

The Effect of Teaching Vocabulary through Synonymous, Semantically Unrelated, and Hyponym Sets on EFL Learners' Retention

Elaheh Sotoudehnama

Associate Professor, Alzahra University, Iran

Faezeh Soleimanifard

M.A. in TEFL, Alzahra University, Iran

Received: May 12, 2013; **Accepted:** October 29, 2013

Abstract

Many textbooks include semantically related words and sometimes teachers add synonyms, antonyms, etc. to the words in order to present new vocabulary items without questioning the possible effects. This study sought to investigate the effect of teaching vocabulary through synonym, semantically unrelated, and hyponym sets based on Higa's (1963) proposed continuum. A total of 120 Iranian intermediate EFL adults were selected and classified into two high and low language proficient learners based on their PET (2003) scores. They learned the vocabulary items based on the three above-mentioned methods. Learners' vocabulary achievement was measured using Paribakht and Wesche's (1993) Vocabulary Knowledge Scale (VKS) in order to assess both the quantitative (number of learnt vocabulary [NLV]) and the qualitative knowledge of vocabulary (depth of learnt vocabulary [DLV]) by administering the same test twice with a two-week interval for obtaining ST and LT results. To address research questions, two independent two-way ANOVAs and two mixed design two-way ANOVAs were conducted. The results revealed that the learners from synonym sets group gained better ST vocabulary achievement quantitatively and language proficiency level proved not to play any significant role in the learners' vocabulary accomplishment based on belonging to any given group. It was also revealed that quantitatively hyponym, semantically unrelated, and synonym set groups were respectively the most effective methods of clustering that lead to less forgetting in LT which supports Higa's proposition.

Keywords: hyponym set, semantically unrelated set, synonym set, vocabulary knowledge scale, number of learnt vocabulary, depth of learnt vocabulary

INTRODUCTION

Learning and teaching vocabulary has received much attention in the past few decades in the field of language teaching (McKeown, Beck, & Sandora, 2012; Schmitt, 2008). Nevertheless, in spite of a great deal of noticeable advancement, the issue seems to be far from reaching consensus considering how learners can learn vocabulary conveniently or how it can be taught effectively.

One of the first important points regarding vocabulary learning is the decision to be made about which words should be taught, in which order, and how they should be presented (Beck, McKeown, & Kucan, 2013). Webb (2007) highlights that researchers and teachers must be very careful in selecting target words, as “the type of words chosen, and their L2 relationships may determine the size of gains” (p.77). A glance into most of the English language textbooks shows that each unit usually contains related words that the teacher should present in one session and the students should learn them all together. *Top Notch Fundamentals* (Saslow & Ascher, 2006), *Interchange Intro* (Richards, 2005), and *Connect2* (Richards, Barbisan, Sandy, 2004) all provide many vocabularies that are semantically related words in a way.

Scholars who have done researches in this area keep two opposite positions in argument: advocates of semantic relationship between words who believe in the facilitative impact of learning semantically related words simultaneously (Channel, 1981; Dunbar, 1992; Neuner, 1992; all cited in Nation, 2000; Hashemi & Gowdasiaei, 2005) and opponents who highlight the obstructive impact (Finkbeiner & Nicol, 2003; Hakkı Erten & Tekin, 2008; Papathanasiou, 2009; Tinkham, 1997; Waring, 1997).

LITERATURE REVIEW

The Role of Clustering New Words in Learning Vocabulary

There are two opposing hypotheses with respect to learning clustered words together in linguistics. Based on the *interference theory*, “as similarity increases between targeted information and other information learned either before or after the targeted information, the difficulty of learning and remembering the targeted information also increases” (Tinkham, 1993, p. 372). Hunt and Elliot’s (1980, cited in Tinkham, 1997) *distinctiveness hypothesis* also relates ease of learning to the

distinctiveness and non-similarity of the information that are going to be learned.

On the other hand, *semantic activation theory* indicates that words are processed in memory through mental structures called nodes. When a node is activated, activation spreads through mutually connected links to other associated concepts (Aitchison, 2003). Thus, if words in mental lexicon are related in such associative networks as Aitchison (2003) cites, then teaching items in lexical sets would possibly assist learning words.

Thus, although it is crystal clear that the meaning of each word in a language is related to some other words in that language and this sense relation is what is needed for understanding new words; there is still controversy whether new items should be taught through relations such as synonymy, hyponymy, etc. or not.

Higa (1963) studied seven kinds of meaning relationships between pairs of words (near synonyms, free associates, opposites, unrelated, connotation, partial response identity, and coordinate) and developed a continuum starting from near synonyms (e.g., *fast* and *rapid*) that are “most interfering” in his term, having unrelated ones in the middle that are “neutral” (e.g., *bread* and *foot*), and ending in coordinates (e.g., *apple* and *pear*) that are “most helpful”. His continuum shows that teaching lexical sets, at least with some particular meaning relations could have obstructive impacts on learning new vocabulary.

Tinkham (1993) found that learning semantically related groups of words (which directly descend from a common superordinate) together interfered with actual learning of the words. He emphasized that when learners were given a list of words that share a common superordinate, they learned more slowly than words not sharing a superordinate. Waring (1997) who replicated Tinkham’s experimental study also obtained the same results.

Moreover, Finkbeiner and Nicol (2003) investigated learning new words paired with their pictures and found that participants translated L2 words learnt in semantic sets more slowly than those learnt in random order. They suggested “simultaneous activation of semantically related lexical items is at the root of the effect” and concluded that presenting semantically grouped words has a deleterious effect on L2 learners’ learning (p.377). Hakkı Erten and Tekin (2008) also reported that in their study, the test completion time was much longer for the semantically related vocabulary items and concluded that “synonyms, antonyms,

hyponyms, or other such relations among words can cause confusion, and thus require extra time and effort” (p.418).

Those scholars who believe that teaching related items together has facilitative impacts have justified their stance with reference to the following findings: It (1) requires less learning to learn words in a set (Neuner, 1992, cited in Nation, 2000); (2) is easier to retrieve related words from memory; (3) helps learners see how knowledge can be organized (Dunbar, 1992, cited in Nation, 2000); (4) reflects the way such information is stored in the brain; and (5) makes the meaning of words clearer by seeing how they relate to and are different from other words in the set (as cited in Nation, 2000, p.6).

Schmitt and Schmitt (2009) assert that organized material is easier to learn and highlight that a great number of related words can be learned in a quite short time. Carter (1987) argues that since words in lexical grids can be defined in relation to each other, their “fine gradations and differences” with respect to their meaning would be plain in a “very efficient and economic manner” (p.7).

The results of Hashemi and Gowdasiaei’s (2005) research showed that the lexical sets students’ gains in their vocabulary depth (VD) and vocabulary breadth (VB) knowledge were more satisfactory than the semantically unrelated ones. Schneider, Healy, and Bourne’s (1998) study also demonstrated that learning related words together was easier than learning unrelated ones, although the retention test results showed difficulty in recalling them in long-term.

The results of Zheng, Kang, and Kim’s (2009) research also supported that a number of semantic relationships including hypernymy, hyponymy, meronymy, and holonymy are effective methods of clustering that can be applied for better learning and retrieving. Higa’s (1963) and Hoshino’s (2010) researches also showed that coordinates or categorical sets (i.e., hyponyms) are helpful and effective in learning.

The Role of Proficiency Level in Learning Lexical Sets

Nation (2000) emphasizes that “learning new words is a cumulative process, with words being enriched and established as they are met again” and adds that learning related words in sets is not good to be used for initial learning but “as learners’ knowledge becomes more established, seeing related words in sets can have a more positive effect” (p.6). Carter (1998) also believes that using word lists would be useful

for beginners; teaching words in context would be more suitable in upper levels of proficiency; and then word sets and grids would be better for advanced learners.

Based on the results of Papathanasiou's (2009) study, it was supported that the presentation of unrelated vocabulary assists learning new L2 words more than related vocabulary at adult beginners' level. Hence, she concludes that at first, it is better to present unrelated vocabulary and then later, at a more advanced level, present semantically related vocabulary. In Hashemi and Gowdasiaei's (2005) research, too, the upper level students showed greater achievement in their vocabulary depth and vocabulary breadth knowledge than their peer lower level ones. Then based on this result, the writers propose that in L2 vocabulary learning, there is possibly a period like the L1 *vocabulary spurt* in which words are learned more quickly and it "probably begins after the L2 learner has built up an initial vocabulary and has reached a higher language proficiency" (p.356).

Zipoli, Michael, Coyne, and McCoach (2010) also highlight that semantically related reviews can be used to "promote high levels of word learning" (p.12). Therefore, it is stressed that working with a group of related words should be delayed till secondary stage of learning L2, during which the learners are ready for taking advantage of making connections and noticing distinctions between lexical sets.

The Retention of Learned Vocabulary through Lexical Sets

Vocabulary is not learned in a linear manner that is only progressing without any backsliding. Learners usually forget material as well and this forgetting is a natural reality about learning. Memory researches have shown that forgetting occurs within a short time after the learning phase and then its speed reduces gradually (Baddeley, 1990, cited in Schmitt, 2008). Additionally, it is widely suggested that the more cognitive energy a learner spends during learning phase, the more likely that person will be for remembering the learned items whenever needed (Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001; Schmitt & McCarthy, 1997). This idea is shaped based on the *Depth (or Levels) of Processing Hypothesis* that highlights a deeper engagement with words would result in better retention (Schmitt, 2008). McKeown, Beck, and Sandora (2012) also mention that "(1) multiple exposures of the words being taught; (2) breadth of information_ definitional and contextual; and (3) engagement

of active or deep processing by getting students to think about the words and interact with them” are features that are effective for enhancing vocabulary knowledge (p. 18).

Considering the retention aspect of teaching vocabulary through semantically related sets, Hakki Erten and Tekin’s (2008) research supported that presenting new words in related sets interfered with learning and retrieving words. Based on the results of the short term posttest and long term posttest, teaching vocabulary in semantically unrelated sets produced better results than teaching words in semantically related sets and also this difference remained the same in second testing. Schneider et al.’s (1998) research also contributed noticeable findings to this area. They initially found that learning related words together was easier; while the results of long term retention test showed that the participants from unrelated vocabulary group were faster and made fewer errors than those from the related vocabulary group. Therefore, they emphasized that “blocking vocabulary by category...may aid initial acquisition but may not yield optimal retention” (p. 86).

Vocabulary Assessment

Vocabulary skill as one of the priorities in L2 teaching requires tests to assess word knowledge of the learners in order to check their progress and meet their needs. Schmitt (1999) justifies that recently scholars coincided that the measure of the vocabulary size by itself is not an adequate description of vocabulary knowledge and “*how well* individual words are known (*depth of knowledge*)” is also needed to have more complete view of lexical knowledge (p.191).

Paribakht and Wesche’s (1993) *Vocabulary Knowledge Scale* (VKS) demands learners to report on their own knowledge of each word while answering five questions that start from recognizing the word, to being able to make a sentence using that word. In order to assess the learner’s word knowledge through VKS, the target words are presented and then learners are supposed to demonstrate their knowledge responding to five categories shown in Figure 1.

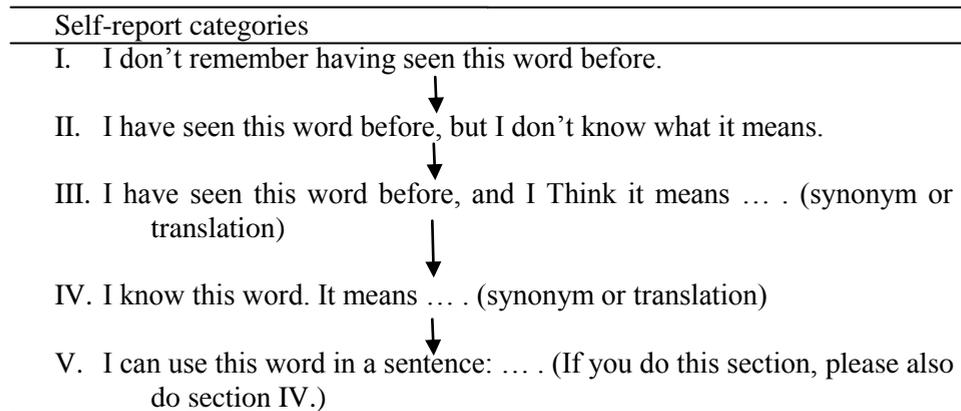


Figure 1: VKS elicitation scale self-report categories (Wesche & Paribakht, 1996, p. 30)

Paribakht and Wesche (1997) mention that vocabulary gains through a course of vocabulary teaching would be both *quantitative* that would be reflected in “the number of words known to some degree versus not known” and *qualitative* that would be obtained from increases in depth of knowledge of given words (p.189). Therefore, two scores based on the learners’ knowledge of the target words would be estimated: (a) *depth of vocabulary* and (b) *known vocabulary*.

PURPOSE OF THE STUDY

So far, the reported background clarifies that there is still lack of consensus among scholars about the advantages and disadvantages of teaching semantically related words. Therefore, this research aimed to obtain the initial (near synonyms), central (unrelated), and final (coordinates) relationships in Higa’s (1963) continuum to study their effects on learning vocabulary adding two new variables: level of proficiency and retention with a different way of assessment from that of Higa’s (1963) i.e., VKS.

The present study concentrated on the effect of teaching words in three different meaning sets, i.e., synonyms, semantically unrelated, and hyponyms on both ST (short-term) and LT(long-term) retention of Iranian high and low proficient EFL learners. Thus, the study intended to answer the following questions:

1. (a) Is there any significant quantitative difference between the short-term vocabulary achievements of high and low language proficient level learners who learn new items through either synonyms, semantically unrelated sets, or hyponyms (as reflected in the number of learnt vocabulary [NLV])?
(b) Is there any significant qualitative difference between the short-term vocabulary achievements of high and low language proficient level learners who learn new items through either synonyms, semantically unrelated sets, or hyponyms (as reflected in the depth of learnt vocabulary [DLV])?
2. (a) Is there any significant quantitative difference between the short-term and long-term retention of the learned items among the learners who learn new items through either synonyms, semantically unrelated sets, or hyponyms (as reflected in the number of learnt vocabulary [NLV])?
(b) Is there any significant qualitative difference between the short-term and long-term retention of the learned items among the learners who learn new items through either synonyms, semantically unrelated sets, or hyponyms (as reflected in the depth of learnt vocabulary [DLV])?

METHOD

Participants

The participants of this research were 120 Iranian learners, ranging in age from 16 to 25 years old who were learning English as a foreign language in Abhar Shokouh Language Institute. They were at intermediate level based on the criteria of the institute and were studying *Top Notch* (Saslow & Ascher, 2006) series at the institute during the summer and autumn of 2010. They were selected from among six intact classes including 143 learners from whom 120 were selected and then divided into two groups of 60 high and low proficient learners based on .05 SD above or below the mean of their PET exam results (appropriate for the participants' level) which the ratio of male to female learners was not controlled in this study. The participants were from six different classes and six teachers taught them the target words.

Instrumentation

Target Words

Forty words—that were all nouns—were selected based on the results of a VKS pretest that separated new words from the already known ones and then they were classified into 10 groups. Each group had one common superordinate (e.g., *shoes*) and four hyponym words that included two synonym pairs (e.g., *sneakers=trainers* and *wellingtons=rubber boots*). As the criterion of relatedness had priority over frequency in this particular study, the lexical sets were checked only for relatedness, not frequency.

Tests

Four tests were administered during this study, two of which were applied before the treatment and the others were given to the learners after that. Initially, the PET test (2003) was administered in order to classify the learners into high and low language proficiency levels. Since the actual level of the participants was intermediate, based on the institute criteria, PET which is a standard test appropriate for an intermediate level was used. Then, a vocabulary pretest was carried out to check whether the words selected to be taught were not known for all participants or not. Another vocabulary test based on Paribakht and Wesche's *Vocabulary Knowledge Scale* (VKS) was administered after teaching 40 words: once within 2 or 3 days after the treatment (ST posttest) and once with 2 week interval (LT posttest) to check the ST and LT effects. This time interval was selected as Mackey and Gass (2005, p. 149) suggest, "Often this is 1 week following the first posttest and then 2 weeks later and even 2 or 3 months later."

Data Collection Procedure

First of all, the PET test (2003) was administered. Since the participants' level, based on the institute's criteria was intermediate, this test was selected. In fact if any other standard test like FCE or TOEFL were selected, due to the participants' level, the normality of distribution could not be achieved. The results showed that there was a normal distribution ($Z=.882$, $p>.05$) and the scores were highly reliable ($\alpha=.91$). Then, 120 learners of all 143 participants of the PET test were chosen: 60 high

proficient learners whose scores fell between 0.5 to 2 standard deviations above the mean score and 60 low proficient learners whose scores ranged from 0.5 to 2 standard deviations below the mean score. Afterwards, both the high and low students were assigned to three different groups in which they were supposed to learn the target words through one of the three different ways of presentation. Therefore, 20 participants from each proficiency level received instruction in one of the three ways of the vocabulary presentation.

All participants in each group took the pretest before the beginning of the instructional phase in order to make sure that they did not know the meaning of the preselected target words in the lists. Based on the results of the pretest, 12 out of 52 words were discarded. The target words were put into three different lesson plans based on the meaning relationships between them to be presented to three different groups of participants (see Appendices A, B, & C). As all 40 words were aimed to be taught in five 45-minute-sessions, each time 8 words were planned to be introduced based on Gairns and Redman's (2006, p.66) suggestion: "It is impossible to be dogmatic about the number of new lexical item that should be presented in a sixty-minute lesson. We would suggest an average of eight to twelve productive items as representing a reasonable input." In this study to assure of learning, the minimum figure suggested, i.e., eight, was considered as the basis. In each session, the synonym groups (high and low) were supposed to learn 8 words that included 4 synonym pairs (e.g., *merchant=retailer*, *receptionist=desk clerk*, *sneakers=trainers*, and *expo=fair*). The hyponym groups had other 8 words that were 4 hyponym pairs (e.g., *negative emotions: woe and fury*, *accused people: culprit and detainee*, *people with extraordinary skills: sorcerer and soothsayer*, and *parts of something: kernel and pinnacle*). And finally semantically unrelated groups received 8 words that did not have any meaning relationships (e.g., *craftsman*, *wellingtons*, *pinnacle*, *bombardment*, *buffoon*, *woe*, *felon*, and *wizard*).

Six intermediate intact classes of Abhar Shokouh Institute, including both male and female learners were selected. The number of participants in these six classes was 22, 25, 24, 25, 23, and 24 which altogether it made 143. Each of the two above- mentioned classes was supposed to be considered for teaching one of the three methods of teaching vocabulary, i.e., semantically unrelated, synonymous and hyponymous sets. In the first pair, including 49 participants, 20 were low, 22 high, and 5 not within the determined range. In the second pair, including 49 again, 23

were low, 21 high, and 4 not within the range and in the third pair including 47, 21 were high, 20 low, and 6 out of the range. Since the minimum number of high and low level participants in two of the pairs was 20, to keep the number of the participants constant in the three sets, the rest of the participants were excluded from the data.

Six teachers taught the target words in these six regular intermediate intact classes. The teachers were all colleagues of the researcher and requested to stick to the lesson plans. The definitions, translations, and examples presented in all classes were the same. After giving the definition and building the meaning relations (except in the case of unrelated sets), the teachers provided some sentence examples for the learners and required them to offer their own examples, as well. After clarifying the meaning of each new vocabulary item, its Persian translation was also provided as an additional help and a check for understanding.

Thus, the semantically unrelated groups learned new items separately and in an unrelated manner through attending to sentence context and clarifying dictionary definition only. Their teachers taught eight semantically unrelated words each session, for example craftsman, wellingtons, pinnacle, bombardment, buffoon, woe, felon, and wizard were presented in the first session through their dictionary definition and two sentences by teacher and several other sentences by learners themselves. After practicing the words through making sentences, the teachers were supposed to elicit translation or provide it themselves in case the learners were unable to suggest correct Persian equivalent. The hyponym groups also followed the same procedure plus associating new items to other words in the list under the same superordinate before starting to practice them through sentence making. And finally, the synonym groups made use of synonymous links in learning synonym words before practicing them in context of sentences.

At the end of the teaching phase, a VKS was administered to all groups two or three days after teaching 40 words. In order to evaluate learners' *depth of learnt vocabulary* and to calculate the *number of learnt vocabulary*, Paribakht and Wesche's (1993) VKS was applied. After two weeks, the same test was administered again in order to check the retention of the target words. The scoring procedure was done based on Wesche and Paribakht's (1996) rubric for interpreting the learners' answers shown in Figure 2.

Self-report categories	possible scores	Meaning of scores
I.	1	The word is not familiar at all.
II.	2	The word is familiar but its meaning is not known.
III.	3	A correct synonym or translation is given.
IV.	4	The word is used with semantic appropriateness in a sentence.
V.	S5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

Figure 2: VKS scoring categories: Meaning of scores (Wesche & Paribakht, 1996, p. 30)

Considering the fact that a learner's *depth of learnt vocabulary* (DLV) score for each word would range from 1 to 5 points and since the total number of the target words in this study was forty, the maximum DLV score for each student in a test was 200 ($40 \times 5 = 200$) and the minimum DLV score was 40 ($40 \times 1 = 40$). Also, in order to measure the students' *number of learnt vocabulary* (NLV), scores 3, 4, and 5 in the DLV were given only one point while scores 1 and 2 received no points this time. Thus, due to the fact that 40 words were assessed in a test, the maximum NLV score could be 40 points for every student ($40 \times 1 = 40$) and the minimum NLV score was 0 ($40 \times 0 = 0$).

Data Analysis

Before running any of the needed analyses for each research question, the necessary assumptions were observed by checking histograms and one-sample K-S for normality of distribution and also box plot for identification of outliers and extremes. Then, to investigate the effect of teaching new vocabulary items through either synonyms, semantically unrelated sets, or hyponyms on vocabulary achievements of learners with high and low language proficiency level, both quantitatively (reflected in NLV scores) and qualitatively (reflected in DLV scores), two independent two-way ANOVAs were conducted. Furthermore, in order to investigate the difference between the ST and LT retention of the

taught items among the language learners who learned new items through either synonyms, semantically unrelated sets, or hyponyms both quantitatively (reflected in the NLV scores) and qualitatively (reflected in the DLV scores), two mixed design two-way ANOVAs were applied.

RESULTS

Results of the First Research Question

The results of the first independent two-way ANOVA (see Table 2) revealed that the type of the vocabulary presentation had a significant effect, $F(2,114)=7.371$, $p=.001$, and the effect size was almost large based on Cohen's (1988) criterion that considers .01=small effect; .06=moderate effect; and .14=large effect (eta squared=.115). Also, the main effect for level reached statistical significance, $F(1,114) = 9.236$, $p=.003$, however the effect size was moderate (eta squared=.075). Moreover, the interaction effect was not significant $F(2,114)=1.531$, $p=.221$.

Table 1: Descriptive statistics for NLV scores in high and low groups

Level	Group	Mean	Std. Deviation	N
High	Synonym sets	32.45	7.323	20
	Hyponym sets	24.25	4.166	20
	Semantically unrelated sets	25.00	8.724	20
	Total	27.23	7.829	60
Low	Synonym sets	25.25	8.650	20
	Hyponym sets	22.85	6.218	20
	Semantically unrelated sets	21.20	8.508	20
	Total	23.10	7.916	60
Total	Synonym sets	28.85	8.711	40
	Hyponym sets	23.55	5.272	40
	Semantically unrelated sets	23.10	8.720	40
	Total	25.17	8.110	120

As the Table indicates, there is difference between the two different levels of language proficiency in NLV. The following Table reveals the degree of this difference.

Table 2: Tests of between-subject effects for NLV scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1500.467 ^a	5	300.093	5.408	.000	.192
Intercept	76003.333	1	76003.333	1.370E3	.000	.923
Level	512.533	1	512.533	9.236	.003	.075
Group	818.067	2	409.033	7.371	.001	.115
Level * Group	169.867	2	84.933	1.531	.221	.026
Error	6326.200	114	55.493			
Total	83830.000	120				
Corrected Total	7826.667	119				

Therefore, it could be concluded that there was a significant difference in quantitative knowledge of learners belonging to three groups in which target words were clustered and presented through synonym, semantically unrelated, and hyponym sets. In order to find the best method of clustering, Tamhane's post hoc test (see Table 3) was conducted due to the result of Levene's test ($p < .05$). The results showed that the participants receiving instruction through synonymous sets outperformed the others. No other significant difference was observed. As far as the level is concerned, the high level outperformed the low level.

Table 3: Tamhane's post hoc test for NLV scores

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Synonym sets	Hyponym sets	5.30*	1.610	.005	1.35	9.25
	Semantically unrelated sets	5.75*	1.949	.013	.99	10.51
Hyponym sets	Synonym sets	-5.30*	1.610	.005	-9.25	-1.35
	Semantically unrelated sets	.45	1.611	.989	-3.50	4.40
Semantically unrelated sets	Synonym sets	-5.75*	1.949	.013	-10.51	-.99
	Hyponym sets	-.45	1.611	.989	-4.40	3.50

The analysis based on the second two-way ANOVA revealed that there was not any significant difference between the DLV scores of the learners from different groups (see Table 5); i.e., synonym, hyponym, and semantically related sets, $F(2,114)=1.219$, $p=.299$, but there was a significant difference for different proficiency levels, $F(1,114)=16.660$, $p=.000$, and the effect size was moderate ($\eta^2=.128$) and no significant difference considering the interaction of the level and group factors, $F(2,114)=.568$; $p=.568$. Table 4 shows the descriptive statistics and the difference between the two levels of proficiency in DLV. In Table 5 the degree of this difference is presented.

Table 4: Descriptive statistics for DLV scores in high and low groups

Level Group	Mean	Std. Deviation	N
High	Synonym sets	155.45	20
	Hyponym sets	141.05	20
	Semantically unrelated sets	146.30	20
	Total	147.60	60
Low	Synonym sets	130.40	20
	Hyponym sets	127.55	20
	Semantically unrelated sets	129.30	20
	Total	129.08	60
Total	Synonym sets	142.92	40
	Hyponym sets	134.30	40
	Semantically unrelated sets	137.80	40
	Total	138.34	120

Thus, teaching new vocabulary items through either synonyms, semantically unrelated sets, or hyponyms did not make any significant difference in ST vocabulary learning of language learners qualitatively, i.e., the depth of learnt vocabulary. Moreover, although all high language proficient level learners gained better achievements, there was not any significant difference between the qualitative vocabulary learning of high and low level learners based on their membership in different treatment groups.

Table 5: Tests of between-subject effects for DLV scores

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	12492.942 ^a	5	2498.588	4.047	.002	.151
Intercept	2296610.008	1	2296610.008	3.720E3	.000	.970
Level	10286.008	1	10286.008	16.660	.000	.128
Group	1505.417	2	752.708	1.219	.299	.021
Level * Group	701.517	2	350.758	.568	.568	.010
Error	70386.050	114	617.421			
Total	2379489.000	120				
Corrected Total	82878.992	119				

Results of the Second Research Question

Based on the results of the first mixed design two-way ANOVA (see Table 7), the effect of time (ST vs. LT) was significant, $F(1,117) = 147.803$, $p=.000$ with a very large effect size (eta squared=.558), and also the interaction of them with different semantic groups the learners belonged to, proved to be significant as well, $F(2,117)=7.105$, $p=.001$ and the effect size was moderate (eta squared=.108).

Table 6: Descriptive statistics for NLV scores comparing ST and LT

	Group	Mean	Std. Deviation	N
STPosttestNLV	Synonym sets	28.85	8.711	40
	Hyponym sets	23.55	5.272	40
	Semantically unrelated sets	23.10	8.720	40
	Total	25.17	8.110	120
LTPosttestNLV	Synonym sets	23.05	7.257	40
	Hyponym sets	21.00	5.487	40
	Semantically unrelated sets	18.28	7.296	40
	Total	20.78	6.959	120

Based on Table (6), there is difference between these short and long term NLV posttests. Next Table is to see if this difference is significant or not.

Table 7: Tests of within-subject effects for NLV comparing ST and LT

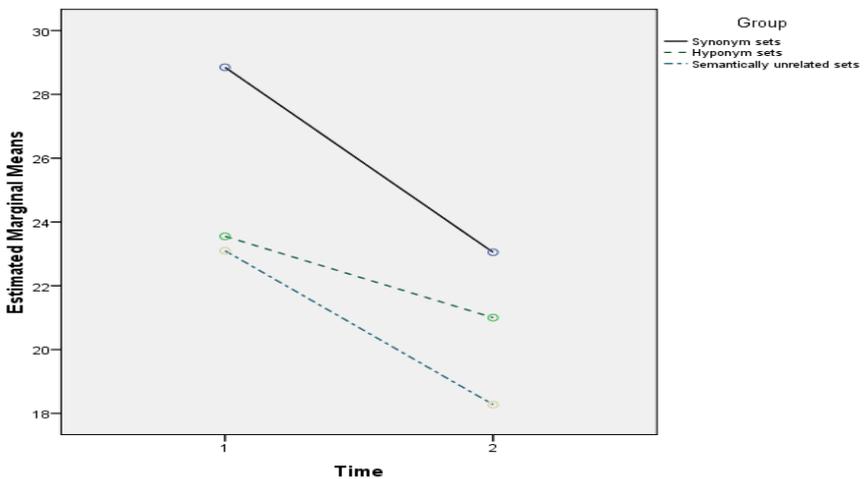
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	1157.204	1	1157.204	147.803	.000	.558
	Greenhouse- Geisser	1157.204	1.000	1157.204	147.803	.000	.558
	Huynh-Feldt	1157.204	1.000	1157.204	147.803	.000	.558
	Lower-bound	1157.204	1.000	1157.204	147.803	.000	.558
Time * Group	Sphericity Assumed	111.258	2	55.629	7.105	.001	.108
	Greenhouse- Geisser	111.258	2.000	55.629	7.105	.001	.108
	Huynh-Feldt	111.258	2.000	55.629	7.105	.001	.108
	Lower-bound	111.258	2.000	55.629	7.105	.001	.108
Error(Time)	Sphericity Assumed	916.038	117	7.829			
	Greenhouse- Geisser	916.038	117.000	7.829			
	Huynh-Feldt	916.038	117.000	7.829			
	Lower-bound	916.038	117.000	7.829			

Due to the result of Levene's Test ($p < .05$), Tamhane post hoc was conducted and as it is shown in Table (8), there was a significant difference between the quantitative scores of learners' ST and LT retention tests in synonym and semantically unrelated sets groups ($p < .05$) and also in synonym and hyponym sets group ($p < .05$). The difference in both cases was in favor of synonym sets group that indicates these learners gained higher scores in LT in comparison with other two groups. On the other hand, there were no significant differences between the scores in hyponym and semantically unrelated sets groups ($p > .05$).

Table 8: Tamhane's post hoc for NLV scores comparing ST and LT

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Synonym sets	Hyponym sets	3.68*	1.450	.040	.13	7.22
	Semantically unrelated sets	5.26*	1.723	.009	1.06	9.47
Hyponym sets	Synonym sets	-3.68*	1.450	.040	-7.22	-.13
	Semantically unrelated sets	1.59	1.495	.645	-2.07	5.25
Semantically unrelated sets	Synonym sets	-5.26*	1.723	.009	-9.47	-1.06
	Hyponym sets	-1.59	1.495	.645	-5.25	2.07

It was also observed that there was less attrition in knowledge of learners who had learned target words through hyponyms and semantically related sets within a two-week interval. Comparing the mean differences of the ST posttest and LT posttest scores, it could be concluded that learners from hyponym sets group lost the smallest amount of their NLV scores while learners from synonym sets group lost the greatest amount of their NLV scores.

**Figure 3:** Estimated marginal means for STPosttest NLV (1) and LTPosttest NLV (2)

The results of the second mixed design two-way ANOVA showed that there is a significant difference between the ST posttest and LT posttest qualitative scores of the learners, $F(1,117)=88.662$, $p=.000$, and the effect size was very large (eta squared=.431). The interaction of ST and LT retention of target words and different semantic groups the learners belonged to also proved to be significant, $F(2,117)=3.378$, $p=.037$, however the effect size was small (eta squared=.055). Finally, there was not any significant difference between the qualitative scores of learners' ST and LT retention tests in any of the groups $F(2,117)=1.076$, $p=.344$ as shown in Table (10).

Considering the means and standard deviations as shown in Table (9), there are differences between the short term and long term DLV posttests. Synonym sets in both short term and long term have the higher means and a higher mean for semantically unrelated sets for short term but a higher mean for hyponym sets in the long term.

Table 9: Descriptive statistics for DLV scores comparing ST and LT

	Group	Mean	Std. Deviation	N
STPosttestDLV	Synonym sets	142.92	27.666	40
	Hyponym sets	134.30	23.292	40
	Semantically unrelated sets	137.80	27.903	40
	Total	138.34	26.391	120
LTPosttestDLV	Synonym sets	132.52	22.967	40
	Hyponym sets	127.55	21.386	40
	Semantically unrelated sets	124.10	23.843	40
	Total	128.06	22.829	120

Table 10: Tests of within-subject effects for DLV scores comparing ST and LT

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	6344.817	1	6344.817	88.662	.000	.431
	Greenhouse-Geisser	6344.817	1.000	6344.817	88.662	.000	.431
	Huynh-Feldt	6344.817	1.000	6344.817	88.662	.000	.431
	Lower-bound	6344.817	1.000	6344.817	88.662	.000	.431

Time * Group	Sphericity Assumed	483.433	2	241.717	3.378	.037	.055
	Greenhouse-Geisser	483.433	2.000	241.717	3.378	.037	.055
	Huynh-Feldt	483.433	2.000	241.717	3.378	.037	.055
	Lower-bound	483.433	2.000	241.717	3.378	.037	.055
Error(Time)	Sphericity Assumed	8372.750	117	71.562			
	Greenhouse-Geisser	8372.750	117.000	71.562			
	Huynh-Feldt	8372.750	117.000	71.562			
	Lower-bound	8372.750	117.000	71.562			

To understand the reason, two one-way ANOVAs were run. The results of the first one indicates that there is no significant difference in the qualitative scores of the learners from three different treatment groups in ST, $F(2,117)=1.082$, $p=.342$. Likewise, the second one proved that there is no significant difference in the qualitative scores of the learners from synonymous, semantically unrelated, and hyponymous sets groups in LT, $F(2,117)=1.386$, $p=.254$. These findings explain why the results of post hoc tests did not show any significance considering different groups of clustering and presenting vocabulary. The main reason for initial significance was *Time* factor not the *Group*. Table 11 provides more information.

Table 11: Pairwise comparison of time factor for DLV Scores comparing ST and LT

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
STPosttestDLV	LTPosttestDLV	10.283*	1.092	.000	8.120	12.446
LTPosttestDLV	STPosttestDLV	-10.283*	1.092	.000	-12.446	-8.120

Hence, although the mean differences were in favor of synonym sets group being compared to other two groups, the statistic was not significant. It can be concluded that there was no significant difference

between the qualitative scores of learners' ST and LT retention tests in synonym, hyponym, and semantically unrelated sets groups after a two-week interval.

DISCUSSION

The results obtained from the first research question highlighted that learning new items through synonym sets resulted in better ST vocabulary learning, thus it was in line with Schmitt and Schmitt's (2009) belief that learning organized material would be easier and a great number of new words could be learned in a quite short time.

The findings from this research question did not support Higa's (1963) continuum by showing that presenting words through synonyms resulted in better ST vocabulary learning; while presenting words through both hyponyms and semantically unrelated sets took the second place without any significant difference with each other. The findings also had some points in common with Hashemi and Gowdasiaei's (2005) and Schneider et al.'s (1998) researches. Synonyms as kind of semantically related sets supported to cause in better ST vocabulary achievement quantitatively; however, hyponyms again as another kind of semantically related sets did not lead to any significant difference in learning words quantitatively being compared with semantically unrelated sets.

Synonym sets' superiority in leading to higher NLV scores in ST showed that the greater semantic relationship between words, the more economic vocabulary learning, at least in ST which was close to Neuner (1992, cited in Nation, 2000) and Carter's (1987) stance in literature. The researchers' informal chat with the learners in synonym sets group revealed that the participants felt excited about learning two synonym new words each time. They believed that it was easier and it could save their time and effort learning them together. Some of the learners believed that synonym pairs were stored together in their mind and whenever thinking about their translation, both words would come to their mind.

The same chat with learners from hyponym sets group showed that they could also remember pair words, but in some cases they were confused about their exact meaning and they needed to think which meaning belonged to which form. This finding supported Nation and Newton's (1997) opinion about the difficulty of learning *hot* and *cold*

together because of mixing their forms and meanings. Learners from semantically unrelated sets group, however, seemed to have normal vocabulary learning experience.

Another finding of the first research question was the fact that teaching new vocabulary items through either synonyms, semantically unrelated sets, or hyponyms did not make significant difference in ST vocabulary learning of language learners qualitatively (DLV). Therefore, although the learners from synonyms group gained higher scores for remembering much more target words, it was not found that they were also significantly superior in their depth of knowledge of those words. The findings of Hashemi and Gowdasiaei's (2005) study however, showed that students belonged to semantically related group gained more satisfactory VB (NLV in this study) and VD (DLV in this study) knowledge, being compared to semantically unrelated ones. There seemed to be differences between the results of this research and their study, because synonym group learners could not gain higher score in DLV like their high NLV scores. The most convincing reason could be the learners' lack of familiarity with VKS. It was the first time in testees' history of language learning that they were evaluated both on how many target words they learned and how well each word was learned. That was the reason they functioned better and gained higher NLV scores that was usual type of checking words' meaning by demanding translation or synonym.

On the other hand, it was obvious from some test papers that some students from synonym sets group could remember the word pairs without remembering the exact meaning and needed knowledge to put them in semantically and grammatically correct sentences. So some learners had obtained 3 points by remembering that *detainee* and *internee* were synonyms just by writing these words in stage 4 of the VKS scale without being able to write a sentence on stage 5 to get the highest score that was expected from them because of their high NLV score. Hence, it was observed that stage 4 which required testees to write a synonym or L1 translation for any given word was not a perfect evaluation of knowing a word by itself. Checking deeper knowledge of target words was possible through stage 5 which required a semantically and grammatically correct sentence including any given word. So clustering and presenting new words through synonyms can improve learning greater number of words in ST but in order to add more depth of

knowledge to this learning activity, something more would be needed to supplement this experience which requires more exposure to the context.

Additionally, the findings of this research question showed that the difference between high and low language proficiency learners' quantitative knowledge of target words was significant. However, the interaction between level and group did not lead to any significant differences in the learners' scores. It can be reasoned that better NLV and DLV scores of the high groups in this study could be as a result of the 'vocabulary spurt' (Hashemi & Gowdasiaei, 2005) which led to easier vocabulary learning for more proficient learners.

As the learners from both levels were adults, reasons like having different motivations for learning with more ease that Papatthaniou (2009) discusses could not be justified by this study. Nation (2000) who points to the cumulative nature of learning new words with words being enriched as they are met again, emphasized that learning related words in sets is not good to be used for initial learning but "as learners' knowledge becomes more established, seeing related words in sets can have a more positive effect" (p.6). Additionally, Carter (1998) also believes that using word sets and grids would be better for advanced learners. These two points could be in line with the results of this study considering the superiority of high proficient learners to the low proficient ones.

The findings from second research question showed that both NLV and DLV scores of the learners in the ST posttest were higher than the LT posttest with a two-week interval for all groups. Moreover, as the learners were not aware of the second posttest, it actually showed the true amount of retained words after forgetting some within two weeks. As Baddeley (1990, cited in Schmitt, 2008) states, forgetting occurs within a short time after the learning phase. Likewise, Schmitt (2000) has pointed out that partially learned vocabulary would probably be forgotten if the learner failed to fix them in mind by deliberate repeated exposure and principled recycling. Therefore, the finding that both NKV and DKV scores were decreased in LT posttest was because of natural forgetting that resulted from not recycling the words during the interval between the two posttests. The results of this research question also confirmed Schmitt and Schmitt's (2009) warning that ignoring recycling would cause forgetting many partially-known words and wasting all the effort already put into learning them.

The findings of this study manifested that learners from synonym sets group gained higher NLV and DLV scores both in their ST posttest and

LT posttest. It was also observed that learners from synonym sets group lost the greatest scores within a two-week interval comparing their ST versus LT retention but this fact did not clear out the superiority for this group in general. Learners from hyponym sets group, however, demonstrated greater retention both quantitatively and qualitatively.

The finding that the lowest forgetting has been found in the case of the learners' NLV and DLV scores from hyponym sets group could be explained based on Hulstijn and Laufer's (2001) key for retention in which learners would be more likely to retain words by paying careful attention to more word properties than only one or two of them. Learners in hyponyms group had to process the lexical sets more deeply by learning two pronunciations, forms, meanings, semantic relationships with each other and the shared superordinate, and also the differences that made them two different words and the similarities that placed them under the same head word or phrase. Thus, the kind of mental processing in the case of learning hyponym sets seemed to be greater than learning synonyms which only helped saving learning time and effort by having two words for one meaning. The same was true about learning semantically unrelated sets that were presented and learned without any semantic relations, so the learners required more processing and paying attention that led to lower forgetting in comparison with synonym sets.

The findings of this research question were not similar to Hakkı Erten and Tekin's (2008) research that showed presenting new words in related sets interfered with learning and retrieving words. Although the synonyms group functioned better in the first posttest and the hyponym group had the lowest forgetting in the second posttest, the semantically unrelated sets group could not take the first place on any of the above mentioned tests.

In the case of the learners' scores from the synonym sets, the results seemed to be similar to Schneider et al.'s (1998) study by causing easier ST learning and greater forgetting in LT, but the difference is here that learners from the synonym sets group in this study still gained better scores in LT posttest compared with those from semantically unrelated sets group. It also supported Bunker's (1988, cited in Hatch & Brown, 1995) tip that synonym words should not be learned with each other is not incorrect at least whenever LT retention is necessary. The findings indicated that recycling the newly learned words would change learning synonym sets to a better method for clustering and learning new words.

Although the findings from the first research question had not anything in common with Higa's continuum, the findings from the second research question demonstrated great similarities with this continuum but in LT retention tests. Higa's continuum was produced based on quantitative scores and comparing the findings from the NLV scores of the learners in this study, it was shown that hyponym, semantically unrelated, and synonym sets groups were the most effective methods of clustering leading to less forgetting from left to right. So hyponyms were the most useful in LT test considering less attrition and this finding was exactly in line with Higa (1963) and Hoshino's (2010) research in which teaching new words through coordinates and categorical lists proved to be rewarding for the learners. Thus, there was less attrition in quantitative knowledge of learners who had learned target words through hyponyms and semantically related sets within a two-week interval.

Finally, one may conclude that NLV is a better criterion for checking the difference in learning vocabulary than DLV. However, lower DLV scores of the learners of this study in comparison with their NLV scores could be because of their lack of familiarity with this kind of testing or the insufficient work on developing sentences with target words in the class.

CONCLUSION AND IMPLICATIONS

The present study yielded the following results in terms of vocabulary learning and teaching process. Considering NLV scores, the learners gained better ST vocabulary achievement being taught in classes in which new items were presented through synonym sets than the other two clustering methods (hyponym and semantically unrelated sets) that brought about similar results.

The findings also indicated that language proficiency played a significant role in learning target words considering both NLV and DLV scores. Thus, high proficient learners from all groups having different semantic relations outperformed the low proficient ones. However, the interaction between language proficiency level and different grouping based on the methods of clustering and presenting new words did not lead to any significant difference in learners' NLV and DLV scores.

Moreover, although the interaction of ST and LT retention of target words and different semantic groups which the learners belonged to

found not to be significant based on their DLV scores, there was a significant difference between the NLV scores of learners' ST and LT retention tests in synonym sets group and semantically unrelated and hyponym sets groups. In fact, although there was greater loss in learners' scores comparing their ST with LT scores, still the NLV mean score of the learners from synonym sets group was the highest. On the other hand, there were no significant differences between the NLV scores in synonym and hyponym sets groups and also hyponym and semantically unrelated sets groups.

Another finding of this study was the difference in the amount of forgetting among the learners from synonym, hyponym, and semantically unrelated sets group. It was found out that the learners from hyponym sets group demonstrated greater retention both quantitatively and qualitatively. Learners from semantically unrelated sets group showed less forgetting in their NLV scores while learners from synonym sets group had less forgetting in their DLV scores to take the second position. Also, the greatest forgetting happened to synonym sets group by noticeable decrease in their NLV scores, and semantically unrelated sets group by getting lower DLV scores.

Considering the impact of teaching vocabulary based on the three stages of Higa's (1963) continuum including the initial (synonyms), central (unrelated sets), and final (hyponyms) levels, the findings indicated that language proficiency level caused differences in NLV and DLV scores of the learners so that high proficient learners in all groups outperformed their low peers. However, language proficiency level did not play any significant role in the learners' vocabulary achievement based on belonging to any given group. It was shown that the results of ST testing did not support Higa's continuum while the findings that supported the existence of the difference in the amount of the learners' retention in different groups after two weeks brought the continuum to the foreground. The learners' NLV scores in hyponyms group showed less forgetting and that was the reason hyponyms were called the most useful semantic relation (by Higa) for learning words and retaining them in the long run. NLV scores from semantically unrelated group took the second place as called neutral in Higa's continuum. And finally, the greatest loss in NLV scores was found through the administration of the same test to the learners from synonyms group after two weeks. The results manifested that synonyms were not as useful over time as they

were in ST and that could be called the reason Higa listed them as the most interfering semantic relation in his continuum.

Thus, on the basis of this study, L2 syllabus designers and textbook writers who select and order words that should be presented in different courses and classes can cluster words under the shared superordinate. They can also avoid synonym sets to be presented and learned together, as most of the books are not meant to teach vocabulary for a short time that would be forgotten easily. As all of the clustering and presenting methods showed not to function differently with high or low proficient level learners, using semantically related sets should not be postponed for higher levels.

Teachers can preferably present the new words through hyponym sets with emphasizing on the semantic relations and the superordinates. The learners who are going to study and learn the vocabulary independently can also use the results of this study, using the books which are organized in this way. The learners can cluster the words into synonym sets if they are supposed to learn a great number of words in a short time and for a special test in the near future. Yet, whenever they wanted to learn words aiming to keep them in mind for a longer time, they should utilize hyponym sets and pay more attention to the semantic relations.

Bio-data

Elaheh Sotoudehnama holds a Ph.D. in TEFL and has been teaching for 20 years in the Department of English Language at Alzahra University in Iran. Her area of interest is teaching skills in general and reading strategies, culture, and motivation in particular.

Faezeh Soleimanifard holds an M.A. in TEFL from Alzahra University and is an English language teacher and TTC trainer in Zanjan Jihad Daneshgahi Department of Foreign Languages and also teaches in Payam-e Noor University of Zanjan. She is interested in teaching skills, metacognition, and discourse analysis.

References

- Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon* (3rd ed.). Oxford: Blackwell.
- Beck, I., McKeown, M., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). New York: Guilford Press

- Carter, R. (1987). Vocabulary and second/foreign language teaching. *Language Teaching*, 20(1), 3-16.
- Carter, R. (1998). *Vocabulary: Applied linguistics perspectives*. London: Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24(3), 369-383.
- Gairns, R., & Redman, S. (2006). *Working with words: A guide to teaching and learning vocabulary*. Cambridge: Cambridge University Press.
- Hakki Erten, I., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36(3), 407-422.
- Hashemi, M. R., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: Lexical-set versus semantically-unrelated vocabulary instruction. *RELC Journal*, 36(3), 341-361.
- Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education*. Cambridge: Cambridge University Press.
- Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 2(2), 170-175. doi:10.1016/S0022-5371(63)80082-1
- Hoshino, Y. (2010). The categorical facilitation of L2 vocabulary learning in a classroom setting. *RELC Journal*, 41(3), 301-312.
- Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 538-558.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. London: Lawrence Erlbaum.
- McKeown, M., Beck, I., & Sandora, C. (2012). Direct and rich vocabulary instruction needs to start early. In E. Kame'enui & J. Baumann (Eds.), *Vocabulary Instruction: Research to Practice* (2nd ed.) (pp. 17-33). New York: Guilford Press.
- Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, 9(2), 6-10.
- Nation, P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 238-254). Cambridge: Cambridge University Press.
- Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT Journal*, 63(4), 313-322.

- Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 9-29. Retrieved from <http://www.teslcanadajournal.ca/index.php/tesl/article/viewFile/623/454>
- Richards, J. C. (2005). *Interchange intro: Student's book* (3rded.). Cambridge: Cambridge University Press.
- Richards, J. C., Barbisan, C., & Sandy, C. (2004). *Connect 2: Student's book*. Cambridge: Cambridge University Press.
- Saslow, J., & Ascher, A. (2006). *Top notch fundamentals*. White Plains, NY: Pearson Longman.
- Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, 16(2), 189-216.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Learning Research*, 12(3), 329-363.
- Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N., & Schmitt, D. (2009). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. In P. Hedge, N. Andon, & M. Dewey (Eds.), *English language teaching: Major themes in education* (Vol.4, pp. 385-397). London: Routledge.
- Schneider, V. I., Healy, A. F., & Bourne, L. E. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 77-90). Mahwah, NJ: Lawrence Erlbaum.
- Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, 21(3), 371-380.
- Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, 13(2), 138-163.
- Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, 25(2), 261-274.
- Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, 11(1), 63-81.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.

- Zheng, H., Kang, B., & Kim, H. (2009). Exploiting noun phrases and semantic relationships for text document clustering. *Information Sciences*, *179*(13), 2249-2262.
- Zipoli, R. P., Jr., Coyne, D. M., & McCoach, D. B. (2010). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education*, *31*(4), 1-13.