The Effect of Semantic and Thematic Clustering of Words on Iranians Vocabulary Learning

Forough Mirjalili
Imam Javad University College, Yazd, Iran
M.A. in TEFL

Ali Akbar Jabbari
Associate prof. Ph.D. in Applied Linguistics
English Department
Yazd Uni. Yazd
Iran

Mohammad Javad Rezai
Assistant Prof. Ph.D. in Applied Linguistics
English Department
Yazd Uni. Yazd
Iran

Abstract
Finding an effective method for vocabulary learning has always preoccupied curriculum developers in general and language teachers in particular. The present study attempts to examine the effects of semantic, thematic and unrelated clustering of words as well as the effects of two instructional approaches, i.e. isolation and context on vocabulary learning. The subjects of the study constituted 90 learners in three proficiency levels, i.e. elementary (1), pre-intermediate (2) and intermediate (3). Each cluster was presented in both isolation and context. Tests were given immediately after each exposure. A three-way ANOVA was used to analyze the data inferentially. The results demonstrated that when the words were presented to the subjects in isolation, they generally recalled the highest number of words from the unrelated set whereas they could recall the highest number of words from the thematic clustering in context. Additionally, the effect of proficiency level did not turn out to be significant.

Definition of key words:
Semantic set: words which share the same semantic and syntactic characteristics, grouped under a common concept (e.g. flower names).
Thematic set: words that are grouped together based on a shared thematic concept (e.g. fishing) and are cognitively associated. They may not share the same syntactic characteristics.
Unrelated set: words which do not share similar semantic and syntactic characteristics and are not associated in any concept.

1. Introduction
Vocabulary is a core component of language proficiency and provides much of the basis for how well learners speak, listen, read and write. Unfortunately, vocabulary teaching and learning were neglected in the past (Renandya & Richards 2002:255). During the active decades of the mid – twentieth century, vocabulary building was not a priority for researchers or curriculum designers in the context of language teaching and learning. Recently, there has been a renewed interest in the nature of vocabulary and its role in learning a language. More emphasis and considerable attention have been paid to vocabulary learning. Various techniques have been introduced and used for teaching vocabulary as a result of which researchers have started testing and evaluating these techniques. Thematic and semantic clustering were among these strategies proposed by educational researchers and psychologists. Currently new vocabulary items are typically presented to ESL / EFL students in semantically and thematically related sets in the current ESL (English as Second Language) textbooks. Concerning semantic clustering, a growing body of research indicates that this widely accepted way of presenting new vocabulary items does not facilitate learning (Tinkham’ 1993; Altarriba and Mathis, 1997; Waring, 1997; Finkbeiner and Nicole, 2003).
Rather, it makes learning more difficult and interferes with the learning of similar words. Evidence in support of concepts such as the Interference Theory and the Distinctive Hypothesis, to be discussed further, strongly suggests that semantic clustering may actually impede rather than facilitate learning. Words can also be clustered thematically. This type of clustering is based on a cognitive relationship between words related to an existing schema in the mind.

The terms such as: eye, nose, mouth and ear provide an example of semantic cluster. A cluster perceived as thematically related would include terms like frog, pond, swim and green (Tinkham 1994). These terms do not refer to semantically similar concepts; however, they cluster around the concept of frog.

The present study endeavors to examine the effect of semantic, thematic and unrelated clustering of words on vocabulary learning on Iranian EFL learners. It also aims to examine the effect of different instructional approaches such as presenting words in isolation and context. The research questions investigated in this study are as follows:

1. Which clustering type (semantic, thematic and unrelated) is more effective in vocabulary learning?
2. Which instructional approach (isolation / context) is more effective in vocabulary learning?
3. What is the effect of proficiency level on vocabulary learning?

2. Background

Reviewing the studies carried out in the past, we come across justifications for using semantic clusters. Seal (1991), the author of American Vocabulary Builder, (1) provides two reasons for his use of semantic clusters. First, he says that they give students the sense of structure they need and second, this organization may help students guess the meaning of new words within the lexical set.

Other arguments that support the use of semantic clustering are pragmatic arguments (Tinkham, 1994). Gairns and Redman (1986) assert that semantic clusters help learners to understand the semantic boundaries; to see where meaning overlaps and learn the limits of use of an item’ (P.32). Hence, semantic clustering helps students to distinguish between semantically related words and it also gives coherence to the lesson. They also believe that semantic clusters form building blocks and can be expanded as students’ progress. It also provides a clear context for practice.

An argument against semantic clustering is related to Interference Theory which was formulated by McGeoch (1942). It is hypothesized that “as similarity increases between targeted information and other information learned either before or after the targeted information the difficulty of learning and remembering the target information also increases”(Tinkham,1993:37). It can be evoked to argue that presenting L2 learners with vocabulary items grouped in semantic clusters actually impedes vocabulary learning rather than acting as a support to learning. It refers to the decrease in retention because of a learning activity that interpolates between original learning and later recall. The theory’s hypothesis is that new knowledge loss or retention is influenced by the nature of subsequently acquired knowledge.

Another piece of evidence against semantic clustering is the distinctiveness hypothesis (Eyseck, 1979), which received considerable attention during the 1980s. This hypothesis considers the ease with which distinctive information is learned. It relates ease of learning to the distinctiveness (non-similarity) of the information to be learnt (Tinkham 1993). The claim is that people remember distinct items better than they remember those that are indistinct. Tinkham hypothesized that new word learning would be greater if the words learned are unrelated.

Concerning thematic clustering, experts believe that speakers organize words as frames or schemas in their mental lexicon. This is done with reference to the speakers’ background knowledge rather than in semantic fields (Fillmore, 1985). These clusters are cognitively rather than linguistically derived. In thematic clustering there is an association concept. According to Fillmore (1985), speakers can be said to know the meaning of the word only by first understanding the background frames. So, words are related by their links to a common background. A theory that is in favor of thematic clustering is schema theory. This theory explains how old information possessed by the learner influences the learning of new information. It aims to explain the way different types of knowledge are learned and people’s interpretation of the world from a psychological perspective Schema is an active organization of past reactions or experiences.
Tinkham (1993) studied the results of learning words in semantic clusters. He explored whether new word learning would be greater if the words learned were unrelated or related. He provided subjects with lists of semantically related (shirt, jacket, sweater) and unrelated words (rain, ear, frog) associated with artificial words. Subjects were tested by recall trials-to-criterion tests. Tinkham found that subjects learned the semantically related English words more slowly and with more difficulty than they learned the unrelated English words.

A replication of Tinkham’s (1993) study was carried out by Waring (1997). Japanese students participated in activities in which they were presented with pairs of words: the original words Tinkham used, translated into Japanese, and an artificial word created under the same guidelines as in Tinkham’s study. In two experiments, Waring provided native-speaking Japanese subjects with six Japanese word-pairs, including three semantically related words sharing a common concept of “clothes” and three unrelated words as stimuli. Responses were Japanese artificial words. In the second experiment, subjects were required to learn two separate sets of six Japanese word-pairs; semantically related words (types of fruit) and semantically unrelated words (such as mountain, television, sky, mouse). Results of trials-to-criterion showed that subjects learned the related word-pairs more slowly than they learned the unrelated word-pairs and that “presenting new words that share a common super-ordinate in a set of words to learn does interfere with learning” (p. 267).

Tinkham (1997) took the topic of semantic clustering one step further and attempted to find a way of clustering words that would enhance vocabulary learning. While semantic clusters such as eye, hand and ear are linguistically derived, thematic clusters such as frog, green, and hop are cognitively derived and may facilitate, or make easier, the learning of new words. The questions of the study were to find out whether semantic or thematic clustering made learning words easier than unrelated groups of words. In the first experiment, the subjects were presented with two six-pair sets of associate pairs. The first list included three semantically related English words and three unrelated English words paired with artificial words. The second list included three thematically associated English words and three unassociated English words paired with artificial words. Subjects studied the lists in two modalities (oral and written) and were given recall and recognition tests. In the second experiment, there were four six-pair lists: semantically related, semantically unrelated, thematically associated, and thematically unassociated words paired with artificial words. Tinkham’s results showed that semantic clustering of L2 vocabulary items was a detriment to learning, while thematic clustering was a facilitator. Moreover, artificial words paired with semantically related words were learned with more difficulty than those paired with thematically related words. On the other hand, artificial words paired with thematically related words were learned more easily than those paired with unassociated words.

3. Methodology

3.1 Participants

Data was collected from 90 participants studying in Kish Language Institute in Yazd. They were 30 elementary, 30 pre-intermediate and 30 intermediate students. All the students were selected randomly from an age group of 18 to 25 and they were all females.

3.2 Materials

The purpose of the study was to compare the learnability of semantic, non-related, and thematic sets of associated pairs composed of English and Persian words. Moreover, the effect of using context in learning words was examined by comparing the learnability of the three groups of words in context as well as in isolation.

Six lists were used, each consisting of eight English words. The first list was related by semantic clustering and included eight English (L2) words for the types of flowers. These words were accompanied by Persian (L1) equivalents (See appendix I). The second list had eight English words with Persian equivalents and they were categorized under the theme of fishing-related terms. They were thematic clusters (See appendix II). The third list contained unrelated words and included eight English words accompanied by Persian equivalents (See appendix III). Students were exposed to the words in the first, second and the third group in isolation.

The second set of words in lists four, five and six was presented in context. The fourth list was related by semantic clustering and included eight English words for garden and gardening. (See appendix IV). The fifth list contained English words with Persian equivalents and they were a thematic cluster under the heading of “moose”. (See appendix V). The last list covered eight unrelated words presented in a context (See appendix VI). All the words both in isolation and in context were accompanied by their Persian equivalents.
Every attempt was made to ensure the resulting lists were equivalent in structure and overall level of difficulty. In order to make sure that the students were not familiar with the words, a test was given before carrying out the study. Words that were intended to be used in the study were given to the students and they were asked to write the Persian equivalents. Any word known by the subjects taking part in the study was set aside for the main study. When the words were in context four minutes was added to this time to allow for the extra cognitive effort needed to understand the context. Although they were familiar with the Persian words, they had not established connections between the English words and their equivalents.

The words in the thematic clusters and in the unrelated sets were varied, and included nouns, verbs, and adjectives; however, given the nature of semantic classes, word class was necessarily held constant in the semantic set, with all eight terms being nouns.

3.3 Procedure

As mentioned before 30 students in three levels of proficiency took part in 6 sessions in the main study. Semantic clusters, thematic clusters and non-related words were presented to the subjects in context and isolation. Each phase was done in a session, adding to six sessions in all. In each session, they were given a handout of two pages. The first page contained the list of words or a context in which the words were included. All the words were accompanied by their Persian equivalents. Subjects were required to study them. For the words in isolation, the subjects had a total of four minutes that is 30 seconds per each item. After four minutes, the participants were instructed to stop referring to the lists, and the immediate recall phase took place. Each subject was required to turn to the following page, which contained the eight English words, and to provide their Persian equivalents. The teacher’s role was limited with no teaching of vocabulary. The subjects were only exposed to the words. In order to eliminate any chance of memorizing the lists as whole rather than learning words, the words were arranged in a different sequence. There was no time limit for this recall test. The average time taken by students was three minutes. The same procedure was used in all clustering types, i.e. semantic, thematic and unrelated for all proficiency levels.

For the clusters presented in context, the subjects were given a text which contained a set of underlined words. They were instructed to read the text and to learn the meaning of the underlined words in eight minutes. The time was adjusted to allow for the extra cognitive effort involved in consulting and integrating the brief story context. In the context recall test, the subjects were required to provide the Persian equivalents of the English words. With the context condition, again there was no time limit for completion of the test.

4. Data Analysis

An analysis of variance three way ANOVA was used to analyze the data collected. Three types of clusterings were included in the study. They were semantic, thematic and unrelated. There were two methods used in the study which were isolation and context. There proficiency levels - elementary, pre-intermediate and intermediate – were also taken into account. These three levels are called level 1, level 2 and level 3 through out the study. According to the mentioned variables in the study a 3 (type of clustering: semantic / thematic/ unrelated) by 2 (methods: isolation / context) by 3 (level: elementary /pre-intermediate/ intermediate) repeated measure factorial design constituted the ANOVA design.

In this study the dependent variable is the vocabulary test scores, while the independent variables are types of clustering, levels of students and methods. Measurement of the dependant variable covers the number of recalled words... Furthermore, post-hoc tests will be carried out for the pair wise comparison of the results. The interaction between cluster, level and method will also be considered in the present study.

4.1 Results

4.1.1 Isolation

In this method, the words were presented to subjects without any semantic or thematic link between the words. The results are as follows.

4.1.1.1 Level 1

1. The subjects recalled more words from the thematic (6.80) and unrelated list (6.80). The least belonged to the semantic list (5.53).
2. The differences were statistically significant between the mean of the semantic list on one hand and the means of the other lists on the other hand.

4.1.1.2 Level 2
Level 2 subjects recalled slightly more words from the unrelated list (6.93) than from thematic (6.90). The words recalled from the unrelated and thematic lists were higher than the semantic (6.16). The differences, however, were not statistically significant.

4.1.1.3 Level 3
1. The subjects recalled more words from the unrelated list (7.30) than the thematic (7.06) and semantic (5.90).
2. The difference between the mean of the semantic list on the one hand and each of the other means are statistically significant.
3. All subjects, regardless of their level, recalled more words when the words were presented to them in isolation.

4.2 Clustering types in isolation
4.2.1 Semantic clustering
Studying the clustering types in isolation, the following findings were found regarding the participants’ recall of the semantic list:
1. Level 2 subjects, recalled more semantically listed words (6.16) than those of level 3 (5.90) and level 1 (5.53).
2. The difference between the means was not statistically significant.

4.2.2 Thematic clustering
The following are the major finding from the tests on recalling thematic words:
1. Level 3 subjects recalled the most number of thematic words (7.06) than level 2 (6.90) and level 1 (6.80).
2. The differences are not significant in the above results.

4.2.3 Unrelated clustering
When it comes to the unrelated words, these are the major finding:
Level 3 subjects recalled more unrelated words (7.30) than level 2 (6.93) and level 1 (6.80). Nonetheless the difference between them is not statistically significant.

4.3 Context
4.3.1 Level 1
The subjects recalled significantly more words from the thematic list (6.80) than the unrelated (4.96) and the semantic list (4.66).

4.3.2 Level 2
Similar to the results of level 1 subjects, level 2 subjects recalled more words from the thematic list (6.93) than the unrelated (6.43) and semantic (5.70). The difference between the semantic and the other two lists was significant.

4.3.3 Level 3
The subjects recalled more words from the thematic list (mean 6.93) than the unrelated (6.53) and semantic (5.80). The difference between the semantic list and the two other groups was significant.

4.4 Clustering types in context
4.4.1 Semantic clustering
Studying within condition results, the following findings were found regarding the subjects’ recall of the semantic list:
Level 3 subjects in the immediate test recalled more semantically listed words (5.80) than those of level 2 (5.70) and level 1 (4.66). The difference between level 1 and the other two groups was significant (p<0.05).

4.4.2 Thematic clustering
The followings are the major findings from the tests on recalling thematic words.
Level 2 and 3 subjects recalled the highest number of thematic words (6.93). Level 1 subject recalled the least (6.80). However, the difference between them was not statistically significant.
4.4.3 Unrelated clustering

When it comes to the unrelated words in context, these are the major findings:
Level 3 subjects recalled more unrelated words (6.53) than level 2 (6.43) and level 1 subjects (4.96). The difference between the means of level 2 and level 3 subjects on one side and level 2 and level 1 subject on the other side was significant.

The above results are summarized in the following table:

Table 1: Descriptive statistics of recalled words [Three levels – Three conditions – Two methods]
5. Discussion

The results indicate that grouping vocabulary items sharing semantic and syntactic characteristics impedes learning and has a negative effect on vocabulary acquisition. The words from the semantic lists were the least to be recalled in all cases and the marks for semantic clusters were significantly lower in all cases except for two. The fact that the semantic words were the least to be recalled is not surprising to those who have done research on the interference theory and distinctive hypotheses. According to the interference theory, as similarity increases between target information and other information learned either before or after the targeted information the difficulty of learning and remembering the target information also increases (Tinkham 1993). It can be evoked to argue that presenting L2 learners with vocabulary items grouped in semantic clusters actually impedes vocabulary learning rather than acting as a supporting implement for learning. It refers to the decrease in retention because of a learning activity that interpolates between original learning and later recall.

The second theory in line with the results is the distinctive hypotheses. This hypothesis considers the ease with which distinctive information is learned. It relates ease of learning to the distinctiveness (non-similarity) of the information to be learnt (Tinkham 1993). The claim is that people remember distinct items better than they remember those that are non distinct. Tinkham hypothesized that new word learning would be greater if the words learned were unrelated. When the words were presented to the subjects in isolation, they recalled the highest number of words from the unrelated clusters in all cases. The results are in line with the interference theory and distinctive hypotheses discussed above. This is also in line with Warring’s (1997) suggestion that in order to completely counteract the interference theory, words should be presented in totally unrelated sets.

When the words were presented to the subjects in context, they recalled the highest number of words from the thematic clusters compared to the semantic and unrelated. The results support the findings of schema theory came up with. Schema theory aims to explain how old information possessed by the learner influences the learning of new information. It aims to explain the way different types of knowledge are learned and people’s interpretation of the world from a psychological perspective. Schema is an active organization of past reactions or experiences.

As it was demonstrated in the result’s section, the subjects had a better performance when exposed to thematic clustering in contrast to the semantic clustering of words. All the researches came up with this conclusion that presenting new words that share a common super-ordinate in a set of words to learn interferes with learning (semantic clustering) while the words that share a common scheme existing in the mind (thematic clustering) are easier to learn and remember. Results showed that L2 learners performed better when the words were presented to them in isolation than in context. The difference between the mean marks gained in isolation was statistically significant in favor of the isolation method.

6. Conclusion

The current study has investigated the role of semantic and thematic clustering in learning lexical items under two instructional approaches of isolation and context for different proficiency levels. The purpose was to find out which clustering type and instructional method can offer a more efficient alternative in acquiring vocabulary. The findings of the empirical study are presented below in terms of the research questions posed.

The study conducted showed that when the words were presented to the subjects in isolation the subjects recalled more words from the unrelated clusters. On the other hand, when the words were presented to the subjects in context, they recalled more words from the thematic clusters. The semantic clustering of words was the least efficient type of clustering to be recalled in both methods.

The subjects, when exposed to the words in isolation, performed more efficiently compared with the words presented in context. The difference between the mean scores gained from the isolation and context method was significant in favor of the isolation method.

Last but not the least, there was no effect of proficiency level in this study since in most cases there was no regular trend in the way the subjects in the three levels (i.e. elementary, pre-intermediate and intermediate learners) performed on the tests. Such a finding implies that the thematic clustering or unrelated clustering in appropriate methods is not merely a for certain proficiency levels; rather the domain of generalization can be extended to different groups of learners.
### Appendix I

**Condition 1: semantic list (isolation)**

<table>
<thead>
<tr>
<th>Target word</th>
<th>L2 equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pansy</td>
<td>زغفران</td>
</tr>
<tr>
<td>Crocus</td>
<td>زنبق</td>
</tr>
<tr>
<td>Iris</td>
<td>زنبق</td>
</tr>
<tr>
<td>Aster</td>
<td>منا</td>
</tr>
<tr>
<td>Lily</td>
<td>سوسن</td>
</tr>
<tr>
<td>Gardenia</td>
<td>ياسمن</td>
</tr>
<tr>
<td>Petunia</td>
<td>اطلسباز</td>
</tr>
<tr>
<td>Geranium</td>
<td>شمعداني</td>
</tr>
</tbody>
</table>

### Appendix II

**Condition 2: Thematic list (isolation)**

<table>
<thead>
<tr>
<th>Target word</th>
<th>L2 equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leister</td>
<td>نیزه ماهیگیری</td>
</tr>
<tr>
<td>Reel</td>
<td>ققره</td>
</tr>
<tr>
<td>Dory</td>
<td>قابق ماهیگیری</td>
</tr>
<tr>
<td>Lure</td>
<td>طمعه</td>
</tr>
<tr>
<td>Cast</td>
<td>انداختن قلاب</td>
</tr>
<tr>
<td>Shoal</td>
<td>انبوه ماهی</td>
</tr>
<tr>
<td>Angling</td>
<td>ماهی گرفتن</td>
</tr>
<tr>
<td>Piscator</td>
<td>ماهی گیر</td>
</tr>
</tbody>
</table>

### Appendix III

**Condition 3: Unrelated list (isolation)**

<table>
<thead>
<tr>
<th>Target word</th>
<th>L2 equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jay</td>
<td>زاغ سیاه</td>
</tr>
<tr>
<td>Wrench</td>
<td>آجر</td>
</tr>
<tr>
<td>Beet</td>
<td>جنگرد</td>
</tr>
<tr>
<td>Bailiff</td>
<td>نکهبان</td>
</tr>
<tr>
<td>Relish</td>
<td>چاشنی</td>
</tr>
<tr>
<td>Skunk</td>
<td>راسو</td>
</tr>
<tr>
<td>Moat</td>
<td>خندق</td>
</tr>
<tr>
<td>Yolk</td>
<td>زده تخم مرغ</td>
</tr>
</tbody>
</table>

### Appendix IV

**Condition 4: Semantic list (context)**

Last summer Jack went to the village for visiting his grandparents. His grandfather was working in the farm with a (1) weeder and his grandmother was making (2) topiary. She was doing it artfully with a (3) shear. He helped his grandfather plant the garden flowers using a (4) trowel. After finishing his job he sat under the (5) pergola and watched his grandparents work. The (6) sprinkler was watering the farm. They also had a beautiful (7) trellis. It was decorated with a (8) nozzle which was shaped like a bird. It was a nice place to rest by.

### Appendix V

**Condition 5: Thematic list (context)**

When I went to the zoo last week, I stopped by the (1) moose’s cage. The cage was next to tall trees. The (2) corpulent animal seemed hungry and was (3) devouring the green hay. I felt sad because it was (4) cooped up in the zoo. I imagined the free moose pulling the (5) sled in the snow at Christmas time with a decorated (6) scut. The moose should (7) roam in the big forest with its flat (8) antlers surrounded by other free animals. It’s not right to keep it here.
Appendix VI

Condition 6: Unrelated list (context)
One day Jack felt (1) 

donsie because he had eaten so much at a party. When they left home they realized that it was (2) 

airish outside since it is the beginning of the winter. Jack drove to an area out of town where he saw a 

Running (3) 

bayou. His kids loved the view of the running water and threw (4) 

dornickets in the water to make the frogs jump. The water looked (5) 

gauamy so he did not drink from it fearing of becoming sick of Malaria. Suddenly his wife saw a (6) 

doty and bright-colored object that looked strange in the ground. Jack got a (7) 

rastus, which he always keeps in his car to use on his farm, and started digging. What they found was a very pretty little metal box, with (8) 

punky wooden handle.

References


