

# Second Language Research

<http://slr.sagepub.com>

---

## The effects of semantic and thematic clustering on the learning of second language vocabulary

Thomas Tinkham

*Second Language Research* 1997; 13; 138

DOI: 10.1191/026765897672376469

The online version of this article can be found at:  
<http://slr.sagepub.com/cgi/content/abstract/13/2/138>

---

Published by:

 SAGE Publications

<http://www.sagepublications.com>

Additional services and information for *Second Language Research* can be found at:

**Email Alerts:** <http://slr.sagepub.com/cgi/alerts>

**Subscriptions:** <http://slr.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

## The effects of semantic and thematic clustering on the learning of second language vocabulary

Thomas Tinkham *University of Guam*

ESL students are often presented much of their new English vocabulary preorganized for them in 'semantic clusters', sets of semantically and syntactically similar words, e.g., *eye, nose, ear, mouth, chin*. Although clustering of this sort facilitates the activities which serve current approaches to language teaching and would seem, at first glance, to facilitate vocabulary learning as well, little or no empirical justification is offered by researchers in support of its employment. In fact, research that might apply, psychological research generated by interference theory, would predict that such clustering of similar items impedes rather than enhances learning.

On a more positive note, a more 'thematic' manner of organizing new L2 vocabulary is suggested by more recent psychological research which would predict that clusters like *frog, green, hop, pond, slippery, croak* would be more easily learnt than groups of unassociated words.

With these predictions in mind, the reported research explored the effects upon L2 vocabulary learning of both 'semantic' and 'thematic' clustering. The results provide a wide range of evidence that suggests that semantic clustering does indeed serve as a hindrance while thematic clustering serves as a facilitator of new language vocabulary learning.

Judging by a recent crop of ESL textbooks, it appears that many if not most ESL students are often exposed to their new language vocabulary preorganized for them in semantic clusters, that is, groups of words which share certain semantic and syntactic similarities. These groupings, for example, *ear, eye, nose, mouth* are referred to by some academics as 'lexical sets' (e.g., Gairns and Redman, 1986) while others (e.g., Marzano and Marzano, 1988) employ the label 'semantic clusters', the term I prefer as it seems to convey more clearly the connection between 'semantic clusters' and 'semantic fields', a concept involving semantic similarity which appears frequently in the literature generated by the study of lexical semantics (see Kittay and Lehrer, 1992).

Such semantically based clusters are often encountered in current, well distributed ESL textbooks. Two examples selected

Address for correspondence: Department of English and Applied Linguistics, University of Guam, Mangilao, Guam 96923, USA.

© Arnold 1997

0267-6583(97)SR104OA

from many which would serve equally well are the cluster *street, road, avenue, boulevard, drive and trail* from *Real-life English 3* (Jolly and Robinson, 1988: 17), and the cluster *bank teller, cashier, clerk, manager, receptionist, secretary, supervisor and typist* from *Interchange 1* (Richards, 1990: 9). As can be seen, the words within each cluster fall under a common superordinate or covering concept (be it labelled or not) and are consequently gathered together as a result of their shared semantic and syntactic characteristics.

Two motivations appear to drive this clustering of semantically related vocabulary items. First, semantic clusters service the methodologies driven by two current approaches to L2 development. For curriculum programmers of a structure-centred persuasion, semantic clusters fit quite naturally into the 'slots' left open in their oral and written substitution activities. *Side by side* (Molinsky and Bliss, 1989: 64), for example, presents *red, orange, yellow, green, blue, purple* and seven more colour labels as possible fillers in the sentence, 'But this is a \_\_\_\_\_ jacket'. In a similar fashion, *New horizons in English 1* (Walker, 1991: 41) offers *carrots, nuts, grapes, pears, peaches, oranges* and three more food labels as possible ways to complete the question, 'Do you like \_\_\_\_\_?'

Following a more learner-centred approach to L2 development, other ESL programmes are guided more clearly by their writers' perceptions of the communicative needs of their students. Such programmes are typically divided into various units responding to either situations in which students might find it necessary to communicate in their new language (e.g., visiting a doctor), or dealing more closely with the language itself, the notions (expressions of time, location) and functions (e.g., requests) which students may wish to communicate (Wilkins, 1976). Although this needs-based approach may appear essentially different from a more language-based approach, following the approach, new words are still often arranged in semantic clusters. The words of these clusters fit, again quite naturally, within the situations, notions or functions being considered. For example, *In contact 2* (Lavie *et al.*, 1991: 55) clusters *bag, bottle, can, cup, glass and dish* as examples of containers in a unit titled 'Are you hungry?' Furthermore, Wilkins himself clusters *tell, inform, proclaim, publish, assert, declare, state, emphasize, argue, know, affirm, maintain, advocate, claim, content and protest* as examples of words serving the notion, 'information, asserted', and suggests that an 'appropriate selection' from these items would fit within a unit of a notional syllabus (Wilkins, 1976: 49).

A second reason prompting the use of semantic clustering of new vocabulary items is the belief that such clustering facilitates the

learning of new words. Gairns and Redman (1986: 31) summarize arguments supporting this belief when they state that grouping words according to their semantic features can 'provide a useful framework for the learner to understand semantic boundaries: to see where meaning overlaps and learn the limits of use of an item'. The idea is that learning the concept 'spoon' is facilitated by learning how spoons are both similar to and different from knives and forks.

The problem, however, is this: although semantic clusters fit quite nicely into two current ESL methodologies and even though they facilitate focus upon semantic similarities and differences among words being learnt there is little or no direct empirical evidence that semantic clustering does in fact facilitate the learning of L2 vocabulary. Moreover, and quite to the contrary, a large body of evidence exists which suggests just the opposite, that the clustering of semantically and syntactically similar words actually inhibits vocabulary learning.

Delving into the psychology research of the first half of this century, one finds an enormous body of literature dedicated to the study of 'interference' and how this phenomenon affects learning and memory. In essence, 'interference theory' postulates that as the similarity between information intended to be learnt and information learnt either before or after that information increases, the difficulty of learning that information also increases. Hundreds of studies supporting and expanding the theory have produced findings which would lead readers to predict that near-simultaneous presentation of semantically and syntactically similar words, e.g., *knife*, *fork* and *spoon*, would impede rather than facilitate the learning of all three (for examples, see Crowder, 1976).

More recently, researchers have studied the 'flip-side' to interference theory and, not surprisingly, have arrived at similar conclusions. The 'distinctiveness hypothesis' (see Hunt and Elliott, 1980; Hunt and Mitchell, 1982) postulates that as the distinctiveness (nonsimilarity) of information to be learnt increases, the ease of learning that information also increases. While not producing anything approaching the size of the body of literature in support of interference theory, these researchers do provide further and more current evidence that should lead one to suspect that the semantic clustering of relatively nondistinct vocabulary items is, in fact, a hindrance to the learning of those items.

The results of research guided by interference theory and the distinctiveness hypothesis, however, should not lead one to believe that clustering *per se* inhibits the learning of new L2 words. The possibility that grouping words in a different way may actually

facilitate learning is raised by the work of lexical semanticists currently investigating hypothesized manners by which speakers of a language subconsciously organize their 'mental lexicon', i.e., the words they know (see Kittay and Lehrer, 1992). While these semanticists would recognize semantic clusters as sets of words drawn from 'semantic fields' within which speakers of a language may organize lexemes in accordance with interlexical similarities, many would also posit that words might be subconsciously organized in accordance with their participation within certain 'frames' or 'schemas', concepts which segmentize a speaker's background knowledge (see Fillmore, 1985). A cluster of words drawn from such a frame or schema might include *frog, pond, hop, swim, green* and *slippery*, words of different parts of speech that are all closely associated with a common thematic concept, in this case, *frog*. Based, as they are, on associative strength, clusters of this sort are cognitively rather than linguistically derived and consequently would appear to fit most easily into learning-centred SLA programmes which are more concerned with learning processes than with linguistic analysis. Not surprisingly, the best example of cognitively based clustering that I can find comes from *Coast to coast 3* (Harmer and Maybin, 1989: 67) of the learning-centred Longman series in which the words *haunted, moonlight, yell, ghost* and *groan* are selected for instruction from a short passage about a haunted house.

Clustering such fright-associated words within a haunted house 'schema' may quite possibly facilitate learning if, as Brewer and Nakamura (1984) claim, schema-related material is more easily learnt than schema-unrelated material, or if, as Anderson and Pearson (1984) contend, mentioning a schema will bring to mind the individual components of that schema.

The possibility, therefore, arises that while most L2 students are struggling to learn new words which have been selected and presented to them in a manner that impedes learning, a different manner of selection and presentation might actually make learning easier.

In an attempt to explore this possibility, this study examines the effects of 'semantic' and 'thematic' clustering upon the learning of new L2 words. The two labels were chosen to fit the terminology associated with distinctions central to the study of lexical semantics (see Markman and Hutchinson, 1984; Kittay and Lehrer, 1992) and are intended to distinguish between two different manners of organizing lexical items: 'semantic clustering' based upon semantic and syntactic similarities among clustered words and 'thematic clustering' based upon psychological associations between clustered

words and a shared thematic concept. While the two categories of clusters are not always mutually exclusive, that is, the members of some clusters might be both semantically and thematically related, many clusters are easily perceived as clear examples of one or the other type. For example, one would intuit that [*apricot, peach, plum, nectarine, pear, apple*] is a different sort of cluster from [*frog, pond, swim, hop, green, slippery*].

Even though these two manners of clustering are identifiable and are both employed in L2 vocabulary instruction, research into the effects of clustering upon L2 vocabulary learning is limited and indirect at best. While a large number of studies generated by interference theory have explored the effects of stimulus and response similarities upon learning and memory, writers interested in L2 learning have not extended this research directly to the learning of L2 vocabulary (for exceptions, see Higa, 1963, for a report of a study of word learnability and Tinkham, 1993, for a report of a study anticipating this research). Nor have researchers explored directly the effects of theme- or schema-based clustering upon L2 vocabulary learning. Given this lack of attention, the overall goal of this study is that of stimulating research into the effects of clustering upon the learning of new L2 vocabulary.

More specifically, however, the intention of this research is to investigate the effects of both semantic and thematic clustering upon the ease with which new L2 vocabulary items are learnt. Given this intention, the study addresses two questions:

- 1) Do L2 students learn semantic clusters of new L2 words with more difficulty than they learn sets of unrelated words?
- 2) Do L2 students learn thematic clusters of new L2 words more easily than they learn sets of unassociated words?

The research consisted of two different experiments, each examining the effects of pedagogical clustering upon L2 vocabulary learning and each sharing the same subjects and similar procedures.

## **I Experiment 1 (mixed)**

### *1 Subjects*

A total of 48 subjects participated in the study. An additional 10 subjects participated in pretesting or dropped out of the research for personal reasons. Half the subjects participated in the studies administered orally and the other half participated in the studies conducted in the written modality. The subjects were all students enrolled in sophomore-level educational psychology courses at a

large mid-western American university and participated in the study in order to fulfil a requirement for a class they were taking. All the subjects were native speakers of English.

## 2 Materials

Experiment 1 consisted of four separate studies, two conducted in the oral modality and two conducted in the written modality. Of the two studies conducted in a particular modality, one study involved the task of recognizing new artificial words and the other involved the task of recalling artificial words given their English meanings. The purpose of each study was to compare, by the independent variable 'condition', the learnability of three-pair sets of associate pairs pairing English and new, artificially created words.

'Condition', the independent variable of interest, was nominally divided according to the following classification:

- *Condition 1*: linguistically related 'semantic clusters': words of the same form-class which directly descend as co-ordinates under a common superordinate concept.
- *Condition 2*: linguistically unrelated sets: words of the same form-class which do not directly descend from a common superordinate concept.
- *Condition 3*: cognitively associated 'thematic clusters': words of different form-classes that, in accordance with the intuition of the author, were likely to be associated with a shared thematic concept.
- *Condition 4*: cognitively unassociated sets: words of different form-classes that were judged not to be associated with a shared thematic concept.

## 3 Two modalities

Half (24) the subjects took tests orally, responding to a word they heard by saying their response. The other half took tests in the written modality, responding to a word they saw by typing their response. Thus, 'modality' was a between-subject variable.

## 4 The two studies administered in the oral modality

The materials for each of the two oral studies consisted of two trials-to-criterion tests each involving a six-pair set of associate pairs pairing an English word with an artificial word. Each six-pair set included three English words from one condition and three from the opposing condition, the six word-pairs mixed together in a

variety of orders across trials. Guided by the research questions, both the recognition and recall studies consisted of one six-pair set which included three semantically related and three unrelated English words and another six-pair set which included three thematically associated and three unassociated English words.

*a English words:* All the English words in the sets were selected in accordance with the four conditions described above and, given the likelihood that the learnability of a new L2 word is affected by the frequency with which its L1 counterpart is used (see Rodgers, 1969), all the English words chosen are ranked between the 1000th and 5000th most frequently encountered words in English by the *Word frequency book* (Carroll *et al.*, 1971). Furthermore, given the possibility that the learnability of a new L2 word is affected by the form-class of the L1 word with which it is associated (see Rodgers, 1969), the form-classes of the English words were held constant across contrasting sets.

While the groups of words that are semantically related may also be thematically associated in some way, the composition and consequently the effects of the two groupings are noticeably different. The semantic clusters consist solely of words of a common word-class (in the case of this research, nouns) while thematic clusters consist of a noun, a verb and an adjective. Also, the words in the thematic clusters were carefully selected for their association with a shared thematic concept.

The English words, arranged by condition, were:

*Condition 1, semantic clusters*

dish            shirt  
bowl          jacket  
plate         sweater

*Condition 2, unrelated sets*

acid           island  
smoke        potato  
roof          beard

*Condition 3, thematic clusters*

beach        library  
sunny        whisper  
swim         quiet

*Condition 4, unassociated sets*

fork          triangle  
count        improve  
brave        sweet

*b Artificial words:* The artificial words which were paired with the English words listed above were created by the researcher. In order to decrease the possibility that one set of artificial words may be, for phonological reasons, more easily learnt than another, the following guidelines were used for the creation of artificial words:

- All artificial words must have two syllables.
- All six-word sets must employ a variety of sounds.
- Within all sets of three artificial words
  - one word must begin with a vowel;
  - another word must end with a vowel;
  - one word must contain a consonant cluster; and
  - either one or two words must place stress on the first syllable.

The sets of artificial words created for Experiment 1, oral modality, were:

oosmid	heegeh	mofay	meykoo
bovahp	dusahh	jetuhn	ejosk
loshae	aeplawk	estaep	fihawp
driguh	bemouf	ushahd	blido
awtej	ayket	sawrik	uhneeb
hanees	grivah	veeksoo	saygahv

*c Reordering of word-pairs by trial:* Since the learning performance of the subjects under the various conditions was measured on a trials-to-criterion basis, for each test the ordering of word-pairs within a trial was changed with each new trial through six trials after which the order returned to the original and the sequence began again. The ordering of pairs within the trials was not random; rather, the researcher altered the ordering of the pairs throughout the trials so that each pair fell in the various positions possible within the set. The ordering also ensured that the last word-pair on one trial did not become the first pair on the next trial.

*d Two different 'tasks':* Each subject was required to hear and recognize the artificial words and say the corresponding English word on two of the tests and hear the English words and recall (say) the corresponding artificial word on the other two tests. Thus, 'task' was a within-subject independent variable.

*e Counterbalancing by 'form':* To counteract the possibility that some pairings of English and artificial words would somehow suggest an obvious connection between the two words of a pair, half the subjects learnt one particular pairing of English and artificial words (Form A) while the other half of the subjects learnt a different pairing (Form B). For both forms, however, the English and artificial words were the same, the only difference being that the English and artificial words were paired differently for the two

forms. Thus, 'form' was a between-subject, counterbalanced independent variable.

*f Counterbalancing by 'order':* Half the subjects who took a particular form of the two tests (Form A or Form B) took those two tests (semantic cluster/unrelated set; thematic cluster/unassociated set) in a particular order, while the other half taking the same form took the two tests in the reverse order. Thus, 'order' also was a between-subject, counterbalanced independent variable.

##### 5 *The two studies administered in the written modality*

The materials, that is the English/artificial word-pairs employed in the two studies administered in the written modality, were virtually the same as those administered in the oral modality. The English words were exactly the same and the artificial words were nearly the same with the only exceptions arising from additional constraints placed upon the spelling of the artificial words, such constraints being applied in an attempt to standardize the typing tasks of the subjects across the four tests. These additional constraints were:

- All words must have five letters.
- No words may contain double letters.

The sets of artificial words employed in the two studies administered in the written modality were:

usmid	hijeh	mofay	meyku
bovap	dusan	jetun	ejosk
losha	aploc	estap	fihap
drigu	bemof	ushod	blido
awtej	eiket	sarik	uhnib
hanis	griva	viksu	sagov

*a Reordering of items by trial:* As in the studies conducted orally the ordering of the six items within trials was altered, each ordering being different from the ones preceding it through a set of six trials after which the ordering procedure began again with the first ordering.

*b Two different 'tasks':* Paralleling tests in the oral modality, each subject was required to see and recognize the artificial words and type the corresponding English word on two of the tests and see

the English words and recall (type) the corresponding artificial word on the other two tests. Thus, 'task' was a within-subject independent variable.

*c Counterbalancing 'form' and 'order':* The independent variables 'form' and 'order' were counterbalanced in the same fashion as they were for the two studies administered orally.

## 6 Procedures

Subjects were tested individually in two different sessions separated by approximately two weeks. The first session involved the production of English words (recognition of new words) while the second session involved the production of the artificial words (recall of new words). Both sessions included the two tests from Experiment 1 as described above followed immediately by the four tests from Experiment 2 as described below.

### 7 The two studies conducted in the oral modality

*a Recognition tests:* Two trials-to-criterion tests were individually administered to 24 subjects, the subjects hearing artificial words and responding by saying the English words which represented the meanings of those new words. One of the tests mixed three semantically related English words with three unrelated English words while the other test mixed thematically associated words with unassociated words. The tests were administered via a recording played on a tape-recorder and subjects were provided a minimum of elaboration by the researcher whose role was limited to that of data recorder marking on a checksheet whether a given response was acceptable or not.

Upon beginning the session, subjects first heard an introduction which explained the purposes of the research and procedures of the testing. This introduction was followed by a short, three-pair sample test resembling the tests subjects were to encounter in the regular experiment.

After the practice test the subjects took the first of the two tests. In preparation for this test the subjects first heard a modelling, in no particular order, of the six word-pairs to be learnt, each item phrased in the form of '*Moshee* means *shirt*'. Following this modelling subjects heard, one by one, the artificial words from the pairs and then correctly said (or did not say) the corresponding English word. More specifically, for each item subjects first heard the artificial word followed by a two-second pause, the sound of a

bell and finally the appropriate English word. Subjects were required to say or at least begin saying the appropriate English word during the two-second pause. Items within a trial were separated by a four-second pause and trials within a test were separated by a six-second pause. Criterion was met when a subject correctly said each of the three words representing one of the conditions on two consecutive trials. Once criterion had been met for one condition, that trial was silently marked by the researcher and the trials continued with all six word-pairs until the subject reached criterion for the second condition. Upon completion of the first test, subjects moved on to the second.

*b Recall tests:* Recall tests were given in a second session approximately two weeks after the session involving recognition tests. The two recall tests resembled the recognition tests except that subjects learnt new sets of English artificial word-pairs, were given the English words as stimuli and responded by saying the artificial words during a three-second pause. Word stress was not considered phonemic and certain variations in vowel coloration were accepted if they appeared to be fossilized within the 'dialect' of a subject. Again, criterion was met when a subject correctly said all three of the words comprising a condition on two consecutive trials.

#### *8 The two studies conducted in the written modality*

*a Recognition tests:* A different group of 24 subjects took the written tests. The tests administered in the written modality paralleled those in the oral modality but were administered visually via PC computer rather than orally via tape-recorder. Subjects first read from the computer monitor a short introduction to the research and the testing procedures of the first session followed by a short, three-item sample test. Upon completion of the sample test, subjects selected from a short menu the 'form' they had been assigned and took the two tests in the order assigned them.

The tests themselves were visual parallels of the oral tests and employed the same materials as well as the same overall procedures. Subjects first saw a list of the six word-pairs for a period of 20 seconds. The subjects then saw, one by one, the artificial words from the list and were allowed five seconds to type correctly the corresponding English words. Each trial of six items was followed by another trial of the same six items in a different order until the subjects had correctly typed all six English words on two

consecutive trials. After completion of the first test, subjects immediately took the second.

*b Recall tests:* The two recall tests were administered in a second session for each subject about two weeks after the first. The recall session resembled the recognition session but employed different words and presented subjects with English words as stimuli to which they responded by typing the corresponding artificial words. Again, the tests were trials-to-criterion requiring two consecutive perfect trials to reach criterion. The subjects were assigned the same 'form' and 'order' as they were for the recognition tests and took the second test immediately after the first.

#### 9 Design and analysis

A mixed factorial design was employed for analysis of the data provided by the two tests (semantic cluster/unrelated set, thematic cluster/unassociated set) within each of the four studies (oral recognition, oral recall, written recognition, written recall). The factor of interest, 'condition', was a within-subject factor as was 'task' (recall or recognition). Between-subject factors were 'modality' (oral or written), 'form' (alternate pairings of English and artificial words) and 'order' (alternate ordering of the two tests within a session). Both 'form' and 'order' were counterbalanced to avoid confounding of effects. The measures of the dependent variable were scores on trials-to-criterion tests.

A mixed MANOVA was employed to determine main effects for the five independent variables as well as any interactions among them. Most importantly, the significance of the differences between contrasting conditions (condition 1 vs. condition 2 and condition 3 vs. condition 4) were also established. Finally, interactions between the contrasts of interest and the independent variables other than 'condition' were established and interpreted.

#### 10 Individual performances

Nonaggregated individual performances were examined and interpreted in an attempt to gain insight into the data. More specifically, the data provided by individual subjects were analysed in order to determine how often subjects learnt a set of interest (semantic cluster or thematic cluster) in more trials, fewer trials or the same number of trials as the opposing set (unrelated set or unassociated set).

**II Experiment 2 (separated)**

The subjects, procedures and design for Experiment 2 were the same as those for Experiment 1. The materials, i.e., words, however, were somewhat different.

Resembling Experiment 1, Experiment 2 consisted of four studies, two conducted orally and two parallel studies conducted in the written modality. Consequently, as in Experiment 1, the materials consisted of six-pair sets of associate pairs pairing English and artificial words, the difference being that while in Experiment 1 the six-pair sets included a mixture of three English words representing one condition and three representing another, in Experiment 2 the six-pair sets were constructed homogeneously, each set employing English words representing only one particular condition. Thus, each of the four studies within the experiment involved the testing of four different word-pair sets, each set reflecting one of the four conditions of interest.

*1 The two studies conducted in the oral modality*

The materials for each of the two oral studies consisted of four tests, each of the tests involving one six-pair set of English/artificial word-pairs representing one of the four conditions of interest.

*a English words:* All the English words were selected in accordance with the four conditions described in reference to Experiment 1 (see above). Given the possibility that the familiarity of a particular English word will affect the learnability of a new word with which it is associated (see Rodgers, 1969), the ranking by frequency of the six English words within a set was standardized in the following manner. Within each set of six English words, one word was ranked among the 1000 most frequently encountered words in English, three were ranked between the 1000th and the 5000th most common words, and two had a ranking of over 5000 according to the *Word frequency book* (Carroll *et al.*, 1971). Furthermore, given the possibility that the learnability of a word is affected by its form-class (Rodgers, 1969), the number of words from a particular form-class was held constant across contrasted sets. All six of the English words in both the semantic clusters and the unrelated sets (conditions 1 and 2) were nouns while within both the thematic clusters and the unassociated sets, two words were nouns, two were verbs and two were adjectives.

The English words included in Experiment 2, arranged by condition, were:

*Condition 1, semantic clusters*

tin	apple
bronze	pear
iron	nectarine
brass	peach
lead	apricot
steel	plum

*Condition 2, unrelated sets*

cigar	paint
wolf	funeral
lace	recipe
stone	market
chain	uncle
fuel	ice

*Condition 3, thematic clusters*

frog	cave
hop	explore
slimy	dark
pond	stalactite
croak	drip
green	cool

*Condition 4, unassociated sets*

cloud	hill
erase	behave
social	stubborn
office	menu
lose	serve
risky	brief

*b Artificial words:* In order to decrease the possibility that one set of artificial words might be, for phonological reasons, more easily learnt than another, the following guidelines were established for the creation of artificial words:

- All artificial words had to have two syllables.
- All sets had to employ a variety of sounds.
- Within all sets of six artificial words
  - three words had to place stress on the first syllable while the other three had to place stress on the second syllable;
  - two words had to begin with a vowel;
  - two other words had to end with a vowel;
  - one word had to contain a consonant cluster; and
  - one word had to contain a vowel diphthong.

The eight sets of artificial words created for Experiment 2, oral modality, were:

aifoht	laebo	neylor	pofahn
tisahl	neesayr	kawvaes	iddek
sokae	taivoo	mochlee	airij
chuhnda	eduhp	bahkel	veehaws
ojin	juhmahk	uhtawf	graetoo
mawzeek	awfrin	oomao	lehdow
ufahsp	aekawd	awfeb	nugaw
paujul	voolahp	koonahp	deepaht
gawrov	roteh	jaesowd	belooc
imeed	mizoy	veklew	kishen
maetow	aybruhg	uhmil	owdley
soobih	berseen	teepoy	uhkess

*c Counterbalancing by 'order':* Again, as in Experiment 1, the ordering of the tests within a study was counterbalanced. In Experiment 2, however, since there were four tests within each study, four orders were established. The 24 subjects taking the tests in a particular modality were divided into four groups of six subjects, each group taking the four tests in one of the four different orders. The four different orders were derived in accordance with a balanced Latin square design.

## 2 *The two studies administered in the written modality*

As was the case in Experiment 1, the materials presented in the written modality were virtually the same as those administered in the oral modality. The English words were exactly the same and the artificial words were nearly the same with the only exceptions arising from additional constraints placed upon the spelling of the artificial words. These additional constraints were:

- Within all sets of six artificial words
  - one word had to be written with four letters;
  - three words had to be written with five letters;
  - two words had to be written with six letters; and
  - two words had to contain double letters.

The eight sets of artificial words employed in the two written sections were:

eifot	laabo	nelor	pofahn
tisall	nessir	kawvas	iddek
sokae	taivu	moslee	airij
cunda	edup	bakel	hevoss
ojin	juhmak	ottaf	gratu
mazeek	ofrem	umao	ledo
ufasp	akod	affeb	nuga
paujul	voolap	kunop	depat
gurrov	rotti	jassod	belooc
imeed	mozoy	veklew	kisen
mato	ebrug	umil	owdley
subih	bersen	tepoy	ukess

## 3 *Qualitative analysis*

Following completion of the tests comprising each session, each subject was asked a set of questions in order to elicit qualitative data that might improve insight into the quantitative data and

analyses. The questions posed to each subject were:

- 1) Which set did you find to be the most difficult?
- 2) Why do you think it was difficult?
- 3) Which set did you find to be the easiest?
- 4) Why do you think it was easy?

### III Results and discussion

#### 1 Experiment 1 (mixed)

*a Descriptive and inferential statistics:* To establish main effects and make the appropriate comparisons incorporating all the data provided by the experiment, the entire data set was combined to include both modalities (oral and written) and both tasks (recognition and recall). Using these data, main effects for 'condition', 'modality', 'task', 'order' and 'form' were established employing a 4 (condition: semantic cluster, unrelated set, thematic cluster, unassociated set)  $\times$  2 (modality: oral vs. written)  $\times$  2 (task: recall vs. recognition)  $\times$  2 (order: semantic cluster/unrelated set first vs. thematic cluster/unassociated set first)  $\times$  2 (form: form A vs form B) mixed multiple analysis of variance (MANOVA) with 'condition' and 'task' within-subject variables and 'modality', 'form' and 'order' between-subject variables. The dependent variable was the number of trials required to reach the criterion of two consecutive perfect trials on a given test.

A significant main effect was found for condition,  $F(3,120) = 10.26$ ,  $p < 0.001$  and for task,  $F(1,40) = 26.37$ ,  $p < 0.001$ . No significant effects were found for modality, order or form.

No significant interaction was found between condition and modality; however, a significant interaction was found between condition and task,  $F(3,120) = 3.26$ ,  $p < 0.024$ , the difference among the conditions being greater in the recognition tests than in the recall tests. No significant interaction was found between condition and form; however, a significant interaction was found between condition and order,  $F(3,120) = 11.74$ ,  $p < 0.001$ , the difference among the conditions being greater in the tests administered first than in the tests administered second. Given the counterbalancing of the design, this interaction had no effect on the interpretation of the comparisons of interest.

To compare effects for the four conditions, data for the five factors were analysed using the MANOVA described above, the contrasts of interest being condition 1 vs. condition 2 (semantic cluster vs. unrelated set) and condition 3 vs. condition 4 (thematic cluster vs. unassociated set). A significant difference was found

between conditions 1 and 2,  $F(1,44) = 57.83$ ,  $p < 0.001$ ; the semantic cluster requiring more trials to reach criterion than the unrelated set, as well as between conditions 3 and 4,  $F(1,44) = 14.50$ ,  $p < 0.001$ ; the thematic cluster requiring fewer trials to reach criterion than the unassociated set.

No significant interactions were found between comparison 1 (semantic cluster vs. unrelated set) and either modality, task, order or form. No significant interactions were found between comparison 2 (thematic cluster vs. unassociated set) and either modality, order or form; however, a significant interaction was found between comparison 2 and task,  $F(1,40) = 7.20$ ,  $p < 0.011$ , with the ease of learning the thematic cluster more pronounced in the recognition tests than in the recall tests.

Table 1 is a means table for Experiment 1. The table organizes means in a fashion which illuminates the variables 'condition', 'task' and 'modality'. Since 'order' and 'form' are counterbalanced nuisance variables, they are not represented in the table.

*b Discussion:* The data presented by Experiment 1 present a clear indication that new L2 vocabulary items arranged in semantic clusters are learnt with more difficulty than new vocabulary items learnt in unrelated sets. The comparison between the alternate methods for grouping new L2 words is clear not only in light of the

**Table 1** Means and standard deviations for scores by condition, modality and task for Experiment 1

	Oral	Written	Combined
<i>Condition 1 (semantic cluster)</i>			
Recognition	8.25 (4.74)	5.17 (2.32)	6.71 (4.01)
Recall	7.50 (2.92)	7.50 (3.67)	7.50 (3.28)
Combined	7.88 (3.46)	6.33 (2.48)	7.10 (3.08)
<i>Condition 2 (unrelated set)</i>			
Recognition	5.58 (2.57)	3.83 (1.69)	4.71 (2.32)
Recall	5.83 (2.62)	6.13 (3.84)	5.98 (3.25)
Combined	5.71 (2.20)	4.98 (2.30)	5.34 (2.26)
<i>Condition 3 (thematic cluster)</i>			
Recognition	4.83 (3.52)	3.83 (1.13)	4.33 (2.64)
Recall	7.42 (3.09)	6.17 (2.22)	6.79 (2.74)
Combined	6.16 (2.81)	5.00 (1.48)	5.56 (2.30)
<i>Condition 4 (unassociated set)</i>			
Recognition	6.17 (4.12)	4.54 (1.25)	5.35 (3.13)
Recall	7.67 (3.41)	6.13 (2.44)	6.90 (3.03)
Combined	6.92 (3.27)	5.33 (1.58)	6.13 (2.66)

strength of the comparison,  $F = 57.83$  with a chance of error of less than one in one thousand, but also its consistency as demonstrated by the fact that there were no significant interactions between the comparison of conditions 1 and 2 and any of the other four independent variables.

The data also present a clear indication that new L2 vocabulary items arranged in thematic clusters are more easily learnt than new L2 vocabulary items arranged in unassociated sets. Again, the strength of the comparison between the two alternate methods of organizing new L2 vocabulary is clearly indicated by the  $F$  ratio, 14.50, and the chance of error of less than one in a thousand.

The difference between conditions 3 and 4 also appears consistent as there were no interactions between the comparison and either modality, order or form. One interaction, however, comparison 2 by task, did appear. A possible explanation for this interaction is that since a particular set of words was assigned to a particular task, the difficulty or ease of learning that word set may be attributed to an inherent but unidentified difficulty in learning the word set itself rather than the task by which the set was learnt.

*c Individual performances:* Regarding the detrimental effect of semantic clustering, an examination of the individual performances of the 48 subjects participating in the research reveals a picture quite similar to that revealed by the statistical analyses of the aggregated data. These individual performances, however, do not paint quite as clear a picture of the benefits of thematic clustering as do the statistical analyses.

Given that there were 48 subjects with each subject providing comparative data twice, once on recall tests and once on recognition tests, a total of 96 individual comparisons may be drawn from the data. Such comparisons contrast, for a given subject, the number of trials needed to learn a semantic cluster to the number needed to learn an opposing unrelated set, as well as the number of trials needed to learn a thematic cluster to the number needed to learn an opposing unassociated set.

An analysis of these comparisons yields an observation that is, on the whole, congruent with observations resulting from the statistical analyses. Of the 96 comparisons comparing the trials needed by individual subjects in order to learn a semantic cluster to the number needed to learn an opposing unrelated set, the semantic clusters required more trials than the opposing unrelated sets on 65 of the 96 comparisons, fewer trials on 12 comparisons and the same number of trials on 19 comparisons. Clearly the comparisons favoured the learning of unrelated sets over the

learning of semantic clusters by a ratio of better than five to one.

The picture for thematic clustering, however, was not so clear. Of the 96 comparisons comparing the trials needed to learn a thematic cluster to the number needed to learn an unassociated set, the thematic clusters were learnt in fewer trials than the unassociated sets on 44 of the comparisons, in more trials on 22 of the comparisons and in the same number of trials on 30 comparisons. While thematic clustering appears to have been beneficial twice as often as it was detrimental, given the 30 comparisons in which it did not seem to make a difference one might argue that thematic clustering was a benefit to learning less than half the time and was actually a detriment over one fifth of the time.

## 2 Experiment 2 (separated)

*a Descriptive and inferential statistics:* To establish main effects and make the appropriate comparisons incorporating all the data provided by the experiment, the entire data set was combined to include both modalities (oral and written) and both tasks (recognition and recall). Using these data, main effects for 'condition', 'modality', 'task', 'order' and 'form' were established employing a 4 (condition: semantic cluster, unrelated set, thematic cluster, unassociated set)  $\times$  2 (modality: oral vs. written)  $\times$  2 (task: recall vs. recognition)  $\times$  4 (order: order 1, order 2, order 3, order 4)  $\times$  2 (form: form A vs. form B) mixed multiple analysis of variance (MANOVA) with 'condition' and 'task' within-subject variables and 'modality', 'form' and 'order' between-subject variables. The dependent variable was the number of trials required to reach the criterion of two consecutive perfect trials on a given test.

A significant main effect was found for condition,  $F(3,96) = 65.34$ ,  $p < 0.001$  and for modality,  $F(1,32) = 9.66$ ,  $p < 0.004$ . No significant effects were found for task, order or form.

No significant interactions were found between condition and any of the other independent variables.

To establish the comparisons of interest, data for the five factors were analysed using the MANOVA described above, the contrasts of interest being condition 1 vs. condition 2 (semantic cluster vs. unrelated set) and condition 3 vs. condition 4 (thematic cluster vs. unassociated set). A significant difference was found between conditions 1 and 2,  $F(1,40) = 89.35$ ,  $p < 0.001$ ; the semantic cluster requiring more trials to reach criterion than the unrelated set, as well as conditions 3 and 4,  $F(1,40) = 20.28$ ,  $p < 0.001$ ; the thematic cluster requiring fewer trials to reach criterion than the unassociated set.

A significant interaction was found between comparison 1 (semantic cluster vs. unrelated set) and modality,  $F(1,32) = 5.65$ ,  $p < 0.024$ , with the difficulty of learning the semantic clusters more pronounced in the oral modality. A significant interaction was also found between comparison 1 and task,  $F(1,32) = 5.45$ ,  $p < 0.026$ , with the comparison more pronounced on recall tests. Significant interactions were not found between comparison 1 and either order or form.

No significant interactions were found between comparison 2 (thematic cluster vs. unassociated set) and either modality, task, order or form.

Table 2 is a means table for Experiment 2. The table organizes means in a fashion that illuminates the three independent variables of interest: 'condition', 'modality' and 'task'. Since 'order' and 'form' are counterbalanced nuisance variables, they are not represented in the table.

*b Discussion:* As was the case in Experiment 1, the data presented by Experiment 2 present a clear indication that new L2 vocabulary items arranged in semantic clusters are learnt with more difficulty than new vocabulary items learnt in unrelated sets. The comparison between the alternate methods for grouping new L2 words is clear in light of the strength of the comparison,  $F = 89.35$

**Table 2** Means and standard deviations for scores by condition, task and modality for Experiment 2

	Oral	Written	Combined
<i>Condition 1 (semantic cluster)</i>			
Recognition	8.50 (4.30)	5.46 (2.60)	6.98 (3.84)
Recall	9.79 (3.51)	7.79 (3.28)	8.79 (3.51)
Combined	9.15 (3.09)	6.63 (2.34)	7.89 (3.00)
<i>Condition 2 (unrelated set)</i>			
Recognition	4.54 (1.53)	3.67 (1.01)	4.10 (1.36)
Recall	4.96 (1.65)	4.25 (1.26)	4.60 (1.50)
Combined	4.75 (1.29)	3.96 (.93)	4.35 (1.18)
<i>Condition 3 (thematic cluster)</i>			
Recognition	4.96 (1.73)	3.58 (1.06)	4.27 (1.58)
Recall	5.58 (1.77)	5.42 (1.72)	5.50 (1.73)
Combined	5.27 (1.57)	4.50 (1.09)	4.89 (1.39)
<i>Condition 4 (unassociated set)</i>			
Recognition	6.17 (1.79)	4.08 (1.21)	5.13 (1.84)
Recall	6.71 (2.46)	6.00 (2.27)	6.35 (2.37)
Combined	6.44 (1.82)	5.04 (1.51)	5.74 (1.80)

with a chance of error of less than one in one thousand.

The experiment, however, indicates that the comparison of semantic clusters and unrelated sets (comparison 1) presents inconsistencies which must be examined. First, a comparison 1 by modality interaction appears in the statistical analysis with the comparison more pronounced in the oral modality than in the written modality. A possible interpretation of this interaction is that a difference in the ease of the testing procedures for the oral and written tests may have contributed to the differences in the scores. More specifically, examination of the individual performances reveals that among the 96 written tests 11 were scored a perfect 'two' (trials-to-criterion) while among the oral tests only one was scored a perfect 'two'. The possibility, therefore, arises that a ceiling effect affected the written scores more than the oral scores, thus compressing the differences on the written scores more than on the oral scores.

Secondly, a comparison 1 by task interaction occurred with the comparison more pronounced in recall tests than in recognition tests. This difference may be given two interpretations. First, the fact that the difference between the two conditions was more pronounced on the recall tests (tests involving stimulus similarity) than on the recognition tests (tests involving response similarity) would be anticipated by and consistent with much of the early research into interference theory, research which led to the observation that stimulus similarity has more of a detrimental effect upon learning than does response similarity (see Crowder, 1976). Secondly, as was the case in Experiment 1, since a particular set of words was assigned to a particular task, the difficulty or ease of learning that word set may be attributed to an inherent but unidentified difficulty in learning the word set itself rather than the task by which the set was learnt.

The data also present a clear indication that new L2 vocabulary items arranged in thematic clusters are more easily learnt than new L2 words arranged in unassociated sets. Again, the strength,  $F(1,40) = 20.28$ ,  $p < 0.001$ , as well as the consistency (no interactions with task, modality, order or form) of the comparison between the two alternate methods of organizing new L2 vocabulary are clearly indicated.

*c Individual performances:* Observations of the individual performances provided in Experiment 2 were more supportive than those provided in Experiment 1. Of the 96 comparisons comparing the trials needed to learn a semantic cluster to the number needed to learn an opposing unrelated set, the semantic cluster was learnt

in more trials than the opposing unrelated set in 80 of the 96 comparisons, in fewer trials in 3 comparisons and in the same number of trials in 13 comparisons. Clearly, the comparisons favour the learning of unrelated sets over the learning of semantic clusters.

Again, however, the picture for thematic clustering was not so clear. Of the 96 comparisons contrasting the trials needed to learn a thematic cluster to the number needed to learn an unassociated set, the thematic cluster was learnt in fewer trials than the unassociated set in 47 of the comparisons, in more trials in 20 of the comparisons and in the same number of trials in 29 comparisons. While thematic clustering appears to have been beneficial far more often than detrimental, given the 29 comparisons in which clustering did not seem to make a difference it may be argued that thematic clustering was a benefit to learning only about half the time and actually a detriment about one-fifth of the time.

*d Qualitative analysis:* Immediately following completion of each session of tests, each subject was asked either orally or by the computer which of the six tests he or she had just taken seemed to be the easiest, why that test seemed to be easy, which test seemed to be the most difficult and why that test seemed to be difficult. In response to the first question, the vast majority of the subjects identified the thematic cluster as the easiest set to learn although some identified the unrelated set and a few identified the unassociated set. The most commonly perceived reason for the relative ease of the thematic sets was they somehow lent themselves to the establishment of associations connecting the English and artificial words. Some subjects, however, claimed that a given thematic cluster was easily learnt because the words 'fit together' or were somehow 'related'. One subject felt that the words in a particular thematic cluster were easy to learn because they could be fit together in a single sentence even though, given a bit of imagination, the words in the other sets could also fit into a single sentence. Other explanations were that the thematically associated words were 'shorter' (untrue), provided phonological or orthographical clues (unlikely) or were just 'cuter'. Many subjects responded that they just did not know why the thematically associated words were easier. The point, however, is that even though they did not often know what made the list easier, most subjects chose the thematic cluster as the easiest and none chose the semantic cluster.

The subjects seemed to have a clearer idea of which test was the most difficult. Virtually all those who responded (a few could not

remember) felt that the semantic cluster was more difficult than the others. Once again, many subjects stated that the semantic cluster was difficult because they had difficulty forming associations or connections between the English and artificial words. A sizeable number, however, felt that the semantic cluster was difficult because the words were 'too similar' or 'all related'. A few subjects claimed that the artificial words were difficult to remember because the English words were 'all the same'.

A few subjects, however, identified the thematic cluster as the most difficult although most of them also identified the semantic cluster as especially difficult. Again, those subjects either claimed they could not form associations between the English and artificial words or thought that the words were too closely related and thus difficult to learn.

#### **IV Conclusions**

The major findings of this research were that semantic clustering of new L2 vocabulary items appeared to serve as a detriment to the learning of vocabulary while thematic clustering appeared to serve as a facilitator of learning. In two separate but parallel experiments yielding similar results, sets of artificial words paired with semantically related English words (e.g., *apple, pear, nectarine, peach, apricot, plum*) were learnt with more difficulty than artificial words paired with sets of unrelated English words (e.g., *paint, funeral, recipe, market, uncle, ice*). In contrast, artificial words paired with the English words comprising thematically related sets (e.g., *frog, pond, green, slimy, hop, croak*) were learnt more easily than artificial words paired with English words comprising unassociated sets (e.g., *cloud, office, risky, social, lose, erase*).

The results of both experiments were presented most forcefully in the statistical analyses of the aggregated data provided by the subjects, analyses which indicated that semantic clustering was clearly detrimental to the learning of new L2 vocabulary across modalities (oral and written) and tasks (recall and recognition) while thematic clustering was clearly beneficial. Providing qualitative support for these findings, subject responses to a short questionnaire intended to explore subject impressions of the relative difficulty of L2 vocabulary learning under the four learning conditions revealed a near consensus that semantic clusters were the most difficult to learn of the four set types while thematic clusters were the easiest (although a few subjects also identified thematic clusters as very difficult).

Examination of individual performances yielded a strong

indication that semantic clusters were detrimental to learning but gave mixed, though generally positive evidence that thematic clusters were beneficial.

Thus, while there was strong and consistent evidence that semantic clusters are learnt with more difficulty than unrelated sets, the evidence that thematic clusters are learnt more easily than unassociated sets, while generally positive, was somewhat less strong and somewhat less consistent, a situation that clearly calls for further research. Also calling for further research is the limited generalizability of the current research: limited generalizability to an expanded stimulus base (more word sets within a particular condition); limited generalizability to evaluation criteria (long-term rather than short-term evaluation); and limited generalizability to other instructional contexts (context-based rather than rote-based learning). These limitations should serve as clear indications that the research does not conclusively demonstrate the effects of semantic and thematic clustering but, rather, calls into question certain current beliefs and instructional practices. Furthermore, the last point is especially important because I do not wish to imply by the format of this research that I support the learning of L2 vocabulary through the rote memorization of six-pair sets pairing L1 and L2 words. The research design was chosen because it seemed to be relatively unencumbered by extraneous variables. Whether or not the reported phenomena are noticeable in other more expanded and less controlled contexts is certainly a question calling for further research.

While the negative effect of semantic clustering upon L2 vocabulary would be anticipated by researchers concerned with interference theory (see Crowder, 1976), and the positive effect of thematic clustering would be anticipated by researchers concerned with the effects of schemata upon learning (see Brewer and Nakamura, 1984), the results of this research would probably come as a surprise to many current writers of ESL texts who, for a variety of reasons, rely heavily upon the employment of semantic clusters in their presentation of new vocabulary.

In conclusion, while the results of this research strongly suggest that semantic clustering impedes the learning of new L2 vocabulary items, the results also present a somewhat brighter possibility, the likelihood that thematic clustering serves as a benefit. Consequently, L2 curriculum writers and programme planners who currently present students with clusters of semantically and syntactically similar new words may need to reconsider such practice and explore, rather, the possibility that they might ease the burden of L2 vocabulary learning by incorporating

thematic rather than semantic clusters of new L2 words into their texts.

## V References

- Anderson, R.C. and Pearson, P.D. 1984: A schema-theoretic view of basic processes in reading comprehension. In Pearson, P.D., editor, *Handbook of reading research*, New York: Longman, 255–91.
- Brewer, W.F. and Nakamura, G.V. 1984: The nature and function of schemas. In Wyer, R.S. and Srull, T.K., editors, *Handbook of social cognition*. Vol. 1, Hillsdale, NJ: Lawrence Erlbaum Associates, 119–60.
- Carroll, J.B., Davies, P. and Richman, B. 1971: *Word frequency book*. New York: American Heritage.
- Crowder, R.G. 1976: *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fillmore, C.J. 1985: Semantic fields and semantic frames. *Quaderni di Semantica* 6, 222–54.
- Gairns, R. and Redman, S. 1986: *Working with words*. Cambridge: Cambridge University Press.
- Harmer, J. and Maybin, D. 1989: *Coast to coast 3*. White Plains, NY: Longman.
- Higa, M. 1963: Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior* 2, 170–75.
- Hunt, R.R. and Elliott, 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. and Mitchell, D.B. 1982: Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8, 81–87.
- Jolly, J. and Robinson, L. 1988: *Real-life English 3*. Austin, TX: Steck-Vaughn.
- Kittay, E.F. and Lehrer, A. 1992: Introduction. In Lehrer, A. and Kittay, E.F., editors, *Frames, fields, and contrasts*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1–18.
- Lavie, N.E., Briggs, S.J., Raht, C. and Denman, B.R. 1991: *In contact 2*. Glenview, IL: Scott Foresman.
- Markman, E.M. and Hutchinson, J.E. 1984: Children's sensitivity to constraints on word meaning: taxonomic versus thematic relations. *Cognitive Psychology* 16, 1–27.
- Marzano, R.J. and Marzano, J.S. 1988: *A cluster approach to elementary vocabulary instruction*. Newark, DE: International Reading Association.
- Molinsky, S.J. and Bliss, B. 1989: *Side by side 1*. Englewood Cliffs, NJ: Prentice-Hall.
- Richards, J.C. 1990: *Interchange 1*. Cambridge: Cambridge University Press.
- Rodgers, T.S. 1969: On measuring vocabulary difficulty: an analysis of item

variables in learning Russian-English vocabulary pairs. *IRAL (International Review of Applied Linguistics in Language Teaching)* 7, 327-43.

**Tinkham, T.M.** 1993: The effect of semantic clustering on the learning of second language vocabulary. *System* 21, 371-80.

**Walker, M.** 1991: *New horizons in English 1*. Reading, MA: Addison-Wesley.

**Wilkins, D.A.** 1976: *Notional syllabuses*. London: Oxford University Press.