Commentary on Four Studies for JALT Vocabulary SIG

Paul Nation

LALS, Victoria University of Wellington
doi: http://dx.doi.org/10.7820/vli.v02.1.nation

Abstract

Four papers by Tomoko Ishii, Emilie Masson, Atsushi Mizumoto and Rachael Ruegg will be presented in morning session of the Second Annual *JALT Vocabulary SIG Vocabulary Symposium* in Fukuoka, Japan on July 2, 2013. As discussant, it is my pleasure to comment upon each manuscript. These four lexical researchers originate from all over Japan: Tokyo, Fukuoka, Osaka and Akita. The range of their lexical topics is almost as wide: semantically related words; loanword usage difficulties; self-efficacy and autonomy in vocabulary learning; and, lexical feedback on writing. After commenting on each paper in turn, I shall present a few suggestions for their future research.

1 Research on vocabulary

The amount of research focused on the teaching and learning of vocabulary has increased enormously. Around 30% of the research on vocabulary in the last 100 years has appeared within the last 10 years. In addition, much of this research is focused on vocabulary learning in a second or foreign language, and we are probably now at a point where the amount of second language vocabulary research outnumbers the amount of first language vocabulary research.

In addition, there is now a keener awareness of the various aspects of research methodology for investigating vocabulary. Two books on research methodology in vocabulary studies have recently appeared (Nation & Webb, 2011; Schmitt, 2010), and there is a great deal of innovation in how vocabulary-focused research is carried out (see Schmitt, 2004 for an excellent example). There is now more discussion of issues like time-on-task, multiple tests of the same words, test items, and factors affecting word difficulty.

There is also a growing body of researchers who have made research on vocabulary their prime focus and who persistently research and publish in this area. This persistence is important, because it allows for the refinement of research techniques and the chipping away at an area of research so that findings are confirmed or challenged, and the research area is looked at in increasing detail. Some of the areas where this is occurring include multiword units, text coverage and comprehension, vocabulary-teaching techniques, word lists, and vocabulary size.

Research on vocabulary has been particularly helped by the very rapid advances in computer technology. As Tom Cobb puts it, the relationship between computer technology and vocabulary is a marriage made in heaven.

The following studies very usefully add to our knowledge of the teaching and learning of vocabulary. They are very different from each other with differing focuses and different research methodologies.

2 The four studies

2.1 Re-examining semantic clustering: insight from memory models by Tomoko Ishii

This study looks at semantically related words to see if the nature of the semantic relationship affects the possibility of interference. The study focuses on visual features – *banana* and *orange* are less likely to interfere with each other because they have very different shapes, whereas *apple* and *oranges* are more likely to interfere because they have a rather similar shape. This likelihood of interference is of course not limited to shape. The study found some support for the idea.

This is a very worthwhile study indeed. It takes the research on interference to a new level, adding much needed detail to an area that needs further research like this. The paper is deliberately short, and I am aware that several of my comments relate to areas that were probably not mentioned because of space limitations.

Because nonsense words were used, no pre-test was needed. The comparisons were between the three different relationships so no control group was needed. I generally recommend three tests for each word when measuring the effects of vocabulary interventions. Would that recommendation apply to this study? Probably yes, because it would be interesting to see if strength of knowledge related in some way to interference. Are partially known words more likely to interfere because of the fragility of the knowledge? This study used an L1 recall test (see the L2 word – write its meaning in Japanese). This is a tough test and more sensitive measures like multiple-choice (with the correct answers to other items as distractors) and interview (a version of the Vocabulary Knowledge Scale) might provide useful insights.

It would be good to see an analysis of the relationships between the words in at least the Tinkham and Waring studies (and perhaps Higa?). My guess is that the researcher has already done this. My guess is that this analysis might support the present study.

I agree with the comment about part of speech in the study of Tinkham (1997). I regard the confounding of part of speech and thematic relationship to be a weakness of the Tinkham study, resulting in an underestimation of the thematic effect.

In the present study there is evidence of a ceiling effect with some learners scoring 18 out of 18, at least initially. I guess this was a small number and did not greatly affect the results.

The choice was made for the study to be within subjects. I think this was the best choice. This raises the possibility of interactions between the items in the three groupings. I wonder if there was evidence of this, for example *stone* and *globe*, *knife* and *chopsticks*. I think Japanese like Thai use classifiers for items of various shapes

and functions. I wonder if the classifiers in Japanese affected learning? The study stands or falls on the quality of the operationalisation of the groupings. Was there any pre-testing that could be done to guarantee the quality of the groupings? This pretesting could involve some kind of classification activity somewhat similar to Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) studies on prototypicality to see what items are seen as being more closely related.

Some of the semantically related items share a physical relationship – *Japanese radish, cucumber, okra; knife, fork, ladle*. Was there evidence of inter-ference between these items?

The test was a receptive recall test (L2 $\geq$ L1). This makes it possible to look at aspects of the process in the sense that wrong translations can be analysed. I was pleased to see the researcher do this to a small degree (p. 4), but I wonder if more analysis is possible with the existing data. Such data provides a compelling evidence for what is causing the difficulty.

I assume that the items were presented in category groupings during the learning. If this was so, it is interesting to speculate on the effect of the tests involving a random ordering. This would certainly increase the difficulty of the task (where testing conditions match learning conditions, recall is better). Or, would this lessen the effect of interference?

I greatly enjoyed reading this study and thinking about it. It focuses on an area where I have special interests, and it takes an innovative question and explores it carefully and well.

Tinkham and Waring used trials to criterion. When re-reading Tinkham's studies a few years ago, I wondered how his subjects would have behaved on a delayed test. Would those who took more trials to learn have better long-term retention because of the extra repetitions and heavier learning burden (depth of processing?) during the learning? I was going to contact Tinkham about this and found to my dismay that he had died during a medical operation. At least he knew that I greatly admired his research although I had never met him.

2.2 How L1 loanwords can create a false sense of familiarity with L2 vocabulary meaning and usage by Marie-Emilie Masson

This paper has a very useful focus. It goes beyond looking at form and meaning relationships and considers the third major aspect of what is involved in knowing a word, use. It builds very usefully on the work of Frank Daulton (2008).

The study looks at the various types of loanword relationships (true cognates, convergent cognates, divergent cognates, close false friends and distant false friends) and their connection with knowledge about how to use the words. Skill at using the word was measured by getting the learners to write a sentence containing the word.

The study found that the easiest words from a form and meaning perspective (true cognates and convergent cognates) were among the worst performing on the measure of word use. This is a very interesting finding – what we are most confident about is our greatest weakness.

The Vocabulary Knowledge Scale as used in the study has come in for a lot of criticism (see Nation and Webb (2011, section 14.9) for references). One criticism is that it is really two scales, in that it combines meaning and use (Step 3: *I know more than one meaning for this word and I can write a sentence using this word*). In the present study it could be argued that having these two aspects present is desirable, because the study looks at both of them and their relationship. It may however be considered a little unfair as the subjects cannot choose between knowing the meaning and using the word. Their choice is both or none.

Two tests are being compared here. The scale test which is a recall test of receptive knowledge, and a productive recall test of the use of the word. There should be some agreement between the tests, but they are testing rather different things. We have a difference of formats and a difference of types of knowledge. What is in common is that the tests are of the same word. The reason that I point out these format differences is that they add to the likelihood of lack of agreement between the tests. Different formats testing knowledge of word meaning for the same words correlate at around 0.7. Format has a strong effect. Having said that, I think that the formats chosen were probably the best ones to use for this study.

I was thinking how this study could also be applied to word card learning. Presumably there would be similar findings – knowing the receptive meaning does not mean knowing the use. As with word card learning, having productive knowledge of the word would probably not mean that the word could be used correctly.

The message of the paper is that receptive knowledge does not mean productive use. Receptive knowledge may blind learners to needing to learn more about the word. There needs to be a range of opportunities to learn how to use loanwords and cognates.

2.3 Enhancing self-efficacy in vocabulary learning: a self-regulated learning approach by Atsushi Mizumoto

This is an interesting study. It has the strong points of being a longitudinal study and involving more than one measure of learning. It also provides me with an opportunity to comment on a couple of features of research that I consider important.

The first is pedagogical significance. Researchers are well aware of statistical significance. Statistical significance is related to the size of the difference between treatments particularly with regard to whether the difference could have occurred by chance without there being any real connection between the treatments being compared and the measure. Pedagogical significance goes a step further and asks if the size of the difference between treatments is large enough for a teacher or learner to get excited about. So it is possible to have statistical significance without pedagogical significance. In the studies on interference, the difference in the number of learning trials for interfering items and non-interfering items is 50–100%. This is a very big difference and is one that teachers and learners would see as having a big effect on learning. If the difference was only one or two trials out of 20, then it would be less recent to get excited about the effect. Pedagogical significance is

important because even if research shows that there is a significant difference between two treatments, if that difference means virtually no change in the amount of effort or work that the learner has to do, then there is no strong reason for adopting the findings.

Let us now look at pedagogical significance in the present study. The amount of learning involved was 10–15 minutes each time for 30 times, around 5–7 hours in total. On the vocabulary measure, this resulted in an increase of 3.74 out of 19 (41 out of 60 were already known) which is approximately a one-fifth increase. This compares with the controls with a 1 out of 19 (approximately one-twentieth) increase. This seems to me enough for pedagogical significance.

Pedagogical significance also involves looking at cost/benefit. That is how much work was needed to achieve how much increase? There are studies where large amounts of time were spent achieving learning rates of around one word per hour (see for example Carlo et al., 2004). It does not make sense to make such a calculation in the present study because the aim was primarily to develop self-efficacy, not just learn vocabulary. However, this study used a text containing vocabulary exercises (I assume that the Schmitt, Schmitt and Mann text was used for the 30 weeks). Vocabulary exercises are generally not very efficient as a means of vocabulary learning. Results from the various involvement load studies for example give a receptive learning return of $<50\%$. Word cards may have been more efficient than the book. What I am calling pedagogical significance consists of effect size and efficiency.

The second issue I want to focus on is the use of questionnaires. While I think that in this study a questionnaire was an appropriate measure, in many studies it is often adopted because of its convenience rather than its validity. Questionnaires typically measure opinions and attitudes. The big problem I have with them is that opinions and attitudes are one step at least away from action. That is, people can say that they believe in certain things, but it is their actions that count and it is not uncommon for stated beliefs and actions to disagree with each other. It seems to me that one of the real challenges in measuring self-efficacy is finding measures that most validly measure self-efficacy. I have no doubt it can be done, but it is likely that the most valid measures will involve individual observation of some kind.

I think the focus on self-efficacy and autonomy in vocabulary learning is very important. Because there are so many words to learn and because deliberate learning will play an important role in this learning because of its efficiency, learners need to take responsibility for their own vocabulary learning.

2.4   A comparison of lexical feedback on writing from peers and a teacher by Rachael Ruegg

This study looked at feedback on vocabulary in writing over several pieces of writing. Is this an experimental or quasi-experimental study? Intact classes were used but the assigning of classes was the result of random placement. Because intact classes were used it is probably quasi-experimental.

In the review of the literature, Laufer and Hulstijn's rating for written assignments is mentioned. This involvement load rating refers to the particular type

of task they used. It is likely that academic writing of the kind involved in this study would be rated even higher.

The major weakness of the study is that we have the amount of feedback but not amount of error. Were lexical errors noticed? What proportions were noticed? In terms of actual errors, what was the proportion of word choice error to word use error? Without this information we cannot assess if the balance of feedback (word choice versus word use) was appropriate or not. The feedback ratio on word choice versus word use was 1:6 by the teacher. Does this reflect the proportion of error or is this out of line with the errors in the writing?

Ideally the judgement of errors would need to be made twice, once by the teacher involved and once by external raters.

The conclusions of the study assume sophisticated vocabulary is better than more frequent vocabulary and that collocational error largely goes unnoticed. The study does not provide evidence for this. There is some research however that supports the idea that more sophisticated vocabulary is seen as a sign of good writing.

3 Future research

The research included in this review covers a wide range of topics. This is a reflection that research on vocabulary is alive and well and that researchers are finding topics that can be usefully researched. I thought I would conclude this review by suggesting avenues for further vocabulary-focused research. Let me do this by using the four studies in this review as starting points.

Research on interference is still really in its infancy. We have the beginning of a range of potentially interfering relationships including lexical sets, opposites and near synonyms. Only the research on lexical sets has been replicated. Higa's findings on near synonyms and opposites have not been replicated. Are there other negative relationships? It would be interested to look at near synonyms in the same detailed way that lexical sets were investigated in Tomoko Ishii's research. Besides Tinkham's (1997) study, there has been no investigation of relationships facilitating learning. Taking such a positive approach could provide useful guidelines for teaching and learning.

Research on loanwords has rarely gone beyond learning isolated lexical items. Marie-Emilie Masson's research suggests very useful directions, namely looking at what other learning needs to occur. Stuart Webb's (2007) research comparing learning with and without context found that a lot of knowledge about a newly learned L2 word could be carried over from L1 and previous L2 knowledge and commonsense. For example, if a learner has some understanding of the countable/ uncountable distinction from earlier study of English, this may not need much if any attention when a new uncountable noun is learned. Does this work with loanwords or is the influence of the L1 so strong that L1 rules and patterns dominate? Does this make loanwords in some ways more difficult than completely new words?

Research on self-efficacy shows great promise for vocabulary learning. Learners need to know a lot of vocabulary and incidental learning needs to be accompanied by more efficient deliberate learning. Vocabulary teaching and

vocabulary activities can only deal with a very small proportion of the words that need to be learned. Atsushi Mizumoto's research shows that with time learners can take on much of the responsibility for deliberate vocabulary learning. We need to know more about what motivates learners to take responsibility for their learning. Change theory suggests that law, reason and involvement are major avenues for bringing about complex change. However, we need to see what practical options in each of these avenues are most effective.

Rachael Ruegg's research is in a virtually unexplored area. While there are many studies of grammatical features in writing, there are very few that focus on vocabulary. The lexical richness studies using the Lexical Frequency Profile (Laufer & Nation, 1995) need to be enriched by discourse studies that examine the role of the vocabulary in texts. I see a relationship between studies like this that look at production and studies that look at lexical coverage for comprehension. Looking at what the words contribute or do not contribute to a particular text is important.

Vocabulary is no longer a neglected area of research and the studies reviewed here testify to the range of focuses that are possible. I look forward to seeing the research reviewed here being published and thus adding more broadly to our knowledge of the field.

References

Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D.N., . . . White, C.E. (2004). Closing the gap: addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39* (2), 188–215. doi:10.1598/RRQ.39.2.3

Daulton, F.E. (2008). *Japan's built-in lexicon of English-based loanwords.* Clevedon: Multilingual Matters.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics, 16* (3), 307–322. doi:10.1093/applin/16.3.307

Nation, I.S.P., & Webb, S. (2011). *Researching and analyzing vocabulary.* Boston: Heinle Cengage Learning.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439. doi:10.1016/0010-0285(76)90013-X

Schmitt, N. (Ed.). (2004). *Formulaic sequences.* Amsterdam: John Benjamins.

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual.* Basingstoke: Palgrave Macmillan.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13* (2), 138–163. doi:10.1191/026765897672376469

Webb, S. (2007). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language, 19* (2), 120–136.