely be stable. Therefore, it is important to know not only the scores at any given time but also the change rates and the scores at equilibrium point in order to make a long-term judgment of vocabulary learning. One caution needs to be made. The equilibrium state looks stable in terms of the overall scores; however, this does not mean there will be no changes in terms of individual word knowledge. The equilibrium simply means that the number of words coming into a particular state is equivalent to the number of words going out of the state. As a result, changes cannot be detected on the surface as long as the number of the words moving in and out of the state is balanced out.

Previous studies have also shown that the decisive factor for change is not initial conditions of vocabulary knowledge but the transitional probability matrices (Meara,

1990). For example, suppose there are two learners with two very different initial conditions with the same transitional probability; the first learner has a very high proportion of known words ($S_1=90$) and a very low proportion of unknown words ($S_0=10$) out of 100 words. The second learner has a low proportion of known words ($S_1=30$) and a high proportion of unknown words ($S_0=70$) out of the same 100 words. Even though the two learners differ very much in their initial conditions of vocabulary knowledge, the transitional probability is exactly the same as the example in Figure 4 shows. The results of the predictions of the two learners' scores turn out to be strikingly similar, as shown in Figure 5.

Figure 5: Examples of the two different initial conditions with the same transitional probability





Even though the two have very different initial conditions, since the transitional probability is the same, both settle into the same equilibrium. Therefore, the key factor determining long-term changes is not the initial conditions but the transitional probability. The initial conditions make a difference only in the timing for entering into equilibrium.

Previous studies have shown high correlations between predictions and actual scores (Horst & Meara, 1999; Meara & Sanchez, 2001; Sanchez, 2000, Yoshii, 2009a). The results tell us that the model can make reasonably accurate predictions about long-term vocabulary scores. However, the number of studies on the model is still very limited and there is a need for verification of the accuracy of the predictions. Therefore, this paper has two goals: one is to verify the results of the previous studies and see how accurately

the matrix models can make predictions for retention of newly learned vocabulary knowledge. The other goal is to explore the possibility of implementing the matrix predictions in the process of intentional learning. The paper investigates whether it is possible to make reasonable predictions when learners try to learn 20 words per week for 10 weeks with the goal of learning 200 target words.

First experimental study

The purpose of the first study was to examine how accurately the matrix models can reflect actual learners' data for the retention of newly learned words. This was to verify the results of the previous studies which showed very close reflections of actual learners' data. The study addressed the following research question: Are there any relationships between actual learners' data and the matrix predictions?

The study was conducted with two groups of EFL learners in Japan. One group consisted of 15 students majoring in English, and the other consisted of 31 Science students; in total, 46 students participated in the study. They had already learned 200 new words.

The target words came from the General Service List (West, 1953). This list contains the most frequently used 2000 English words. In order to determine the target words, the

Yoshii, M. (2011). Malaysian Journal of ELT Research, Vol. 7(1), p. 94-117. www.melta.org.my

researcher first gave a list of 1000 words to the learners. This list contained the second 1000 words of the GSL which, the researcher had determined, might not be known to the participants. Then, the researcher compiled the results and produced the top 200 words which this particular group of students did not know. Some samples of the target words appear in Table 2.

Table 2: Samples of the target words

plaster	vowel	basin	thorn
uppermost	grind	gaiety	ditch
elastic	inn	ounce	rejoice
hindrance	saucer	shilling	oar
dine	heap	axe	theatrical
inward	patriotic	cliff	paw
amongst	hurrah	widower	redde
sow	mill	verse	plow
brick	omit	hay	wreck
cork	sting	slippery	ripen

The students had just finished intensive learning of the 200 words in a month. The students were tested 3 times to see how well they could retain the words. The first test was given one week after the end of the intentional study of the words. The second test was administered one week after the first test. The students, then, had one month's break. The last test was given after the break, which was 7 weeks after the first test. This

provided ideal conditions for a retention study since the learners most likely would not go over what they had learned during the break.

The students were given self-rating tests where there were two states: One was State 1 (“I know this word”) and the other was State 0 (“I don’t know this word”). A simple computer programme was created for the self-rating tests and the students rated each word by clicking “Yes” (S1) or “No” (S0) on the computer programme.

The learners’ performance was compared with the predictions which were based on the transitional probability calculated from the first and the second tests. In order to make reasonable comparisons, it was first determined whether or not there were any outliers in the data. The difference between the predictions at the last test and the actual scores were checked. The mean difference was $M=21.9$ and the standard deviation was $SD=21.0$. The results showed 3 outliers; therefore, the data from the rest of the 43 participants was used for the final analysis.

Figure 6 shows the results of Study 1 with the actual scores represented by the line and the matrix predictions both at the last test and at the equilibrium represented by the bars. The data are arranged in the order of the actual scores with the highest scores on the left and the lowest on the right.

Figure 6: Actual scores (AS), matrix prediction at the last test (MP-L), and matrix prediction at equilibrium (MP-E) data in Study 1

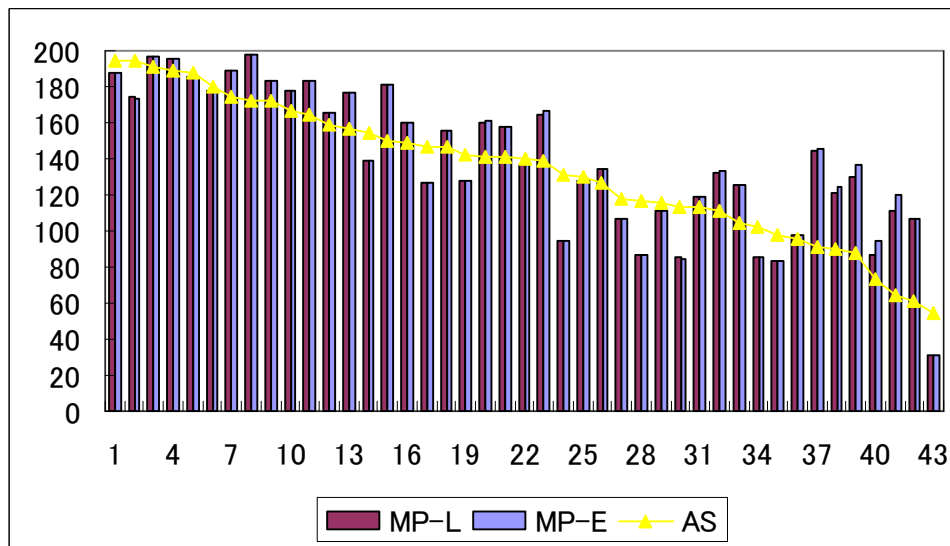


Figure 6 shows how well the predictions reflect the actual learners' data. Overestimations, which are represented by bars going over the line, and underestimations, which are represented by bars going below the line, seem to balance out, indicating an overall close relationship between the actual data and the predictions: Pearson's correlation between the actual scores and the matrix predictions at the last test was $r=.85$, $p=.000$; the correlation between the actual scores and the matrix predictions at equilibrium was $r=.83$, $p=.000$; the relationship between the predictions at the last test and the equilibrium was almost identical, $r=.99$, $p=.000$. This shows that 7 weeks was enough time for the vocabulary scores to settle into equilibrium states. The results verify the previous studies

and show that the matrix predictions can make accurate predictions for retention of newly learned words.

Second experimental study

The second study investigates how accurately the matrix model can reflect the actual data of learners' vocabulary test scores in the process of intentional learning. The study addressed the following research question: Are there any relationships between the actual data and the matrix predictions at the last test and at equilibrium?

The actual data were taken from 27 EFL learners in Japan, who were English-major university students. They learned 20 words per week intentionally for 10 weeks, covering 200 words by the end. The process of selecting the target words was the same as in Study 1. The researcher compiled the top 200 most likely unknown words for the students based on the vocabulary survey of the second 1000 words from the GSL.

Self-rating tests were administered every two weeks over the course of 10 weeks. The students took 5 self-rating tests over the course of the study. There were two states in the rating tests: One was State 1 ("I know this word") and the other was State 0 ("I don't know this word"). The same computer program as in Study 1 was used for the self-rating

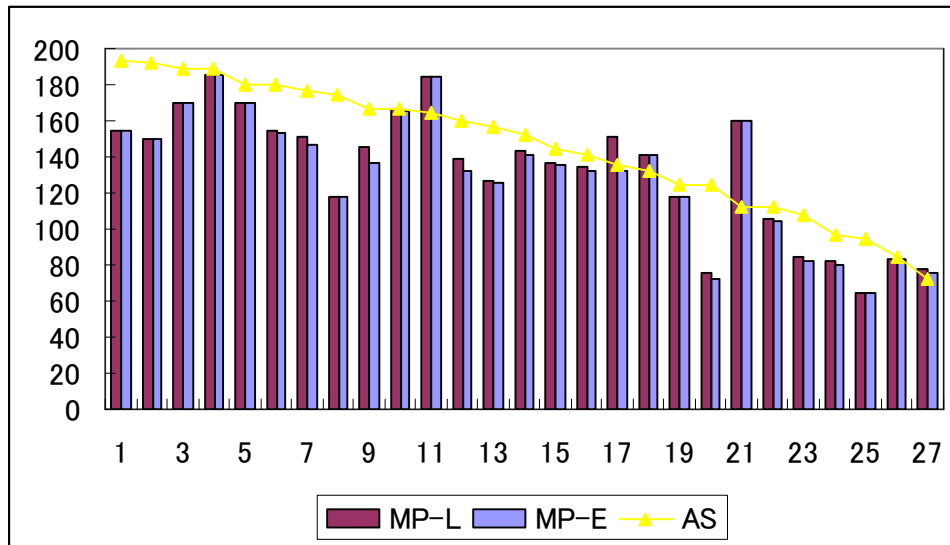
test. The students rated each word by clicking “Yes” (S1) or “No” (S0). Even though the students learned 20 words each week, each test consisted of all the 200 target words.

For the comparison of vocabulary test scores for the matrix predictions and actual data, correlations were used. In Study 1, there were only 3 tests and the first two tests were used to calculate the transitional probability and make the predictions at the third test. In Study 2, there were 5 tests, and this allowed the researcher to make a choice of calculating the transitional probability based on the data from Test 1 and Test 2 (T1-T2) or the data from Test 2 and Test 3 (T2-T3). There is a possibility that data sets from T2-T3 might be more reliable than those from T1-T2. Learners go through the test for the first time at Test 1, and may need some time to get used to the test itself. Therefore, the transitional probability taken from T2-T3 might be more reliable than that from T1-T2. In order to determine the most appropriate data sets, the correlations of actual scores at the last test (Test 5) and matrix predictions based on T1-T2 transitional probability and T2-T3 probability were compared. The results of the correlations were much higher for T2-T3 ($r = .86$) than for T1-T2 ($r = .78$); therefore, the T2-T3 transitional probability was chosen for analyzing the data in this present study.

The results of actual scores (AS) at the last test (Test 5), matrix predictions based on the T2-T3 transitional probability at the last test (MP-L), and the matrix predictions at equilibrium (MP-E) are shown in Figure 7. The actual scores of all the 27 learners are represented by the line in the graph and the scores were arranged in order of higher scores

on the left and lower scores on the right. The matrix prediction scores at the last test and at equilibrium are represented by vertical bars.

Figure 7: Actual scores (AS), matrix prediction at the last test (MP-L), and matrix prediction at equilibrium (MP-E) data in Study 2



The predictions appear to underestimate the actual scores. In order to determine the overall accuracy of the predictions, existence of any outliers in the data was examined first. This was done by looking at the differences between the actual scores and the matrix predictions. The average difference was $M = 20.2$ and the $SD = 15.8$. No outlier was detected in the data; therefore, all the 27 participants were included for the final analyses.

As can be seen, the matrix predictions at the last test and matrix predictions at the equilibrium time were almost identical and both underestimated the actual scores overall with the exception of a few learners. The results showed extremely high correlations between actual scores and matrix predictions: the correlation between the actual scores and the matrix prediction at the last test was $r = .80$, $p = .000$; the correlation between the actual scores and the matrix prediction at equilibrium was also $r = .80$, $p = .000$; and the correlation between the matrix prediction at the last test and the matrix prediction at equilibrium was $r = .99$, $p = .000$. All the correlations were extremely high, indicating the accuracy of the matrix models predicting the scores at the last test and at the equilibrium. The results show that predictions can be made accurately for the process of intentional learning.

Discussion and implications

The results of the two studies tell us that matrix models can accurately predict long-term vocabulary scores both in retention and in intentional learning. What are the implications of the studies? What do the results really mean to us?

First, they give us a new perspective on the way we assess learners on their vocabulary performance. An assessment is usually given at the end of a certain task or a certain

learning period such as a school term or semester. Vocabulary tests are given to show how many words learners know at a given time. It is all reactive since the test results reflect what learners know or have learned. However, the matrix prediction opens up a new avenue for assessment. Tests can be given two times at the beginning of a learning period. Based on the two data sets and the transitional probability calculated from them, teachers could predict future performance. This means that assessment may function in a proactive rather than reactive way. Teachers, for example, can use the tool to give appropriate advice to learners regarding their future vocabulary development based on the predicted data. If a learner's predicted score is desirable, the instructor simply provides the learner with affirmation of his or her learning and can encourage the learner to keep up with the good work. If a learner's prediction is not desirable, the instructor could share that information with the learner and provide necessary early intervention.

Secondly, the matrix models might help teachers provide specific advice to learners on their vocabulary learning rather than simply making general remarks telling learners to work harder. Teachers not only could make reasonable predictions about vocabulary scores to learners, but might also be able to make predictions on the change patterns of vocabulary knowledge based on the transitional probability. A transitional probability displayed in a matrix could be translated in terms of patterns of changes as in Table 3. As seen in the Table, change patterns can be classified in the four boxes in the matrix: two

no-changes and two changes. No changes can be further classified as remaining as S0 or S1; and changes can be further classified as plus and minus change.

Table 3: Possible change patterns in Transitional probability matrix

		Time n + 1	
		S0	S1
Time n	S0	S0 → S0 (No change)	S0 → S1 (Plus change)
	S1	S1 → S0 (Minus change)	S1 → S1 (No change)

This would enable teachers to give more specific advice to learners about the characteristics of their learning, specifically, about the change patterns of vocabulary knowledge. For instance, if a learner has a high proportion of S0→S0 (S0 stays in S0) pattern, a teacher could point out the learning difficulty of certain words which tend to remain as S0. The teacher could investigate further the possible causes of such a high proportion of this pattern. If the cause derives from a problem relating to a particular group of words, the teacher may be able to analyze them and give specific guidance for learning those difficult words. If a learner has a low S1→S1 (S1 stays in S1) portion, meaning that the words he or she learns tend to be forgotten easily, teachers could work with the learner, finding ways to retain the knowledge of the words she or he learns. This

type of new assessment of change patterns has the potential to help teachers provide more specific and tailor-made advice on their vocabulary learning to learners.

Conclusion

This paper has discussed the potential that matrix models have for vocabulary assessment. The studies on matrix models have also revealed the need for future studies. First, more verification is needed to examine the validity of the predictions made in this study, particularly in an intentional learning setting. This study was built on the assumption that each treatment, in this case the learning of 20 words, was the same in quality. The number of words for each treatment was the same; however, the difficulty level of learning may differ for each set of 20 words and even for each word. The results of the study, therefore, need to be verified in similar settings, preferably involving more learners and for longer periods than 10 weeks.

Future studies need to investigate the transitional probability matrix itself. For example, the assumption regarding the stability of the transitional probability needs to be verified. Accumulation of learner data from future studies would give us insights into the probability. For example, the probability may have certain ranges, and this information would be useful for conducting simulations. It would be worth examining whether

transitional probability itself can give us enough information on prediction of change patterns. If so, profiling learners can be much more productive than conducting simulations each time.

Research on matrix models has just begun. This type of theory-based research has much room for growth. The theory requires verification from data with learners. The results will confirm, reform, or reject the theory. More theory-based studies are needed for the advancement of vocabulary research and second language acquisition.

This research was supported by a Grant-in-Aid for Scientific Research (C) (1) 2008-2010 from the Japan Society for the Promotion of Science (16520347).

References

- Bradley, I., & Meek, R. L. (1986). *Matrices and society*. Harmondsworth: Penguin.
- Coady, J., Magoto, J. Hubbard, P., Graney, JU., & Mokhtari, K. (1993). High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 217-228). Norwood, JN: Ablex.
- De Groot, A. & Keijzer R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign language learning and forgetting. *Language Learning*, 50(1), 1-56.
- Yoshii, M. (2011). *Malaysian Journal of ELT Research*, Vol. 7(1), p. 94-117. www.melta.org.my

- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46-64). Norwood, NJ: Ablex.
- Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Journal*, 56(2), 308-328.
- Joe, A. (1995). Task-based tasks and incidental vocabulary learning: A case study. *Second Language Research*, 11(2), 159-177.
- Joe, A. Nation, P., & Newton, J. (1996). Vocabulary learning and speaking activities. *English Teaching Forum* 34(1), 2-7.
- Jones, F. (1995). Learning an alien lexicon: A teach - yourself case. *Second Language Research*, 11(2), 95-111.
- Laufer, B. (1997). The lexical plight in second language reading: Words you dont know, words you think you know, and words you cant guess. In J. Coady & T. Huckin (Eds.) *Second language vocabulary acquisition* (pp. 20-34). Cambridge: Cambridge University Press.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3) 307-322.
- Laufer, B., & Paribakht, T. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48 365-391.
- Meara, P. (1990). Matrix models of vocabulary acquisition. *AILA Review* 6, 66-74.
- Meara, P., & I. R. Sanchez (2001). A methodology for evaluating the effectiveness of vocabulary treatments. *Reflection on language and language learning*. Edited by M. Bax & J-W Zmert. John Benjamins Publishing Co. Amsterdam. pp.267-278.

- Newton J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, 11(2) 159-177.
- Parry, K. (1993). Too many words: Learning the vocabulary of an academic subject. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46-64). Norwood, NJ: Ablex.
- Sanchez, I. R. (2000). *Matrix models of second language acquisition*. Unpublished doctoral dissertation, the University of Wales, Swansea.
- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning* 48(2), 281-317.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 577-607.
- West, M. (1953). *A general service list of English words*. Longman, London.
- Yoshii, M. (2009a). Investigation into word knowledge change: Comparison of simulation data and actual data. *Language Issues: The International Journal of the Academic Information and Media Center*. Prefectural University of Kumamoto, 13-15(1), 1-22.
- Yoshii, M. (2009b). Matrix model-wa Mochiita Simulation Study: Chookiteki-na Goi Chishiki Henka-no Kenshoo. [Matrix model simulation: An investigation into long-term vocabulary change.] *Journal of the Faculty of Letters. Prefectural University of Kumamoto*, 15(68), 65-79.