Category Effects of Vocabulary Presentation in L2 Vocabulary Acquisition in Novice Learners

by

Amanda Elizabeth Wilcox

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Arts

Auburn, Alabama
August 6, 2011

Second Language Acquisition

Approved by

Almitra Medina, Chair, Assistant Professor of Spanish
Gilda Socarrás, Assistant Professor of Spanish
Alejandro Lazarte, Assistant Professor of Psychology

Abstract

Vocabulary learning is a crucial factor in second language acquisition (SLA), yet there is no consensus as to how to present vocabulary in a way that facilitates learning. Little theory and scant empirical evidence support the claim that presenting words in semantic clusters facilitates vocabulary acquisition, yet most text books present SL vocabulary in this way. Another branch of research is developing that seems to assert just the opposite- that in the beginning stages of SLA- grouping vocabulary words semantically actually impedes the acquisition of new L2 words. The purpose of this study was to investigate the effects of categorizing vocabulary on the acquisition of L2 in novice learners. Participants were presented with vocabulary that was randomly, semantically, or phonologically grouped and comparisons were made as to which vocabulary words were acquired more quickly and retained longer over time. Results show that the semantically clustered (and not phonologically similar) words were indeed more difficult to learn, and that performance was improved when vocabulary was presented in random and/or phonetic groups.

Table of Contents

List of Tables

List of Abbreviations

EFL   English as a Foreign Language

ESL   English as a Second Language

SL   Second Language

SLA   Second Language Acquisition

LTM   Long-Term Memory

STM   Short-Term Memory

WM   Working Memory

L1   First Language (Native Language)

L2   Second Language

TA   Think-aloud

CHAPTER 1 SEMANTIC CLUSTERING

Vocabulary learning is a crucial factor in second language acquisition (SLA). Second language (SL) learners must acquire hundreds of words at minimum in order to be able to communicate in the target language at the most basic level. Increasing vocabulary knowledge in novice learners should not only help the learner be able to communicate more effectively, but also help the learner comprehend a broader range of input from the target language. The question then arises as to how to package vocabulary in the most beneficial way to facilitate learning as well as better comprehension.

Many recent studies pertaining to SL vocabulary acquisition have been designed to test the overlying assumption that vocabulary should be presented in groups of words that are related semantically, or groups of words whose meaning would fall under one superordinate concept (Tinkham, 1993, 1997; Finkbeiner & Nicol, 2003; Waring, 1997). Most of the authors of these studies point out that the vast majority of available textbooks organize words in this way, under headings such as "body parts" or "kitchen utensils", word groupings that are linguistically linked by meaning. There is some theory that seems to support clustering vocabulary presentation in this way. But, there is also growing evidence and theory that suggest semantic clustering impedes SL vocabulary acquisition at the novice level. This literature review includes a general overview of some of the theory and evidence that both conflicts with and supports the commonly held assumption that presenting vocabulary in linguistically linked groups facilitates its acquisition.

Theoretical Foundation

Semantic field theory is based on the assumption that rather than being organized in lists of random words, vocabulary is cognitively organized by interrelationships and networks

between words, i.e., the mind classifies vocabulary by making connections in meaning; these connections in meaning are semantic fields (Channell, 1981). Channell interprets this to mean that words that are close in meaning are literally located closer to each other in the mental lexicon. The conclusion drawn from this theory implies that since vocabulary is organized in the mind into groups of words that are linked in meaning, it should be presented to students in groups of semantically related words to facilitate learning (Channell, 1981). Jullian's 2000 article supports this notion as an effective method of classroom presentation for L2.

Two concepts dominate the opposite perspective, that of interference theory and the distinctiveness hypothesis. Dating back to the 1931 first-language study conducted by McGeoch and McDonald, Robert Waring (1997) summarized the research behind interference theory (Baddeley, 1990; Higa, 1963, 1965), which shows that traces in memory often compete with each other. Waring describes interference theory as the idea that "when words are being learned at the same time, but are too 'similar' or share too many common elements, [in terms of meaning] then these words will interfere with each other thus impairing retention of them." (1997, p. 262, brackets added). Thus, he concludes vocabulary should not be presented in sets that are grouped under one superordinate concept.

The distinctiveness hypothesis focuses on differences rather than similarities and comes to what amounts to the same conclusion. Since similarity seems to confound the mind, distinctiveness should help organize it. Thomas Tinkham (1997) summarizes the research (Hunt & Elliot, 1980; Hunt & Mitchell, 1982) behind the distinctiveness hypothesis, which proposes that "as the distinctiveness (nonsimilarity) of information to be learnt increases, the ease of learning that information also increases" (p. 140, parenthesis in the original). This hypothesis also leads to the conclusion that vocabulary should be presented in a nonrelated fashion so that

the mind is presented with information organized in such a way that it is conducive for learning. As will be discussed later, even the theory that supports semantic clustering must acknowledge that the differences among a set of semantically related words are the distinguishing factors. Tinkham's research complements that of both Jullian (2000) and Channell (1981) by including a further distinction of semantic clustering, namely that of thematic clustering.

When referring to how the mind organizes lexical items, Tinkham states that semantic clustering is "based upon semantic and syntactic similarities among clustered words and 'thematic clustering' is based upon psychological associations between words and a thematic concept" (1997, p. 141-42). Semantic categories are linguistically based and thematic clusters are cognitively based. Though these two types of clusters will not be mutually exclusive (some concepts could be related both semantically and thematically) the nuance of difference between the groupings of 'apricot, peach, plum, nectarine, pear, apple', and 'frog, pond, swim, hop, green, slippery' is still apparent. One other defining characteristic of a thematic cluster is variable word class. While semantic clusters fall into categories like nouns (example above) and verbs (example below), thematic clusters contain a mixture of verbs, adjectives, adverbs and nouns. The purpose of Tinkham's research was to help fill a gap that continues to exist regarding the effects vocabulary clusters have on language learning. Clearly there are a number of ways in which vocabulary words can be grouped for presentation during second language learning, yet other than Tinkham's thematic contribution, current research seems to lack examination of the effects of types of clustering other than semantic and non-semantic.

Evidence

In her article on second language vocabulary acquisition, Channell (1981) suggests using semantic field theory (Lehrer, 1974) and componential analysis (Nida, 1975) in order to facilitate

vocabulary acquisition. Semantic field theory describes the vocabulary of language as existing in interrelating networks and relationships, i.e., semantic fields, as opposed to long lists of random words. She lists familial words (mother, father, sister, brother, etc) and verbs of movement (walk, stroll, run) as examples of semantic fields.

Channell suggests that since the mind seems to categorize vocabulary based on semantic similarity (Miller & Johnson-Laird, 1976), categorizing vocabulary into semantic fields will help the learner acquire it more easily. As support for this premise, speech errors in native language speakers, i.e., slips of the tongue, tend to be a blend of two words that come from the same semantic field. For example, instead of saying "asked", a speaker might say "invited" in the sentence "We have invited him to…asked him to buy crisps" or "herrible" when trying to say horrible or terrible (Channell, 1981, p. 118). She uses componential analysis in order to differentiate which meanings are shared by words and which connotations distinguish them from one another. During this componential analysis, the meaning or meanings of words are broken down in to pieces known as semantic components. Once the semantic components are derived, it becomes more apparent which aspects of meaning words share and which aspects are not shared, thereby distinguishing one word's meaning from another, providing a systematic way of describing vocabulary words.

According to Jullian (2000), because upper-intermediate and advanced learners of English as a Second Language (ESL) tend to over use their basic set of core vocabulary, they often experience 'a lexical acquisition plateau' and fail to comprehend input that varies even slightly from their core vocabulary learning (p. 37). Their oral production tends to sound very basic and often childish. In order to combat the overuse of a small, fixed vocabulary, Jullian sought to dissect the meanings of many close words so learners could communicate more

specific meanings. Since the learners are being presented with almost totally new vocabulary, they use dictionary definitions which are more specifically articulated once distinguished from a superordinate, more general term. It is important to note that what makes these more specific terms accessible to the learner is the difference, albeit subtle since they are synonyms, between the new vocabulary words.

Jullian cites Channell's (1990) study which suggests that concerning production, "access is via meanings and there is evidence for a model in which words with like meanings are 'close together' in accessing terms" (p. 39), Jullian interprets this to mean that vocabulary teaching should involve associations between semantically related terms.

Jullian reported on classroom procedures used in her Bachelor of Arts English class of native Spanish-speaking students to help them gain a wider range of vocabulary that can be categorized under a broad, classifying term. The first broad terms explored were "hit, beat, and strike". In the first stage of the case study, the students came up with thirty-five terms that were in some way similar and that in some way mean "to hit", like flog, clobber, spank, smack, punch, etc. Again, students were distinguishing terms by their differences.

In Stage 2 students classified the words according to common distinguishers and created categories for the vocabulary terms, even though some terms had overlapping categories.

In Stage 3, students created a network based on the semantics of the words with the different networks drawn to reflect how the information was organized in their minds. During Stage 4, learners tried to identify the various contexts in which the terms would probably be used, which meant that the students thought of corresponding subjects, objects, adverbs, and compliments. For example, "spanked" is usually a punishment by a parent inflicted upon a child and "flog" entails the use of a whip in an old military setting. This exercise allowed students to

examine the grammatical and pragmatic uses of each word.

The fifth stage prompted students to write sentences that illustrate the context and the new vocabulary and write sentences where the new vocabulary was used incorrectly to help reinforce the semantic connotations and differences of each word. For example, "I was so mad I slammed the tree." (Jullian, 2000, p. 43). Next, the students were asked to use the vocabulary in a figurative way. Finally the students chose their own words, mostly adjectives and adverbs, and did research which was reported to the class.

According to Jullian, the results of this three-day classroom activity helped students gain a much broader range of vocabulary pertaining to "to hit" and taught them to research meanings for themselves so they could be incorporated into their L2 speech. Students were able to recognize the tendency to form L2 speech based on L1 structures and through this activity they gained not only more intelligent-sounding vocabulary but also needed practice and polishing in L2 production. There are scant laboratory investigations that support the idea of grouping words semantically, though.

Schneider, Healy, and Bourne (2002) found that only in a very specific circumstance, did learners show an advantage with semantically related vocabulary. Sixty-four native-English speaking college students participated in two experiments which compared semantic (blocked category) versus non-semantic (mixed) presentation methods as well as tested the effects of the methods on vocabulary retention and transfer. Participants were tested, in both translation directions, at intervals on pairs of French and English words that were presented in either a blocked category or mixed fashion.

Half of the participants were trained using English words as the cue (with French responses) and the other half were trained using French words as the cue (with English

responses). The "practice schedules" (presentation of blocked or mixed vocabulary) varied as well as the stimulus. On the initial day of the experiment participants were given three training trials which were followed by an immediate test. Twenty-five words were presented on a computer screen in groups of five. Once the group of five words was presented for training, participants were tested on that group. This pattern of presentation/assessment was repeated until all 25 words had been presented and tested.

A week later, in the second session, participants were retrained and tested again; half of the participants were retrained in the same manner as the first session while the other half was trained using the opposite set of cues and responses. During the second session participants were presented with new five-word groupings of the same mixed-order words or a new order for the five word groupings of the words blocked by category. After completing the relearning process participants were tested again, this time responding to the same cue words presented in retraining, one by one. Since some students in Session 1 were presented with a different cue/response task than in Session 2 (i.e., participants were trained with an L1 stimulus and eventually tested on an L2 or vice versa), Schneider et al. were able to analyze the transfer of vocabulary that occurred between sessions.

The effects found pertaining to the blocking and mixed presentation of vocabulary hinge on translation direction and session. Only those participants given French cues during learning (Session 1), not during relearning, showed a "sizeable advantage" (p. 427) in producing the semantically related vocabulary words. On average, participants given the French cue for semantically related words responded correctly three-quarters of the time during the first session in comparison to those who were given the English cue for semantically related words. On average the latter responded correctly slightly less than half of the time (M= .43). In the mixed

category, the results are less striking with the mean proportion correct being .66 for the French cue and .42 for the English. Unlike the learning phase, during relearning there is very little difference in the mean proportion correct between the related and nonrelated words practice schedules with French cues (M= .77 blocked, M= .76 mixed) or English cues (M= .81 blocked, M= .78 mixed). The results indicate that participants had more difficulty learning the mixed sets of words and producing the L2 vocabulary initially, but that the groups presented with mixed sets or that were asked to produce an L2 response retained better over time. Overall, this is an extremely limited finding in support of semantic clustering and seems to lend more support to the idea that category presentation has little effect on SL vocabulary learning.

These findings support another study which showed that L1 vocabulary words are more likely than L2 vocabulary words to stimulate conceptual representations rather than lexical representations i.e. for novice learners, L2 words are more connected to their L1 counterparts than to the concept behind the word. The L2 –conceptual representation connection develops and is strengthened over time (Kroll, Michael, & Sankaranarayanan, 1998). Narrowing a potential set of responses to those that could fall under one superordinate category is the natural effect of grouping L1 responses in this way; conversely, grouping L2 responses in this way does little to narrow potential responses. The remaining results of the second session showing little discernable advantage for either practice schedule are also consistent with Kroll and Stewart's (1994) study which found that list organization had no effect on tasks involving the production of L2 words. If word list organization has no effect on learners' ability to produce a given set of vocabulary, then semantically clustering vocabulary should have no effect on vocabulary acquisition, one way or the other. Yet, there is a building body of evidence that seems to demonstrate, at least at the novice level, that presenting vocabulary in semantically related

groups tends to slow down acquisition, at least in the short term.

After examining several ESL texts, Tinkham (1993) observed that the vast majority of those texts presented vocabulary in "lexical sets" or "semantic clusters," both terms referring to groups of vocabulary that would fall under one common superordinate concept. Tinkham states that this method of presentation has very little empirical support for its use in the facilitation of vocabulary learning, referencing interference theory and the distinctiveness hypothesis as possible reasons for semantic clustering impeding rather than encouraging vocabulary acquisition. The purpose of this study was to test the hypothesis that participants learn L2 vocabulary more easily when it is presented in sets of unrelated words than when it is presented in semantic sets.

Twenty subjects, ranging in age between 16 and mid-40s, all non-native advanced-level English speakers, were presented with six pairs of words −one English word and one artificial word within each pair; half of the words were related to each other semantically, while the other half were unrelated in meaning. All of the English words were nouns that were already in the lexicon of each participant. Tinkham performed two experiments to test his hypothesis.

In the first experiment, Tinkham used a trials-to-criterion test to see how easily each participant could learn the English/ artificial pairs in sets of six −three related English words and three nonrelated English words− paired with nonwords. The trials-to-criterion test was given in two forms. Form A presented an English/artificial word pair while Form B presented the pair in reverse order. Each trial was presented as an audio recording and participants responded orally.

The results of the first experiment clearly show that the groups of unrelated words required fewer trials to reach criterion than did the semantically related words, even though the related and unrelated pairs were mixed together in the trials. Criterion was met once a participant

correctly said the artificial word/English word for all of the words in a given condition. A significant main effect ($F(1,18) = 22.5$, $p < 0.001$) for relatedness was found, that is to say it took participants on average 4.5 more trials to reach criterion in the related group than in the unrelated group in Form A of the assessment which presented the English/nonword pair order. The results from Form B (reverse order) follow the same trend though to a lesser degree. For this experiment, form was a between-subjects variable while relatedness was a within-subjects variable.

Similarly, the purpose of the second experiment was also to examine if the participants learned a set of semantically related English words more easily than they learn unrelated English word pairs; however, participants in the second experiment were exposed to six-pair sets rather than the three-pair sets. As with the first experiment, the trials-to-criterion form was also used.

The results supported those of the first experiment. The sets of six word-pairs that presented unrelated English words required fewer trials to criterion than the sets of related English words. Again a significant main effect for relatedness was found ($F(1,16) = 22.8$ $p < 0.001$) showing that related sets required more trials to reach criterion. Form was found to have any significant effect. Interestingly, although the subjects were diverse in cultural background, age and education level, the results consistently showed that participants learned sets of unrelated words more quickly than related words. Four years later, Tinkham extended this research.

In two separate, parallel experiments, using the same sophomore-level psychology student participants, Tinkham (1997) explored the effects of thematically versus semantically clustering words on second language acquisition. Although much SLA research (Hunt & Elliot, 1980; Hunt & Mitchell, 1982; Waring, 1997) indicates that vocabulary acquisition is impeded by semantic clustering, but thematic clustering seems to facilitate vocabulary acquisition.

Tinkham's two research questions examine 1) if students learn semantic clusters of vocabulary (e.g., dish, bowl, plate) with more difficulty than nonrelated words (e.g., island, potato, beard) and 2) if students learn thematically clustered words (e.g., beach, sunny, swim) more easily than sets of unassociated words (e.g., triangle, improve, sweet).

In this study, 48 native English-speaking sophomore-level educational psychology students participated. One half participated in the oral modality and the other half in the written modality. In a trials-to-criterion format, students were presented with six pairs of words; each pair consisted of one English word and one artificial word. Three of the English words were semantically clustered and all were from the same word-class; the other three were thematically clustered and consisted of a noun, a verb, and an adjective. In the oral modality, participants were asked to hear the artificial word and verbally express its English equivalent and vice versa, that is hear the English prompt and say aloud the artificial word. In the written modality, participants were asked to recognize the English word and type its corresponding artificial word and vice versa. Criterion was met when individual participants produced the correct answer for each of the three words on two separate trials. Students were asked to answer questions following the study to contribute to the researcher's analysis.

The results of each experiment mostly support Tinkham's predictions. Tinkham tested the following conditions as an independent variable: 1) linguistically related sets (semantic clusters) 2) linguistically unrelated sets, 3) cognitively associated sets, and 4) cognitively unassociated sets. Data indicates that a significant difference ($F(1, 44) = 57.83$, $p < 0.001$) was found between semantic clustering and unrelated words, showing that it took longer to reach criterion when testing semantically related words. Clustering words thematically also reduced the number of trials needed to reach criterion in comparison to learning a set of unrelated words ($F(1, 44) =$

14.50, $p < 0.001$). Responding to the researcher's questions referred to above, students commented that some of the semantic sets "were all the same" (p. 379) which made it more difficult for them to remember the corresponding artificial word. Most participants, although they may not have been able to explain why, indicated that the thematically based sets were easier to learn.

The purpose of Waring's (1997) study was to replicate that of Tinkham (1993), but to use subjects of an L1 other than English so as to make more generalizations about this effect to other languages. Waring (1997) examined multiple second language course books and found that often vocabulary is presented in sets of words that fall under a single common superordinate word. This method of presentation seems to be based in methodology and not on research. Waring cites interference theory (McGeoch & McDonald, 1931; Baddeley, 1990) as a possible reason why words are more difficult to retain when presented in this manner, since the clustered words share too many common elements of meaning.

Waring created artificial words (L2) to use with his 20 native Japanese-speaking participants. Participants were presented with six word-pairs, three of which shared a superordinate concept like 'clothes' along with three pairs of unrelated words. Waring also used the trials-to-criterion method. Using data for form, or the order of the pairings, as a between-subjects variable and relatedness as a within-subjects variable, the MANOVA results show that participants were able to learn unrelated words in fewer trials than related words supporting the hypothesis that semantic clustering interferes with learning. The significant main effect ($F(1, 18)$ $= 18.9, p < 0.001$) found for relatedness reflects the greater difficulty in meeting criterion in the semantically related set. Waring notes that several students commented that it was difficult to learn the related pairs because they seemed "all jumbled up" (p. 267). He also noted a common

finding that 25% of the errors made when pairing related words were from the same superordinate category, where as this occurred only 5% of the time for nonrelated words.

Finally, Waring notes that these effects are most likely strongest for beginning learners as they have no store in which they can base their new learning. It could be argued however, that since beginning language learners have no 'store', confusion of words should be less likely. As stated earlier, L2 vocabulary acquisition requires a new form-meaning connection. If nearly the same semantic field is being called upon by multiple stimuli, it seems reasonable that learners would have more difficulty remembering the distinguishing features between words as well as the words themselves in the beginning stages of acquisition. When L2 words are too similar in terms of meaning, they seem to compete with each other causing confusion rather than clarity.

Finkbeiner and Nicol (2003) acknowledge that there is some L1 research that supports the idea that learners benefit from semantic categorization of vocabulary presentation based on the premise that similarity between words facilitates learning and encourages the learner to notice distinctions between words leading to a better understanding of them (Gairns & Redman, 1986). Finkbeiner and Nicol reference numerous ESL textbooks that seem to *assume* that this method of presentation facilitates L2 vocabulary acquisition −'assume' because there is little empirical evidence in L2 to support this claim. Learning an L2 requires the creation of new form-meaning connections not needed in L1. Since L2 vocabulary acquisition requires first creating and then strengthening the new L2 form-meaning connection it is differentiated from remembering a list of L1 words. This notion does not account for the fact that acquisition of new L1 vocabulary also requires attention to the form-meaning connection.

Taking longer to acquire a new set of semantically linked vocabulary words in comparison to a new set of nonrelated words as shown by a number of L2 studies (Higa, 1963;

13

Nation, 2000; Tinkham, 1993, 1997; Waring, 1997) however, does not necessarily mean acquisition is impeded because learning under these circumstances (i.e., semantically clustered vocabulary) may actually make the words easier to process than non related words and, in turn, use once they are learned. On the other hand, it is possible that something that makes words more time-consuming to acquire may make them more difficult to retrieve later on. In fact, semantic interference does seem to play a significant role even in novice L2 learners (Altarriba & Mathis, 1997; LaHeij, Hooglander, Kerling, & van der Velden, 1996), which means that the words that were more difficult to learn were more difficult to access and produce. That being said, there is no data from this experiment that accounts for the long term retention of these words; therefore, the role of semantic clustering in L2 development still remains unclear.

Finkbeiner and Nicol's (2003) study consisted of a training phase where participants learned new labels for familiar terms, followed by a test phase that required participants to retrieve those labels in a translation task in both directions (i.e., L1-L2 and L2-L1). The semantic grouping of the words was manipulated during both phases (training and testing) of the experiment.

Forty-seven monolingual English-speaking undergraduates were exposed to 32 novel words that were paired with a picture of a familiar concept such as an animal, kitchen utensil, body part, or piece of furniture. Each student participated in two, 45 minute sessions over a five-day-period. Each session consisted of both an L2 vocabulary training session, followed by a recall activity and a bidirectional translation task, for a total of four blocks of translation.

During the training session each participant heard a recording of the L2 (novel word) and saw a corresponding picture on the monitor followed by a repetition of the L2 word. In the semantically clustered training condition, items were blocked into groups of eight which were

presented four times during each session. In the unclustered condition, the word order was scrambled within a given block and each block was also presented four times in random order. During the recognition task, half of the picture-label pairs were correct, the other half were incorrect; participants had to choose 'yes' or 'no' depending on whether they believed the label was correct or not. Participants were told the accuracy of their responses as well as their response times. Next, students were given the translation task and were required to speak L2 equivalents into the recording device as quickly as possible. Half of the students translated from L1-L2, the other half in the reverse direction in an attempt to counterbalance differences that could occur due to direction of the translation.

Data from about half of the participants were unusable since they did not meet the 80% accuracy criterion. Of the remaining twenty-four participants, translation in the forward direction (L1-L2) took 238ms longer in the semantically clustered condition ($F(1, 20) = 12.60$, $p = 0.002$) than those exposed to the random grouping of the same L2 words. Translation in the opposite direction was 146 ms slower for those who were presented with semantically related words ($F(1, 20) = 7.46$, $p = 0.013$). It is clear that the semantic category effect was negative in this experiment in terms of speed of recall.

In order to explain these category effects, Finkbeiner and Nicol (2003) suggest that the constant conjuring of the same *lemma* (lexical hub of form, meaning and syntax) repeatedly during the semantic clustering training condition caused interference and delayed translation in both directions. This research converges with other studies and suggests that it takes longer to retrieve newly acquired L2 words when presented semantic clusters. Most recently, this branch of research has been continued by being applied directly to SLA using L2 vocabulary instead of novel/nonwords.

In an attempt to determine whether semantic or nonrelated groupings were most beneficial to students, Papathanasiou (2009) tested two ways of presenting vocabulary. Thirty-one native Greek-speaking children studying intermediate English as a Foreign Language (EFL), and 32 native Greek-speaking beginner EFL adults participated in the experiment.

Over the course of three weeks, participants met two times per week for 50 minutes. Students were assessed immediately after the final lesson and again two weeks later. The two groups of native Greek speakers were presented with sixty words in English in one of two ways: semantically-related sets of nouns or semantically-unrelated sets of nouns.

Participants were presented with ten words at a time. The first stage of the presentation, which she termed Noticing, lasted ten minutes. Students had vocabulary on small cards and were encouraged to look at them and try to recall meanings. During the next fifteen minutes −the Retrieval stage− students looked at a set of L2 words and tried to orally recall the L1 equivalents. Retrieval was repeated five to six times per student. During the Generation phase, participants did matching and fill-in-the-blank activities for practice.

After the three phases, students were given a pencil-and-paper test and asked to recall the L1 when provided with the L2. Results show that beginning-level adults in the unrelated-word presentation group performed significantly better than those in the related group. The average score for correct responses on the short-term test was about 23 for the related words and about 31 for the unrelated word grouping. The results are similar for the delayed test showing an average correct score of about 20 for the related words and about 27 for the unrelated grouping. The children showed no significant difference in test scores in either the posttest or the delayed posttest. Papathanasiou explained that the adults could have been more motivated to learn new vocabulary because of a need to learn English that the children may not have had or felt.

Papathanasiou (2009) concluded that perhaps educators should begin to rethink presentation practices at the beginning level, at least for adults. She also suggests that more research is needed to validate these results and that other contexts are worthy of examination.

Given this body of conflicting evidence, further investigation is clearly warranted in order to elucidate the effect of vocabulary presentation. Research that supports semantic clustering (Channell, 1981; Jullian, 2000; Schneider et al., 2002) consists mostly of case-studies, except for Schneider et al. (2002), whose finding for semantic clusters in L2-L1 word translations has yet to be corroborated. Tinkham's (1993, 1997) research strongly supports organizing words into nonrelated or thematically related groups, but the implications of his investigations along with those of Waring's (1997) and Finkbeiner and Nicol (2003) are limited since the words being tested were artificial. Furthermore, these authors did not test over the long term, which restricts the conclusions that can be drawn for L2 development. Papathanasiou (2009) presents a similar argument using English words as the L2 with adults. She found that adult beginning-level ESL students scored significantly higher in both the posttests and delayed posttests containing unrelated vocabulary. However, further research probing into the effect of semantic clustering is certainly warranted in order to address the paucity of experimental research in this strand of investigation.

CHAPTER 2 PHONOLOGICAL CLUSTERING: ANOTHER METHOD OF VOCABULARY

PRESENTATION

Another, parallel, branch of research that could potentially complement that of L2

vocabulary acquisition regards novel word learning in L1. When dealing with learners that are at

the beginning level of an L2, it seems reasonable to think that at least *some* of the mechanisms

for L1 vocabulary acquisition could be applied to L2 learning. A new word is a new word, be it

in L1 or L2. One of the most fundamental levels of L1 novel word acquisition requires

consideration of short and long term memory (LTM) and moving information from one to the

other. Donald Hebb's (1949) postulates provide a model for the effects of repetition over time,

learning, and the move of information from short term memory (STM) to LTM. Their

implications in learning have been explored in much research; pertinent to this study, however, is

L1 novel and nonword learning. The following brief literature review is provided in order to

justify the testing of another potential method of vocabulary presentation

Evidence

Short term memory is heralded as a precursor or first step in the process of moving

information from STM toward LTM during acquisition, and repeated presentation of information

increases retention levels (Baddeley, 1990; Melton, 1963; Peterson, 1959). With this in mind,

Szmalec, Duyck, Vandierendonck, Mata, and Page (2009) present a study that connects STM

and LTM theory to the Hebb effect and produce what they call "a laboratory analogue of novel

word learning" (p. 435).

In an effort to understand how the mechanisms involved in short-term retention of verbal

information are linked to learning novel word forms, Szmalec et al. (2009) designed two

experiments that tested the hypothesis that the "long-term serial order memory component of

18

word-form learning depends on the same mechanisms as those underlying the Hebb repetition effect" (p. 436). The study correlates the learning observed in an immediate serial recall task after the repeated presentation of a list −of numbers, words or serial data− to the processes that contribute to the development of long-term phonological lexical representations by using the Hebb effect as a link between the STM and LT memory. The Hebb effect is explained as the effects of repetition: that recall performance significantly increases for sequences that are repeated than for those sequences that are not. Hebb's (1949) model shows that the repeated sequences encoded in STM generate traces in LTM that outlive STM. The recall of the repetition gradually generates a stable representation in LTM. The phonological nature of this type of learning underlines its importance in general language processing. Frost (1998) and Duyck (2005) even argue that lexical entries are mainly of a phonological nature in both L1 and L2. It seems, then, that Hebb's model could be pertinent in L2 acquisition as well.

In two experiments using the same 42 native Dutch-speaking first-year university students, Szmalec et al. (2009) attempted to demonstrate that Hebb sequences enter the mental lexicon in the same way that novel word forms do. Participants were presented with sequences of nine nonsense syllables for immediate serial recall. Each syllable was constructed as a consonant and vowel pair (CV). CVs were arranged into two blocks of 36 sequences of nine CVs, every third of which was repeated, meaning that 24 different "filler" sequences were presented to each participant while one sequence was repeated twelve times −the Hebb sequence. The CVs in the filler sequences were rotated so that participants would be exposed to the Hebb CVs and the filler CVs an equal number of times in order to dispel any question as to participants' familiarity with the CV.

Centered on a touch screen monitor, each CV was presented individually and serially.

After each sequence, participants were asked to recall the order of the previously presented

sequence. This was accomplished by the participant attempting to touch CVs, in the order

presented. The number of correct responses increased significantly between trials in the Hebb

sequence versus the filer sequences. In order to build on these results the second experiment used

Hebb sequence nonwords, words formed out of three CV pairs from the Hebb sequence.

Participants were given an auditory lexical decision task in order to see if any lexical

representations had been made with the Hebb sequence material, since any representation formed

would reflect the phonological representations created during Hebb learning by participants.

For the purpose of clearing the STM of the participants, administrative tasks were

performed during the five minutes between experiments. In the lexical decision task (LD),

researchers created four different types of stimuli: CVCVCV nonwords formed from

nonrepeated filler sequences from the first experiment, CVCVCV nonwords from Hebb

sequences, nonwords that were not CVCVCV but still consisted of six letters, and finally, Dutch

words of equal length. The stimuli were presented through closed earphones and participants

were asked to indicate by pressing a response box (on the touch screen) as quickly as possible

indicating whether a stimulus word was a word or a nonword.

As predicted, the reaction times for Hebb nonwords was reliably slower −meaning it took

participants longer to reject Hebb nonwords as nonwords since they seemed to have formed

some type of phonological lexical representation. This result is demonstrated by a main effect for

condition, $F(1,39) = 5.65$, $n^2_p = .13$, $p < .05$. Participants seemed to form sub-vocal phonological

representations of the words on their own since they were only exposed to visual representations

of the CVs and sequences. The strong associations between short-term serial recall and novel

word learning that exist seem to be connected by the Hebb repetition model. This study

demonstrated that what is learned during immediate serial recall of phonological information creates a cognitive representation similar to that of a newly acquired word thus, a laboratory analogue of novel word learning. This research compliments that of Mosse and Jarrold (2008) who conclude that "acquisition of novel word forms may rely on the domain general ability of representing serial order in working memory" (p. 441). This type of learning is implicit; these laboratory results attempt to show patterns and mechanisms of learning of which the participant has little control or awareness.

Using a sample of 42 five- to six- year old English speaking children, Mosse and Jarrold (2008) correlate the magnitude of Hebb learning with word and nonword paired associate learning. The aspects of their study pertaining to the relationship between Hebb learning and novel/nonword learning reinforce the findings of Szmalec et al., (2009) and further support a rational for connecting this branch of research with vocabulary presentation in L2. Mosse and Jarrold (2008) found that only nonword learning was significantly related to the magnitude of Hebb learning, which could suggest that this process or type of learning might be functional in an L2 model as well.

For the paired-associate word and nonword learning portion of the experiment, 42 participants were presented with two sets of four images of "novel creatures" across both tasks. In the paired-associate word learning task, these images were linked with familiar names: Michael, Simon, Thomas, and Peter; the participants were asked to learn these associations. The presentation consisted of an image appearing on the computer monitor while the name of each creature was heard from the computer. Once the presentation was complete, participants viewed all four creatures and were asked to identify each one. In this experiment each child participated in a total of ten trials unless he/she correctly named all of the images during two consecutive

21

trials; testing was not continued in this case. Using nonwords instead of familiar ones, the nonword task was performed in the same manner the as word learning task. The phonemes of the familiar names mentioned above were the repetitive factor; they were rearranged and used to create the nonwords for this task.

According to its authors, this experiment is the first of its kind to assert a "direct connection between Hebb learning and new word learning" (p.512) possibly because a serial ordering-mechanism is at work in this process.

The aim of the current study is not to replicate the Hebb effect as that is well beyond its scope; however, the aim is to use the implications of this parallel line of research to justify testing phonological category effects on vocabulary presentation in an L2. Two trains of thought support the application of this L1 research in an L2 laboratory setting. First, it is common in both first language acquisition and SLA to use nonwords for research purposes. In L2 these nonwords represent a novel word of which participants can have no way of having previous knowledge. In L1 research, nonwords can represent results for new word acquisition. Since both branches of research employ the use of nonwords, it is not a stretch in logic to see a connection between brain mechanisms used in general language acquisition as seen in Hebb research and SLA investigations pertaining to methods of vocabulary presentation and acquisition.

Second, Hebb research shows that when a series has a repetitive aspect and when participants are required to recall a series, learning is facilitated. Individual words can be perceived as a series of bits of information, as seen in Szmalec et al. (2009), which makes application to SL vocabulary acquisition easy. If recalling a series with a repetitive aspect facilitates non/novel word learning in L1, then it seems reasonable to investigate if L2 words can be acquired in the same or a similar way.

Since the L2 that pertains to the current study is Spanish, separating a series or word into CV combinations deconstructs them into phonemes. Could repeating a combination of phonemes in a group of words be enough to trigger and then create associations that carry novel recalled words from STM into LTM? Since there are virtually no studies that investigate the above question in L2, this study not only provides further evidence to support or refute current research on the role of semantics in vocabulary learning, but also provides preliminary results in a new line of research on the effects of phonological categorization.

Research Question

The following research question will guide my study:

What are the effects of categorization (+/- semantic, +/-phonological) on L2 vocabulary learning among novice-level learners of Spanish immediately after practice and over time?

Hypothesis

This study categorizes vocabulary into the following four groups: [+semantic, -phonological], [-semantic, - phonological], [-semantic, + phonological], [+semantic, + phonological. Based on the findings of Tinkham (1993, 1997), Waring (1997), Finkbeiner and Nicol, (2003), and Papathanasiou (2009), I predict that vocabulary presented in the semantically clustered conditions will be more difficult to learn in the short term than vocabulary presented in a semantically nonrelated fashion. According to Schneider et al. (2002), this difficulty could facilitate better long-term retention. I also predict that the phonologically similar groups will be easier to learn than the semantically clustered set since the participants will be exposed to the repetition of serial information that is conducive to novel word learning (Mosse & Jarrold, 2008; Szmalec, et al., 2009). And, even though there is no precedent from which to speculate, it seems

reasonable that since one of the basic features of the Hebb effect is its participation in the process

of moving information from STM to LTM, participants will better retain those word-pairs.

CHAPTER 3 METHODOLOGY

Participants

Fifty-four native-English speaking Auburn University students taking Elementary

Spanish I originally attended the first session, but only 38 of those participants remained in the

final participant pool. Selection for inclusion in the study was based on meeting three criteria:

attending all three sessions, claiming English as their native language, and scoring 0 on the

pretest. Of the 38, six other participants took part in a think-aloud (TA) protocol where their

verbal thoughts were recorded as they completed the second session (for more details about the

protocol, see "Think-aloud protocol" below). In an effort to avoid any type of interaction, these

participants' data were collected separately their scores are not included in the following

analyses. The purpose of the recordings was to attempt to obtain a verbal account of how

participants were remembering the various categories of vocabulary and will be used in the

discussion. Data from the remaining 32 participants was used in the statistical analyses.

Experiment Design and Materials

In a pre-post-delayed design, with time and category as independent variables, four

categories were tested using sets of five words that were semantically linked or not, and

phonologically similar or not: [+semantic, -phonological], [-semantic, - phonological], [-

semantic, + phonological], [+semantic, + phonological]. Participants were given an English

prompt and were asked to produce the written Spanish equivalent which according to Schneider

et al. (2002), is a more difficult task than producing the L1 equivalent and should aid in long-

term retention. All participants completed the materials in the same computer lab and were

exposed to all four conditions in a randomized order.

During the first session, participants signed the consent form, filled out back ground

questionnaires designed to collect information on their language experience, and completed a pretest on E-Prime (2.0). The first session lasted about 20 minutes. The pretest was an L1-L2 translation task of 30 vocabulary words (see appendix A) of the same class, nouns. In particular, 20 were target words (one set of five words for each of the four conditions as performed by Schneider, et al. (2001)), and 10 were distractor words. The distractor words are words typically found in an introductory textbook. The target words were chosen and grouped into four categories. The first group, +semantic, -phonological ([+S-P]), all fall under the common superordinate "tools". The second group, -semantic, - phonological ([-S-P]), was chosen based on their lack of similarity both in meaning and sound. In both phonological groups, the beginning of each word was repeated. The third group, -semantic, + phonological ([-S+P]), is a grouping of words that do not share meaning, but as stated *do* share common beginning sounds. The final group, +semantic, + phonological ([+S+P]), share not only the common beginning of the word, but also share meaning in that each one pertains to torment or torture.

Each English prompt was presented one at a time in lower case type. Participants were asked to type in the Spanish form of the word if they knew it, or think they knew it. Otherwise, they were asked to type a question mark (?) in the response field. Participants were not given any feedback.

During the second session, one week later, participants were trained only on the target vocabulary. On a computer monitor using E-Prime, word pairs along with the sentence "_____ means_____" as seen in Tinkham (1993, 1997) appeared on the screen for approximately two seconds following Schneider et al. (2002). Tinkham's sentence was modified slightly for the +S–P group to "A_____ is a _____". This modification was made to make sure participants did not confuse nouns in that category with verbs. Once participants

viewed a conditioned 5-pair set of vocabulary words, they were assessed on that set of words. During this partial assessment, participants saw an English prompt and were asked to produce the Spanish equivalent. Once the test on those five words was complete, participants continued on to the next condition for a new set of five word-pairs. Participants were tested again, in the manner stated before, and the process was repeated until all 20 target words were presented and tested. That was considered one trial. Participants did three trials, with the order of words within the 5-pair sets and the order of sets for each trial randomized, following the model provided by Schneider, et al. (2001). At no point during any of the trials were the participants given feedback, but the intermediate recall of the vocabulary followed models in Schneider et al. (2002) and Finkbeiner & Nicol (2003) and is also supported by Nation (2001). Nation affirms that having to recall a meaning when seeing or hearing a particular cue strengthens the form-meaning connection (Papathanasiou, 2009). After the third trial, participants took the immediate posttest which consisted of the 20 target words in random order presented in the English form one by one on the computer screen. Participants typed in the Spanish equivalent. Words were not presented in the same order as in the preliminary pretest. The second session lasted approximately 20 minutes as well.

Approximately two weeks later, participants returned to take their final assessment, the delayed post-test. Students were assessed as in the first session, via E-Prime using English prompts in lowercase letters centered on a computer screen. This session also lasted about 20 minutes. Participants were assessed on the 20 target vocabulary words. Papathanasiou (2009) also tested her participants two weeks after completing all of the training in an effort to see what was retained over that time period. This will help expand on previous results that did not include a long-term component. A summary of the research design is shown in Figure 1.

Scoring Procedure

Two separate scoring methods −strict and lenient− were used and analyzed. In the strict scoring method, participants were awarded 1 point for exact reproduction of the target word and 0 points otherwise. No partial credit was given for spelling inaccuracies. Comparisons were made in the number of correct answers for each condition, and the change over the course of the sessions. In the lenient scoring procedure, participants received 1 point for accurately producing 2 out of 3 syllables in three-syllable words and 1 out of 2 syllables in two-syllable words. Zero points were awarded otherwise. The lenient scoring procedure is included in an attempt to account for a broader definition of learning. Any significant findings for both scoring methods are reported at the $p = < 0.05$ level.

Think-aloud protocol

In the think-aloud protocol, participants were asked to verbalize their thoughts while practicing the target vocabulary during the second session. Each participant received special instructions and practiced thinking aloud prior to beginning the target vocabulary task. During the think-aloud training session participants were given an example TA i.e., they read a passage of what a TA might sound like, using a simple math problem as an example. Then participants were given three Spanish-English word-pairs and were asked to say whatever passed through their minds as they tried to remember the new words. Once training was complete, participants began the target vocabulary task. Oral data was collected using headsets and Audacity software.

CHAPTER 4 RESULTS

Since each participant scored 0 in the pretest, data from that assessment was not used in

the analysis. Data from the strict scoring procedure[1] from the 32 non-TA participants were

submitted to a 4x2 repeated measures analysis of variance (ANOVA), with category ([+S+P], [-

S+P], [+S-P], [-S-P]) and time (post and delayed) as the within-subjects variables. A significant

main effect was found for time, $F(1,31) = 19.837$, $p = .000$, and category $F(3, 31) = 101.27$, $p =$

.000, and a significant Time x Category interaction was also revealed, $F(3,31) = 7.711$, $p = .000$.

The effect size (partial eta squared) was considered large in each case. See Table 4 for the

ANOVA results, Table 5 for descriptive statistics, and Figure 2 for a plot of the means over time.

The main effect for time shows that participants scored significantly lower in the long-term as

would be expected.

In order to gain a clearer understanding of the interaction between categories, separate

one-way ANOVA tests were carried out for the immediate and delayed posttests with category as

the independent variable. The analysis of the posttest scores showed a significant main effect for

category, $F(3, 31) = 17.414$, $p = 0.0$. See Table 6 for ANOVA results and Table 7 for descriptive

statistics for the posttest. The [+S-P] group of vocabulary words were by far the most difficult to

learn (M =1.38) in the posttest; that difference was significant at the .05 level when compared to

all three other groupings. Though the [-S-P] group fared the best (M= 3.03) in the posttest,

---

[1] The data from the lenient scoring procedure was also submitted to repeated measures ANOVA. A significant main

effect was found for time, $F(1,31) = 219.84$, $p = .000$, and category, $F(3,31) = 23.77$, $p = .000$, and a significant

interaction between time and category was also found, $F(3,31) = 8.23$, $p = .000$. However, the same patterns

emerged from this analysis as from the strict scoring procedure. Therefore, tables and charts display results from the

strict scoring procedure only.

analysis also showed that both phonological groups, [-S+P] (M = 2.16) and [+S+P] (M =2.38) were significantly improved when compared to the [+S-P] cluster, but that neither phonological cluster was significantly different when compared to the other. The [+S-P] category fared worse than the other three categories. Though the +phonological categories fared better than the [+S-P] category, they still did not fare as well as the [-S-P] category.

Finally, the effects of category as an independent variable in the delayed posttest were analyzed. It is important to note that retention was generally poor for all 4 categories. This is understandable since participants only practiced the words for 20 minutes two weeks prior to the delayed posttest assessment. Every student also stated on the post-experiment questionnaire that they did not practice these words in any way during the break. Even with the poor retention, however, there was also a significant main effect for category in the delayed posttest, $F(3,31) =$ 3.559, $p = .017$. The results of the delayed posttest follow the tendencies presented earlier, but vary slightly. The [+S-P] category still had the lowest average (M =.06) of all the categories, but this was not significantly different from the retention levels of the [-S+P] category ($p = .161$). And though the [-S-P] category again had the highest average (M = .47) overall in the delayed posttest, this average was not significantly different from either of the phonologically similar categories, [-S+P] $p = .109$, [+S+P] $p = .521$ respectively. It seems that in the long term, participants remembered very little in general, but significantly more than the [+S-P] in the [-S-P] and [+S+P] categories. See Table 8 for the descriptive statistics.

Discussion

The research question investigated in this study sought to observe effects of categorization on L2 vocabulary learning. Categorization seems to have a strong impact on acquisition of L2 vocabulary among novice level Spanish learners in the short term and some

effect in the long term. These results provide some support for the presented hypothesis. First, the results partially support the notion that semantically clustered word groupings are more difficult to learn than nonrelated groups of words. This difficulty in the short-term did not, however, translate into better long-term retention as had been originally predicted. The two phonologically linked categories consistently had significantly higher means than the [+S-P] category, and consistently and significantly lower means than the [-S-P] category, but were not significantly different from one another. The following discussion attempts to explain these findings.

The semantic set in question, tools, should have been a fairly easy one to learn, but it proved more difficult for participants. Many participants stated in their post-experiment questionnaires that this particular group was "harder", but their responses were mixed as to why that was true. About half stated it was because Spanish forms were so different from their English forms; the other half did not know why they seemed more difficult. The lenient scoring procedure provides information that neither clarifies nor convolutes the stated results, but rather confirms them. Even with additional credit for partially remembering words in the posttest, the means of the [+S-P] group was significantly lower compared to all three other groups in the posttest ($p = .000$ [-S-P], .000 [-S+P], and .000 [+S+P]) and on the delayed posttest when compared to the [-S-P] group ($p = .003$ [-S-P], .037 [-S+P], and .000 [+S+P]).

Though Channell (1981) and Jullian (2000) both provide compelling arguments concerning how words are organized and then stored in the brain, they seem to make a hasty conclusion. Even if the brain stores vocabulary in semantic fields, it does not necessarily follow that vocabulary should be presented in that fashion. In fact the opposite seems to be becoming more and more apparent. It is possible that even though the mind may tend to store accumulated

vocabulary in semantic fields, it is prepared to receive words that need to be organized in that way, not to receive words that have already been organized. And, it makes sense that the brain would be accustomed to varied input. In the beginning years of L1 acquisition words are not exclusively presented to babies and children in semantic groups. It could be argued that babies are exposed to some semantic sets, when parents or care-givers talk to their babies about body parts, narrate feeding time, etc. But, considering that the baby has no store with which to confuse the related vocabulary, it seems unlikely that semantic fields exist at that point, so that interference would play a small role, if any, during the beginning stages of L1. (It would be difficult to make this argument apply to beginning L2 learners who are already fluent and literate in their L1.) Yet, children acquire an extremely large and varied vocabulary by the time they are school aged. Once the child attends some type of formal education, words are at times presented to them in a semantically organized way for later recall. At this point in L1, it would be difficult to argue that semantic fields do not exist for the learner, therefore interference and semantic similarity could begin to pose a problem at this stage. For second language learners, presenting vocabulary in semantic groupings should at least be varied with other methods. Presenting nonrelated vocabulary, since it significantly out preformed [+S] in every case except the long-term, and possibly phonologically similar words instead could provide a wider variety of vocabulary to students which could combat the "plateau and often childish sounding" speech to which Jullian (2000) referred.

In comparison to the semantically grouped words, both phonologically similar word groups demonstrated significantly improved acquisition after practice, whether they were semantically linked or not. From this perspective, it seems fair to say that the phonological category effect was positive in the short term when compared to the [+S-P] group.

This is an interesting new finding because it is a reminder of the role phonology plays in SLA, especially in novice learners. Participants were not primed in any way about the types of categories they would be exposed to, yet the similarities between words seem to be enough to have aided in their acquisition of those words. It should be obvious that phonology would be an important factor in SLA, but as mentioned earlier there is a lack of study of its effects on vocabulary acquisition. This lack is often repeated in the classroom where other than communication in the target language, there is little explicit attention given to sounds and corresponding letter combinations. Grouping the words together based on phonology makes that aspect obvious to the learner. And, repeating that aspect seems to increase the chances of acquisition. It seems possible now that some of the mechanisms involved in L1 novel-word learning could also be involved in novice-level L2 word learning. It is curious though, that there was no difference between the two phonological categories in the immediate posttest or the delayed posttest since one group ([+S+P]) had the confounding factor of semantics as well as the phonological word similarity. This might be explained by word choice.

A closer look at the [+S+P] category reveals that it is the only group that contained two words whose meaning was changed by a single letter *maza* and *mazo*. Not only are they only a letter apart, but several students responded on their post experiment questionnaire that these words were particularly easy to remember because they were so similar in spelling and in terms of their English equivalents, mace and mallet. If two of the five words of that group were particularly easy to remember[2], it makes sense that their acquisition/retention would be increased as well, perhaps unfairly so, when compared to its non-semantically related counterpart, [-S+P], thus creating no significant difference between the two groups, when maybe there should be.

---

[2] A post hoc examination of the immediate posttest scores of the number correct between words in the [+S+P] group support this notion. Twenty-nine participants correctly responded with *maza* and 23 responded with *mazo*. In comparison only 5, 7, and 12 participants responded correctly to *mazmorra*, *mazorca,* and *mazazo*, respectively.

Even so, the increased means of the [-S+P] category when compared to [+S-P] indicate that phonological clustering might be beneficial to L2 Spanish learners at the novice level

On the other hand, since the [-S-P] group nearly always presented significantly increased means compared to all three other groups, it could be said that phonological clustering inhibited learning when compared to the [-S+P] group in the short term.

It is expected that over time recall will decrease. In this case it decreased substantially, where even in the best performing category, the average was about half of a word. The [+S-P] group had a significantly lower mean than both the [-S-P] and [+S+P] categories, but not a significantly lower mean than [-S+P], implying that the [+S-P] and the [-S+P] categories were about equally as difficult to recall. This long term finding may also be explained by continuing to look at the [+S+P] category.

It could be argued that this group of words was not, strictly speaking, a semantic cluster. Upon first glance of the [+S-P] word list it is easily discernable that it is a list of tools. The words in the [+S+P] category are linked in meaning, one participant called them "mean" words, the superordinate that dominates them is not as clear as it is in the [+S-P] group. It might even be argued that the [+S+P] group could be considered as a modified version of Tinkham's thematic clusters, though the words in this study are all of the same word class. If the mean of the [+S+P] group was unintentionally inflated because of word choice, then there would probably be little long-term genuine difference between those three groups (i.e., [+S+P], [-S+P], [+S-P]) at all in the strict scoring in the delayed posttest. And though the [-S-P] was significantly improved when compared to [+S-P], it did not fare better statistically than the phonological groups. The fairest way to interpret this seems to be that, minimally, there is some advantage to grouping words randomly in the long term, and this advantage may also be found to an extent in the phonological

groupings (see below).

In the real world partial recollection is more valuable than no recollection at all. And, in this case, it seems pertinent to look at the lenient scores for the delayed posttest. Using the lenient scoring method, partial learning was accounted for which, after such a long break with no new practice or priming before the final assessment, seems warranted. The results of this analysis actually mimic the results seen in the strict scoring of the immediate posttest: both phonological categories had significantly higher means when compared to the [+S-P] word group. Neither phonological category had a significantly different mean from the [-S-P], so the phonological categories were arguably as easy to remember as the [-S-P] category in the long term. These results provide an even stronger case for the proposal that grouping vocabulary either randomly or phonologically could facilitate long-term retention.

Several interesting observations were made after listening to the recordings made from those six participants who took part in the think-aloud protocol. First there was an extremely marked difference in recording content between participants. For example, either the participants consistently did nothing other than repeat the Spanish and English words and say "This____ means _____," for the duration of the practice session, or the participants attended to the phonological or orthographical aspects of the word and tried to make some type of sense out of it. That is to say, if the target word was *trueno,* then a participant might say something like, "the letters are almost the same in both words, thunder…there's a storm, true, no?" Notice that when the space is removed from "true, no?" *trueno* is formed. Another example was "*taladro* is drill…that sounds like 'la drill o'…*taladro*…" More than one participant used the double *m* in hammer to help them remember *martillo,* which has a double *l*. The point here is that students looked for similar sounds and spellings to help them remember the different sets of words most

of the time. But, one participant also said, "*mazo, mazo , mazorca…*that's a lot of *m*'s…I'm

confused." This statement might help explain the difference in means between the [-S-P] and the

[–S+P] and [+S+P] groups. Perhaps it is because participants had difficulty distinguishing the

words orthographically in the phonological groups, instead of having difficulty distinguishing the

concepts, as seemed to be the case with the [+S-P] group.

<center>Conclusion</center>

The results of this study support those of Tinkham (1993, 1997), Waring (1997),

Finkbeiner and Nicol (2003), and Papthanasiou (2009). Considering the fact that the nonrelated

group fared best consistently throughout this study, it seems fair to conclude that for novice

learners of Spanish, this method of presentation best facilitated learning in a laboratory setting.

Methodologically, therefore, textbook writers should consider the building evidence in support

of this nontraditional presentation method. Likewise, given the pedagogical implications on

long-term retention, teachers—at least at the novice level—should consider varying current

methods of vocabulary presentation.

 In terms of the phonological categories, this study provides preliminary results that

potentially connect theory pertaining to brain mechanisms used in general L1 novel word

learning to L2 acquisition. Considering the fact that the phonologically grouped words showed a

marked improvement compared to the semantically clustered not phonologically similar words, it

seems fair to conclude that this presentation method calls for further research. Future research

could select a stricter set of semantically linked words that are not so close between languages

which may better distinguish the differences in learnability of the two phonological groups. Also,

sets of semantically linked words could be chosen by learners in a preliminary session. For

example, a group of participants not included in the final sample size could be given several

<center>36</center>

superordinate categories and asked to list the first five words they think of that are associated with that term, thereby creating clusters that learners would consider to be semantically related. It should also be kept in mind that semantic fields differ between languages, meaning that what constitutes a semantic cluster in L1 may not constitute a semantic cluster in L2. These considerations would enable the researcher to create semantic groupings that are clear and sensible to his or her particular population, since it should be understood that semantic fields will vary between them.

New sets of words with phonologically similar endings, instead of beginnings, or with phonological sounds not available in L1 (the sounds "*tin*" and "*ma*" are both present in these participants' L1, but *ñ* would not have been) should be examined in order to see if those similarities in words are noticeable and helpful in vocabulary presentation and acquisition. Research could also explore listening and speaking components to investigate whether or not the phonology is a confounding or clarifying factor in these modes of vocabulary presentation.

This study could also be altered by giving participants a chance to briefly review the material prior to taking the delayed posttest. This might give researchers a better indication of what the participants learned as well as better reflect how assessments are executed in the classroom. Results from studies like these could eventually translate into better understanding of how the brain acquires new vocabulary and where the line is drawn between L1 and SLA. Understanding how vocabulary is best presented could help teachers and learners alike in achieving their shared goal of second language acquisition.

References

Altarriba, J., & Mathis, K. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36*, 550-568.

Baddeley, A. D. (1990). *Human Memory Theory and Practice*. A Division of Simon & Schuster, Inc. Needham Height Mass.

Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal 35*(2), 115-22.

Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 1340-1359.

Finkbeiner, M. & .Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics, 24*(3), 369-83.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychology Bulletin, 123*, 71-99.

Gairns, R., & Redman, S. (1986). *Working with words: A guide to teaching and learning vocabulary.* New York: Cambridge University Press.

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley: New York.

Higa, M. (1963). Interference effects of intralist and word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 2*, 170-175.

Hunt, R. R. & Elliot, J. M. (1980). The role of nonsemantic information in memory: orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General, 109*, 49-74.

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 81-87.

Jullian, P. (2000). Creating word-meaning awareness. *ELT Journal 54*(1), 37-46.

Klein, Raymond, (1999). The Hebb Legacy. *Canadian Journal of Experimental Psychology, 53,* 1-3.

Kroll, J. F., Michael, E., & Sankaranarayanan, A. (1998). A model of bilingual representation and its implications for second language acquisition. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign Language Learning: Psycholinguistic studies on training and retention (pp. 365-395)*. Mahwah, NJ: Erlbaum.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174.

La Heji, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal Memory and Language, 35*, 648-665.

Lehrer, A. (1974). *Semantic Fields and Lexical Structure.* Amsterdam: North Holland.

McGeoch, A. & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology, 43*, 579-588.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1-21.

Mosse, E. K. & Jarrold, C. (2008) Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *The Quarterly Journal of Experimental Psychology,*

61(4)*, 505-514.

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

Press.

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal,*

9, 6-10.

Nida, E. A. (1975). *Componential analysis of meaning.* Mouton: The Hague.

Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT*

Journal, *63*(4), 313-322.

Schneider, V., Healy, A. & Bourne, L. (2002). What is learned under difficult conditions is

hard to forget: contextual interference effects in foreign vocabulary acquisition,

retention, and transfer. *Journal of Memory and Language, 46,* 419-440.

Seal, B. D. (1991) Vocabulary learning and teaching. In M. Celce-Murcia (Ed.), *Teaching*

English as a second or foreign language (2nd ed., pp 296-311). Boston: Heinle & Heinle.

Szmalec, A., Duyck W., Vandierendonck, A., Mata, A., & Page, M. (2009). The Hebb

repetition effect as a laboratory analogue of novel word learning. *The Quarterly*

Journal of Experimental Psychology, 62(3) 435-443.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language

vocabulary. *System, 21*(3), 371-80.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of a second

language vocabulary. *Second Language Research, 13*(2). 138-63.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication.

System, 25(2), 261-74

Table 1

Summary of articles that contribute to the rational for semantic clustering.

Article	Theory	Hypothesis	Implications in teaching	Results
Channell (1981)	Semantic field theory: Vocabulary of language exists in interrelating networks created by linguistic meaning which makes semantically related words geographically closer to each other.	Since the brain categorizes words into semantic fields, presenting them in that way should be helpful to learners.	Componential field analysis: Students analyze sets of words (based on shared/non-shared components of meaning) while comparing and contrasting the various shades of meaning between them. This technique was used in various classroom experiments with intermediate to advanced learners.	Channell stated that students gained a better understanding of the meanings of words and broader vocabulary when they are presented and analyzed in semantic groups.
Jullian (2000)	Cites Channell (1981)	Studying vocabulary that would fall under one broad, classifying term will facilitate vocabulary learning.	Her students also used componential field analysis during classroom activities.	Jullian stated that students gained a broader range of vocabulary pertaining to the target classifying term (to hit).

Table 2

Summary of articles that pertain to semantic clustering

Study	Participants	Tasks	Design	Tests	Results
Tinkha m (1993)	20 NN adult advanced-level English speakers	6 pairs of English/ artificial words	Half of the English words were semantically grouped (Condition 1), half of the words were unrelated (Condition 2). Form A presented English/artificial pairing while Form B presented artificial/English pairing	Trials-to-criterion was met when participant correctly said all three words from a given condition and data was taken until both conditions met the criterion.	Unrelated words required fewer trials to criterion than did related words.
Tinkha m (1997)	48 Adult N English Speakers	6 pairs of English/ artificial words	Condition 1 included "semantic clusters" of words of the same class (nouns). Condition 2 included linguistically unrelated words of the same class (nouns). Condition 3 included "thematic clusters" of words of different classes. Condition 4 included unassociated words of different classes.	Trials-to-criterion tests were employed, including responses in both written and spoken modalities. Once participants reached 100% accuracy twice consecutively during recall/ recognition tests in any given condition, data were obtained and continued to be collected until criterion was met in all conditions.	Data showed that participants took longer to reach criterion when being assessed on semantic clusters versus nonrelated word sets and that participants took fewer trials to meet criterion for "thematically clustered" words than nonrelated words.
Waring (1997)	20 N adult Japanese speakers	6 pairs of Japanese /artificial words	Half of the Japanese words were linked by a single superordinate word (Condition 1, the other half of the words were unrelated (Condition 2). Form A presented Japanese/artificial word paring while Form B presented artificial/Japanese word pairing.	Trials to criterion was met when participant correctly said all three words from a given condition and data was taken until both conditions were met.	Unrelated words required fewer trials to criterion than did related words.
Finkbei ner & Nicol (2003)	24 N Adult English speakers	32 picture/l abel pairs (English/ artificial word)	Half of the pictures were presented in semantic blocks; the other half was presented in a nonrelated fashion.	A translation (bi-directional) task (oral) and a recognition task (written) were employed. Once participants met 80% accuracy on wither modality, criterion was met.	Translation from L1 to L2 took longer in the semantically clustered condition. L2 to L1 translation was also slower for the semantically related condition.
Schnei der, Healy, Bourne (2002)	64 Adult N English speakers	25 English/ French pairs	Words were either blocked (semantically linked) or mixed (unrelated). Participants were trained/assessed in one session and trained/assessed in the second session in order to obtain longer term retention results.	Half of the participants were trained on with French cues, the other half with English cues. During the second session (1 week later) the (re)training took place in reverse order of the precious session for	Participants given L2 cues during Session 1 showed a sizable statistical advantage in producing the semantically related vocabulary. The remaining results show little advantage for either

				half of the participants.	presentation method.
Papath anasiou (2009)	31 N Greek speaking adults/children	60 English/ Greek word pairings	Half of the words were presented in semantic groupings, the other half in nonrelated groupings	After three weeks of training, participants were assessed using paper-and-pencil translation tests (L2 to L1)	Adults scored higher on unrelated groupings of vocabulary than on semantically related vocabulary in both the initial and post test. The children showed no discernable difference in scoring.

N=Native NN=Nonnative

Table 3

Studies showing the Hebb effect in novel word learning.

Study	Participants	Tasks	Design	Tests	Results
Mosse & Jarrold (2008)	45 five-to-six year olds, N English Speakers	8 pictures/ words	Participants were presented both 8 pictures that were associated with words and 8 pictures that were associated with novel words	Participants were prompted with a picture and asked to respond with the word/novel word	Data analysis showed that Hebb learning (serial recall of phonemes) was reliably correlated with nonword verbal learning and not correlated with word learning.
Szmalec, Duyck, Vandierendonck, Mata, & Page (2009)	42 Adult N Dutch speakers	36 sequences of 9 CV pairs	12 of the 36 sequences were repeated (Hebb sequence) in immediate serial recall tests.	Participants were presented with words created from the Hebb sequences and performed a lexical decision task on whether the word they were presented was an authentic word or an artificial one.	As predicted, participants had more difficulty rejecting nonwords made from Hebb sequences as authentic words than other nonwords.

Table 4

Repeated Measures ANOVA: Category and Time

Source of variability	*df*	Sum of squares	Mean squared	F	*p*	Partial eta squared
Category	3	44.781	14.27	101.207	.000*	.766
Time	1	35.168	11.723	19.637	.000*	.390
Category*Time	3	12.574	4.191	7.711	.000*	.199

*p < .05

Table 5

Descriptive Statistics for Average Correct Over Time

Category	Mean	SD
[+S-P]	.719	.121
[-S-P]	1.75	.147
[-S+P]	1.203	.166
[+S+P]	1.375	.142

Note: N=32; Min=0; Max=5

Table 6

ANOVA Results for Category in the Immediate Posttest and Delayed Posttest

	df	Sum of squares	Mean squared	F	*p*	Partial Eta squared
POS	3	44.781	14.927	17.414	.000*	.360
DL	3	2.961	.987	3.559	.017*	.103

*p < .05; POS = Immediate Posttest; DL= Delayed Posttest

Table 7

Average Correct for Each Category in the Immediate Posttest

Category	Mean	SD
[+S-P]	1.375	.226
[-S-P]	3.031	.252
[-S+P]	2.156	.281
[+S+P]	2.375	.245

N=32; Min=0; Max=5

Table 8

Average Correct for Each Category in the Delayed Posttest

Category	Mean	SD
[+S-P]	.06	.246
[-S-P]	.47	.671
[-S+P]	.25	.672
[+S+P]	.38	.554

N=32; Min=0; Max=5

Figure 1

Experimental Design

Background questionnaire, pretest

↓↓ (Approx. 1 week) ↓↓

Introduction and practice of 20 target words
Immediate posttest

↓↓ (Approx. 2 weeks) ↓↓

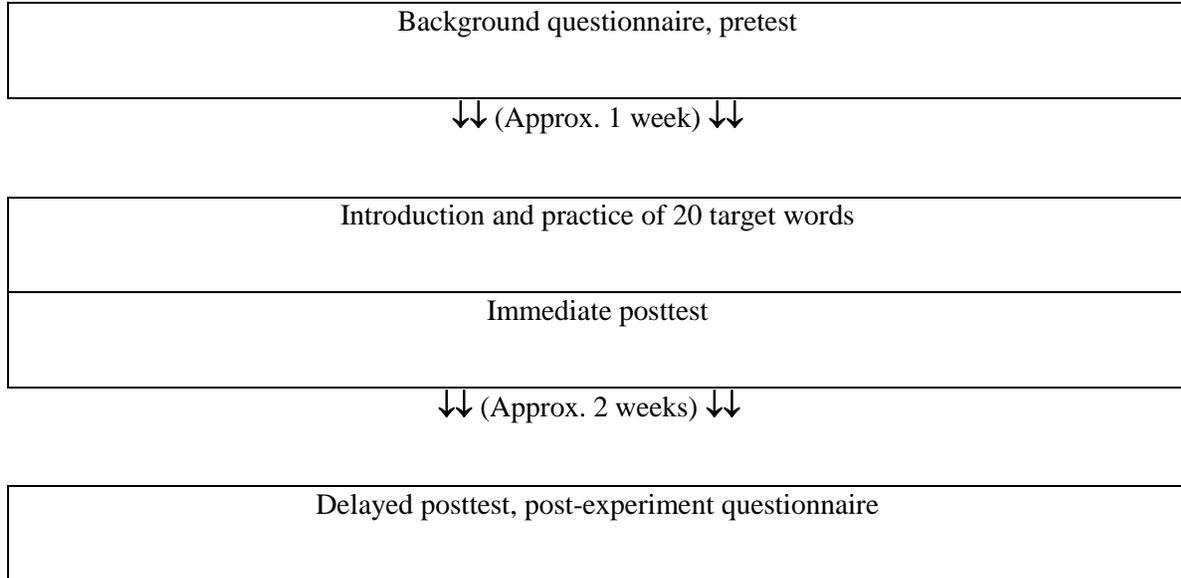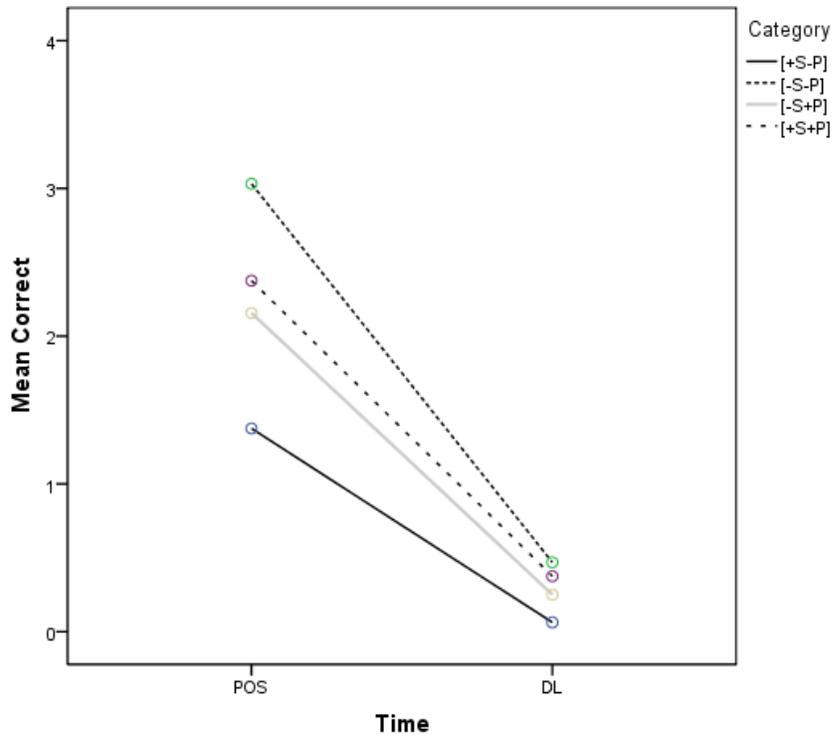| Delayed posttest, post-experiment questionnaire |
| --- |

Figure 2

Plot of Average Correct Over Time

Note: POS = Immediate Posttest; DL = Delayed Posttest

Appendix 1

Word categories:

<u>Practice Segment 1:</u>   **[+S-P]**

Drill – taladro   A *taladro* is a drill.

Hammer - martillo A *martillo* is a hammer.

Chisel - escoplo   An *escoplo* is a chisel.

Rachet - carraca   A c*arraca* is a ratchet.

Saw – serrucho   A *serrucho* is a saw.

<u>Practice Segment 2:</u>   **[-S-P]**

knuckle – nudillo   *Nudillo* means knuckle.

broom – escoba   *Escoba* means broom.

thunder – trueno   *Trueno* means thunder.

deer – venado   *Venado* means deer.

sword- espada *Espada* means sword.

<u>Practice Segment 3:</u>   **[-S+P]**

inkwell – tintero   *Tintero* means inkwell.

cheap red wine – tintorro   *Tintorro* means cheap red wine.

water tank – tinaco   *Tinaco* means water tank.

large earthenware jar – tinaja   A *tinaja* is a large earthenware jar.

wood pile – tinada   *Tinada* means wood pile.

<u>Practice Segment 4:</u>   **[+S+P]**

mace – maza   *Maza* means mace.

mallet – mazo   *Mazo* means mallet

dungeon -- mazmorra   *Mazmorra* means dungeon

torture – mazorca*   *Mazorca* means torture

thump/blow – mazazo   *Mazazo* is a thump or blow

*This word can be used in this way in Argentina according to two sources I have found. This particular meaning is derived from a Rosista military group named Mazorca.

Appendix 2

Background Questionnaire

Directions: Answer every question. If a question does not apply to you, write "N/A".

1. Participate ID#_____
2. Age_____
3. Sex: M / F
4. Class year: Freshman / Sophomore / Junior / Senior
5. Major_____
6. Minor_____
7. What do you consider your native language(s) to be? _____
8. How old were you when you were first exposed to Spanish? _____
9. On a scale from one to seven, seven being *excellent*, please select your level of proficiency in terms of *understanding spoken Spanish by a native Spanish speaker.*

    1   2   3   4   5   6   7

10. High school Spanish (circle all that apply or add to the category *other* if applicable. If you took no Spanish classes while in high school, circle N/A here):

    Spanish I   Spanish III   Spanish V

    Spanish II   Spanish IV   AP Spanish/IB Spanish

    Other: _____

11. College Spanish. Indicate if you have previously taken this Spanish class during college.

    _____ Elementary I

12. Use of Spanish outside of Spanish class during high school years:

13. Travel to Spanish-speaking countries:

Country:	Length of time:	Reason (e.g. vacation, Spanish language course, dad worked there):	Use of Spanish while in the country:

14. Other languages studied or spoken:

Language:	Length of time studied or spoken:	Use of the language (in what ways do you or did you used to engage in the language?):

15. Other language experience not mentioned above:

Appendix 3

Post-experiment questionnaire

Participant ID:_____

Thank you so much for participating in this experiment! Please take a few minutes to fill out the following questionnaire regarding your experience and opinions. Please be as candid as possible. Your responses are an important part of this research and are greatly appreciated!

1. Did you notice any patterns in the list of words?_____

    What pattern(s)?_____

2. Did you feel like some words were more difficult to learn than others? _____

  To assist you with remembering the words, here is a list of those you learned:

deer – venado   hammer - martillo chisel - escoplo   rachet - carraca

Saw – serrucho   knuckle – nudillo   broom – escoba sword- espada wood pile – tinada

inkwell – tintero cheap red wine – tintorro   water tank - tinaco

large earthenware jar – tinaja drill – taladro thunder – trueno   maza- mace

mazo- mallet mazmorra- dungeon mazorca* - torture   mazazo- thump/blow

    If yes, which words seemed more difficult?_____

    Why do you think they were more difficult?_____

    _____

3. Were some words easier to learn than others?_____

      If yes, which seemed easier?_____

Why do you think they were easier?_____

_____

4. Do you think any of these groupings helped you learn the vocabulary words?_____

   If yes, explain which groupings and how they helped you. _____

   _____

   _____

5. Did you look up the meaning of any of the words outside of the laboratory?_____

6. Did you ask a teacher, friend, or anyone about these words?_____

*The word mazorca generally means corn cob, but in Argentina, this word also can also mean

torture.