A Comparison of Long-term Recall Accuracy for Semantically Related and Unrelated Vocabulary Items

Michiko Masaki
Osaka International University

Brad Visgatis
Osaka International University

This paper reports on a pilot study investigating the effect of vocabulary list organization on long-term item recall. Two 30-item lists and related multiple-choice tests were developed. One contained sets of related words, the other unrelated words. Students received one list, were tested in week two, then received the other list for testing in week three. Three months later, they were tested on both lists. Results show no significant difference in recall of the items.

この論文は、語彙リストの構成が長期記憶に与える影響を調べる研究の、パイロットスタディについて報告するものである。意味的に関連のある３０語のリストと意味的に関連のない３０語のリストを作成し、それらのリストに対応した、４択のテストを用意した。学生は、一方のリストを受け取り、１週間後にそれに対応したテストを受けた。同時に、もう一方のリストを受け取り、更に１週間後にそれに対応したテストを受けた。３か月後、両方のリストにあった合計６０語から成る４択のテストを実施した。その結果、意味的に関連のある語のリストと意味的に関連のない語のリストが長期記憶に与える影響には、有意な差はなかったということがわかった。

Introduction

There is growing interest in investigating vocabulary in foreign language teaching and learning and several important texts have recently appeared (Nation, 1990, 2001; Read, 2000; Schmitt & McCarthy, 1997). Moreover, the autumn 2003 issue of TESOL Quarterly was devoted to corpus linguistics. Finally, at JALT2003 there were seven presentations that directly addressed issues of vocabulary.

Language acquisition involves learning new vocabulary. The action of learning is achieved through mental processing (Hayes, 2002). The most common type of mental processing used by adults for learning vocabulary is symbolic representation. In the case of foreign vocabulary, the word

is linked to first language vocabulary. The actual link may be according to any number of shared characteristics, including: sound, spelling and use domain. Success in memorization is related to both the degree of effort as well as to the actual "learning burden" (Nation, 2001) of the particular words. Nation explains that "the more a word represents patterns and knowledge that learners are already familiar with, the lighter its learning burden" (pp. 23-24), and hence the easier it will be to learn.

Given the above, it would seem reasonable to assume that material designers would develop vocabulary study materials that would seek to reduce the mental processing load in order to facilitate item memorization. Moreover, it would also be reasonable for teachers to adopt similar aims when preparing materials for students to focus on. One example of a technique for accomplishing this is to link vocabulary items so that they can be memorized as a set or chunk rather than individual instances.

Indeed, many of the newer vocabulary study materials try to present words in ways that can aid in their memorization (Furutou, 1995; Kazahaya, 2003; Oubunsha, 1994). In many cases this consists of presenting each new English item together with a few synonyms or examples in Japanese, thus supposedly reducing the mental processing that is required for linking new words to previously known ones. That said, the particular organization of the lists is still predominantly according to frequency or alphabetical (usually within a class of words: noun, adjective, etc.).

Moreover, many teachers continue to adopt these organizational plans (alphabetical or frequency-based) when developing or presenting new vocabulary lists for out of class study, especially in classes where teachers focus on test or reading skills. This trend in materials development is often reinforced by other types of readily available supplementary resource material, such as corpora (i.e., the British National Corpus (Leech, Rayson, & Wilson, 2001)) or basic word lists (JACET, 1993).

While previous research has often focused on the learnability of individual words, (A. M. B. d. Groot & Keijzer, 2000; P. J. M. Groot, 2000) not much looks at how the organization of the particular vocabulary list in question impacts upon learning. It is this area where our interest lies. Specifically, we seek to compare long-term recall for items introduced through two equivalent but differentially-organized vocabulary lists, one consisting of semantically-related items and one of non-related items.

We set the following as our working hypothesis: Words introduced by way of a vocabulary list consisting of semantically related items will be more accurately recalled than words introduced by way of a non-semantically related vocabulary list. Secondarily, we seek to gather information about the general characteristics of long-term English vocabulary recall by university students.

Realizing that this is a complex issue to study, we divided it into three stages (see Table 1). This paper only reports results from stage 1 of our investigation.

Method

Any number of variables can affect performance, including motivation, preparation time and manner, instruction, and familiarity with the instrument type. In order to identify the effect that the organization of a particular vocabulary list has on long-term recall, it is essential to control whatever effect

Table 1. Stages

Stage	Time	Contents
1	June 2003 ~ October 2003	Pilot test stage
2	October 2003 ~ January 2004	Wider administration to pinpoint patterns
3	April 2004 ~ July 2004	Revised study administered to wider group of subjects

other variables may exert. This principle guided many of our implementation decisions.

Our overall plan consisted of the following steps.

1. Develop two equivalent vocabulary lists, one of semantically related words (related) and one of non-semantically related words (unrelated)
2. Select groups of subjects and inform them about the study
3. Distribute the first list and tell students to study for a test the following week
4. Administer the first test
5. Distribute the second list
6. Administer the second test
7. Administer the final test

Vocabulary list construction

Two vocabulary lists were constructed. List A contained 10 sets of three semantically related items, for a total of 30 items. The second, List B, consisted of 30 unrelated items. We attempted to make both lists equivalent in all other possible aspects (see Table 2): average item length in syllables and letters, distribution of abstract and concrete items, proportion of syntactic category (nouns, verbs, adjectives, and adverbs). Where equivalency was not attainable, we attempted to make List A the more difficult. Items in both lists were arranged alphabetically.

We adopted three types of semantic relation for List A items: Synonymous (S), Shared-Domain (SD), and Ideational (I) (see Table 3). While we felt these relations to be valid, we surveyed 11 teachers in order to verify if our assumptions were correct. We found that they did not perceive the same distinction between synonymous and ideational relations as we had posited. In the same survey, we also asked the teachers to indicate whether items were abstract (A) or concrete (C). Their responses validated our judgments (see Table 4).

We did not verify whether subjects had studied the vocabulary items previously. However, when we constructed the lists, we chose low-frequency items based on the British National Corpus (Leech et al., 2001) and used our experiences teaching these and similar students to make the final determination. The final word lists and item test results are given in Table 6.

Table 2. List equivalencies

Aspect		List A (Related Words)	List B (Unrelated Words)
Length in syllables	Minimum	1.00	1.00
	Average	2.63*	2.27
	Maximum	5.00	5.00
	Std. Dev.	1.07	1.08
Length in letters	Minimum	4.00*	3.00
	Average	7.23*	6.93
	Maximum	12.00	12.00
	Std. Dev.	2.14	2.35
Quality	Abstract	15*	12
	Concrete	15	18
Part of speech	Noun	6	6
	Adjective	3	6
	Verb	15	12
	Adverb	6	6

* = aspects of greater difficulty

Table 3. Definitions

Relation (Code)	Definition
Synonymous (S)	Words that could be used interchangeably in certain circumstances without greatly altering the meaning.
Shared-Domain (SD)	Words commonly used within the same general domain or in conjunction with a particular item, such as science, clothing, bookbinding, etc., and most often refer to parts of some whole entity.
Ideational (I)	Words that refer to some characteristic or trait of a shared domain or entity, but which are not synonymous.

Table 4. Verification of relation and abstractness

#	Item A	Item B	Item C	A	C	Ours	S	SD	I	Ours
1	Convoluted	enigmatic	byzantine	7	4	A*	1	1	9	S
2	Resolutely	determinedly	doggedly	6	5	A	10	1		S*
3	Confiscate	abscond	impound	1	10	C*	3	3	5	I
4	Stamen	sepal	petal	1	10	C*		10	1	D*
5	Sidle	limp	slither	1	10	C*	2	2	6	I*
6	Cogitate	ponder	mull	7	4	A*	10		1	I
7	Aura	mien	demeanor	8	3	A*	4	3	4	I
8	Hinder	impede	stymie	6	6†	C	11			S*
9	Amiably	hostilely	cordially	3	8	C*		2	9	I*
10	Foster	engender	facilitate	6	5	A	6	3	2	I

* = areas with agreement † = one teacher marked both as correct

Table 5. Schedule by group

Group	1	2	3	4	5	6
Subjects	19	20	13	17	12	12
Week 1	23-Jun-03		25-Jun-03		27-Jun-03	
	List A	List A	List A	List B	List B	List A
Week 2	30-Jun-03		2-Jul-03		4-Jul-03	
	Test A	Test A	Test A	Test B	Test B	Test A
	List B	List B	List B	List A	List A	List B
Week 3	7-Jul-03		9-Jul-03		11-Jul-03	
	Test B	Test B	Test B	Test A	Test A	Test B
Week X	20-Oct-03		8-Oct-03		17-Oct-03	
	Test AB					

Table 6. Word lists and item test results

List	Item	Set	Characteristic	Test AB		Test A		Test B		Change	Note
			Concrete (C) or Abstract (A)	IF	∑	IF	∑	IF	∑	Test AB – Test A or Test B	
A	Abscond	01a	C	36.6%	93	79.1%	91			-42.6%	
A	Confiscate	01b	C	26.1%	92	70.7%	92			-44.6%	
A	Impound	01c	C	69.6%	92	87.6%	89			-18.1%	
A	Amiably	02a	C	64.5%	93	93.3%	89			-28.7%	
A	Cordially	02b	C	31.9%	91	76.9%	91			-45.1%	
A	Hostilely	02c	A								*
A	Aura	03a	A	52.7%	93	95.7%	93			-43.0%	
A	Demeanor	03b	A	48.4%	91	78.0%	91			-29.7%	
A	Mien	03c	A								*
A	Byzantine	04a	A	75.8%	91	84.6%	91			-8.8%	
A	Convoluted	04b	A	52.2%	92	77.2%	92			-25.0%	
A	Enigmatic	04c	A	56.0%	91	86.8%	91			-30.8%	
A	Cogitate	05a	A	20.7%	92	63.7%	91			-43.1%	
A	Meditate	05b	A	59.8%	92	52.7%	91			7.0%	
A	Mull	05c	A								*
A	Determinedly	06a	A	52.7%	93	72.7%	88			-20.0%	
A	Doggedly	06b	A	48.4%	91	75.0%	92			-26.6%	
A	Resolutely	06c	A	34.4%	93	72.2%	90			-37.8%	
A	Facilitate	07a	A	25.8%	93	71.7%	92			-45.9%	
A	Engender	07b	A	34.4%	93	74.7%	91			-40.3%	
A	Foster	07c	A	54.8%	93	82.8%	93			-28.0%	
A	Hinder	08a	C	68.8%	93	85.9%	92			-17.1%	
A	Impede	08b	C	40.7%	91	73.6%	91			-33.0%	

A	Stymie	08c	C	59.8%	92	86.8%	91			-27.0%	
A	Limp	09a	C	59.8%	92	82.6%	92			-22.8%	
A	Sidle	09b	C	59.8%	92	74.4%	90			-14.7%	
A	Slither	09c	C	66.3%	92	75.8%	91			-9.5%	
A	Petal	10a	C	55.9%	93	79.3%	92			-23.4%	
A	Stamen	10b	C	78.5%	93	91.3%	92			-12.8%	
A	Sepal	10c	C	45.2%	93	76.1%	92			-30.9%	
B	Avidly		C	65.6%	90			92.4%	92	-26.8%	
B	Awe		A	68.5%	92			84.9%	93	-16.5%	
B	Choppy		A	44.6%	92			80.6%	93	-36.1%	
B	Coalesce		A	34.8%	92			75.0%	92	-40.2%	
B	Concoct		A	52.7%	91			84.8%	92	-32.0%	
B	Congregate		C	41.3%	92			65.2%	92	-23.9%	
B	Deviously		A	50.0%	92			81.7%	93	-31.7%	
B	Devotedly		C	47.3%	93			76.1%	92	-28.8%	
B	Engrave		C	26.9%	93			77.2%	92	-50.3%	
B	Ennui		A	63.7%	91			82.8%	93	-19.1%	
B	Evaporate		C	51.6%	93			79.3%	92	-27.7%	
B	Farce		A	38.5%	91			78.5%	93	-40.0%	
B	Flit		C								*
B	Florid		A	46.2%	93			79.6%	93	-33.3%	
B	Fracture		C								*
B	Gill		C	63.0%	92			84.9%	93	-21.9%	
B	Haphazardly		C	48.9%	92			70.7%	92	-21.7%	
B	Horrible		A	71.0%	93			79.1%	91	-8.2%	
B	Implicate		C	45.7%	92			68.1%	91	-22.5%	
B	Malaise		A	40.0%	90			64.1%	92	-24.1%	
B	Murmur		A	59.1%	93			83.9%	93	-24.7%	

B	Patella		C	38.7%	93			75.8%	91	-37.1%	
B	Probe		C	46.2%	91			66.7%	93	-20.5%	
B	Prow		C	50.0%	92			69.6%	92	-19.6%	
B	Reciprocal		C								*
B	Revere		C	23.9%	92			61.3%	93	-37.4%	
B	Segregate		C	69.6%	92			75.3%	93	-5.7%	
B	Slur		C	53.8%	91			76.9%	91	-23.1%	
B	Snub		A	53.3%	92			71.0%	93	-17.7%	
B	Vociferously		C	51.7%	89			77.4%	93	-25.7%	

* = excluded items

Selection of subjects

Six groups of subjects, all native speakers of Japanese, were drawn from five intact groups of students in first and second year English courses at two coeducational universities in the Kansai area of Japan (see Table 5). One intact group was split so there would be three paired groups. Each paired group was taught by one teacher to provide consistency. Data on the ratio of women to men or first year to second year was not collected. Due to the high attrition rate common in long-term studies, of the 143 initial subjects, complete data sets could only be collected from 93.

Development of the testing instrument

To evaluate the short- and long-term recall, three multiple-choice tests were created. The rationale behind this test format was twofold: it was familiar to students and it would take a short time to administer.

The tests were four-option English to Japanese matching tests. One 30-item test (Test A) assessed short-term recall of related words from List A. The second 30-item test (Test B) assessed short-term recall of unrelated words from List B. A third 60-item test (Test AB) was developed to check long-term recall. It contained all of the items from the other tests. All items were randomized. No steps were taken to estimate the reliability of the instruments.

Unfortunately, during copying, various problems occurred that were caught too late to correct (i.e., printing artifacts that resembled strike-outs) and we excluded those items (three each from Test A and Test B), leaving 54 out of the original 60 items.

Results

The data (see Table 7) clearly shows that overall recall declined between the pre-tests (Tests A and B) and post-test (Test AB). Test A had 78.58% correct, Test B 76.41%. Overall pre-test average was 77.49%. Test AB percent correct was 50.48%. On

Table 7. Composite changes

	Test AB	Test A	Test B	Average Test A & Test B	Test AB - Test A	Test AB - Test B	Test AB - Test A & Test B Average
All Item Totals							
Minimum	20.65%	52.75%	61.29%	57.02%	-45.93%	**-50.29%**	
Average	**50.48%**	**78.58%**	**76.41%**	**77.49%**	-27.49%	-26.53%	**-27.01%**
Maximum	78.49%	95.70%	92.39%	94.05%	7.04%	-5.70%	
Relatedness Totals							
Related	51.09%	78.58%			**-27.49%**		
Unrelated	49.87%		76.41%			**-26.53%**	
Characteristic Totals							
Abstract Average	49.87%	76.00%	78.84%	77.42%	-26.13%	-28.97%	**-27.55%**
Concrete Average	50.65%	80.97%	74.46%	77.71%	-30.31%	-23.81%	**-27.06%**

average, recall scores dropped -27.01%. Individual items ranged from an increase of 7.04%, to a decrease of -50.29%. Only one item showed a positive change between pre- and post-test.

A comparison of post-test recall accuracy for related and unrelated items showed little difference, decreasing -27.49% for items from the related list and -26.53% for items from the unrelated list. Examination of long-term recall of abstract versus concrete words also showed little difference, with -27.55% decline for abstract items and -27.06% for concrete items. Overall performance varied somewhat by group (see Table 8).

For some subjects, general English proficiency scores (G-TELP) were available. We examined these for correlation with Test AB related items and Test AB unrelated items. Results (SS = 23) indicated a correlation of $R^2 = -.0003$ between G-TELP with related and $R^2 = -.0018$ with unrelated items.

No tests for significance were carried out due to the reasons addressed below.

Table 8. Short and long-term recall by group

Group	Test A (Related Items)	Test B (Unrelated Items)	Test AB (All Items)	Test AB (Related Items)	Test AB (Unrelated Items)
1	80%	85%	49%	54%	50%
2	76%	73%	50%	55%	47%
3	81%	51%	44%	65%	34%
4	54%	69%	37%	35%	41%
5	84%	90%	65%	56%	74%
6	87%	89%	60%	60%	58%

Discussion

First, the data fails to indicate any clear difference between long-term recall of items from the List A (related) or List B (unrelated). This argues in favor of rejecting our main hypothesis. Second, there was a natural decline in the recall accuracy between the pre-test (Tests A and B) and post-test (Test AB). One anomalous reading was the increase in recall for the word *meditate*. This is most likely due to learning that occurred between the pre-test and post-test phases. Third, general English ability showed no relationship to which type of item (related or unrelated) would be recalled. Fourth, there was some difference in performance between groups that is probably related to specific ways in which teachers directed their students, such as the type of feedback to students before or after the test (i.e., “please study harder next time”).

In light of these results, we reviewed our original procedures and have discovered two serious flaws:

1) As part of the effort to eliminate the effect of teaching on recall of items, we presented related and unrelated items in the same way: alphabetically. As such, students may not have discovered the relationship between items and, hence, not benefited from the reduced mental processing burden related words should have provided.
2) As performance on the post-test can be considered exclusively related to degree of preparation for both of the pre-tests, any unbalance in that preparation will skew the results. We did not have clear protocols for insuring that study for both of the pre-tests was equal.

Conclusion

As a pilot study, the errors we found in design and implementation are insightful into the process of designing viable projects. The next stage of the project, currently underway, has revised the design and implementation in light of what we discovered. Other aspects will be dealt with during the final stage, which will begin in April of 2004.

Of interest, however, is that despite only a short memorization period, students still had fairly good recall after three months. This is encouraging. However, as a record of the amount of time spent on study of the word lists by the students was not kept, it would be difficult to extrapolate this information to other learners.

Research in the type of word list to provide learners for optimal recall is an intriguing and potentially insightful area of research. We hope that our experiences in the design of the pilot study will inform others and help them establish their research projects.

References

Furutou, A. (1995). *Daigaku nyuushi ukeru eitango sokuratesu 2088* [University entrance examination English vocabulary Socrates 2088]. Tokyo: Gakushu Kenkyuukai.

Groot, A. M. B. de, & 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 vocabulary learning and forgetting. *Language Learning, 50*(1), 1-56.

Groot, P. J. M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology, 4*(1), 60-81.

Hayes, N. (2002). *Psychology.* London: Hodder & Stoughton.

JACET (Ed.). (1993). *JACET 4000 basic words.* Tokyo: JACET.

Kazahaya, H. (2003). *Sokudoku eitango joukyuu hen* [Rapid reading English vocabulary upper level edition]. Shizuoka: Zoushinkai Shuppansha.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English.* Edinburgh: Pearson Educational.

Nation, I. S. P. (1990). *Teaching & learning vocabulary.* Boston: Heinle & Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Oubunsha. (1994). *Eiken gokaku no tame no jun'ikyuu hisshutan•jukugo 2100* [Essential vocabulary for passing the STEP Test]. Tokyo: Oubunsha.

Read, J. (2000). *Assessing vocabulary (First ed.).* Cambridge: Cambridge University Press.

Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.