Applied Psycholinguistics **24** (2003), 369–383
Printed in the United States of America
DOI: 10.1017.S0142716403000195

Semantic category effects in second language word learning

MATTHEW FINKBEINER and JANET NICOL
University of Arizona

ADDRESS FOR CORRESPONDENCE
Matthew Finkbeiner, Communication 302, PO Box 210025, University of Arizona, Tucson, AZ 85721-0025. E-mail: msf@u.arizona.edu

ABSTRACT
The present study addresses a long-standing assumption in the field of applied linguistics: that presenting new second language (L2) vocabulary in semantically grouped sets is an effective method of teaching. Participants learned 32 new L2 labels for concepts from four different semantic categories in either a related or unrelated condition. At test, participants translated words in both translation directions. We found a semantic interference effect both during the encoding of information into memory and during the retrieval of information in translation. We discuss these findings in terms of theoretical models of L2 lexical representation and development, as well as in more practical terms of L2 curriculum design and vocabulary instruction.

A long-standing question within the field of second language acquisition (SLA) has been how second language (L2) vocabulary should be taught: in semantic groups or not? Several researchers and textbook authors assume the efficacy of presenting learners with new L2 vocabulary in semantically related sets. For example, Gairns and Redman (1986), in their well-known book *Working with Words,* argue that grouping words by meaning "can provide greater precision in guiding students towards meaning, and in helping them to define the boundaries that separate lexical items" (p. 32). Similarly, Seal (1991) explains that when words are learned in semantic sets, "the learning of one item can reinforce the learning of another," as well as facilitate understanding because "items that are similar in meaning can be differentiated" (p. 300). These authors assume at least two benefits associated with learning words in semantic sets: the similarity between items (a) serves to facilitate the learning task and (b) causes the learner to notice fine-grained distinctions between words, which leads to a better understanding of the words that are learned.

A brief examination of several of the current English as a second language textbooks reveals that the practice of introducing new vocabulary in semantic sets is a popular one. For example, in one widely used textbook, *Vistas* (Brown, 1991), the "target vocabulary" for chapter 1 consists of items necessary in a classroom (e.g., paper, pen, pencil, chalk, blackboard, eraser). In another equally popular textbook, *ExpressWays* (Molinsky & Bliss, 1996), family members (e.g., husband, wife, mother, father, brother, sister) and places in the community (e.g., airport, bank, post office, mall, park, library, museum) serve as the target vocabulary in the first two lessons.

Although many SLA theorists and practitioners endorse (implicitly or explicitly) the seemingly sensible position that teaching new L2 vocabulary in semantically grouped sets is an effective method of teaching, there is actually very little empirical evidence to support this position. The body of literature often cited in support of presenting learners with semantically grouped words includes (monolingual) memory studies that involve the following: first there is a study phase, in which subjects are given a series of words (all of which are well known to the subjects) and told to memorize them; then there is a test phase, which requires subjects to either recall all the words from the study phase or to recognize which words from a list of words had appeared in the study phase. Such studies have found that semantically grouping the study words facilitates later recall or recognition (Bousfield, 1953; Cofer, 1966; Cohen, 1963). In a closely related set of studies, it has been shown that the use of category labels during the study phase and/or during later recall of word lists has improved performance (Lewis, 1971; Tulving & Pearlstone, 1966; Tulving & Psotka, 1971).

Some SLA researchers have interpreted these findings as providing support for the idea of using semantic sets in L2 vocabulary teaching. For instance, Schmitt (1997) argues that, just as grouping works to enhance recall of words in list-recall tasks in native speakers, "there is no reason to believe it does not do the same for L2 learners" (p. 213). It is not entirely clear, however, how a task that requires subjects to recall already well-known words that appeared on a study list is similar to the task of learning brand new L2 labels for already established concepts that have well-established first language (L1) labels. In a recall task, subjects could benefit from semantic grouping because it provides an organizational schema. If the subject notices that the study items all belong to a particular semantic category, at test the participant simply has to generate a list of appropriate exemplars for that category and check each one off against a record of items established during the study phase. If a generated item matches one of the subject's record of items, then she or he can "recall" it for the experimenter. This effectively turns the recall task into a recognition task, which is much easier to perform. Recall is much more difficult when participants are unable to make use of an organizational schema (Bower, Karlin, & Dueck, 1975).

In contrast, in an L2 vocabulary–learning task, it is unclear how awareness of the category label would be beneficial. The learning of new L2 labels for already established concepts requires the creation of new form–meaning connections. A strong, stable, one-to-one connection between a concept and its L1 form needs to give way to a one-to-two connection between concept and both the L1 form and L2 form (c.f. Kroll & Stewart, 1994; Potter, So, Von Eckardt, & Feldman, 1984). The precise nature of these connections has been the subject of much debate since the early work of Kolers (1963, 1966), but it is reasonable to suppose that the connection between concepts and L2 forms is quite weak in the initial stages of acquisition (Talamas, Kroll, & Dufour, 1999). The process of L2 vocabulary learning could thus be characterized as one of first establishing and then strengthening the new L2 form–meaning connections. This is clearly quite different from remembering a series of L1 words.

Thus, it is not surprising that studies that focus on the learning of new labels show a different effect of semantic grouping. In studies that record the number of learning trials needed to reach a predetermined learning criterion, it has been shown that participants take longer to learn new labels for sets of semantically related items than for sets of semantically dissimilar items (Higa, 1963; Kintsch & Kintsch, 1969; Nation, 2000; Tinkham, 1993, 1997; Underwood, Ekstrand, & Keppel, 1965; Waring, 1997). Note, however, that although learning semantically related words appears to take *longer,* it is possible that words learned under these conditions are learned *better* for the purpose of actual language use (e.g., the retrieval of vocabulary during production and comprehension). That is, the very difficulty associated with learning the new labels may make them easier to process once they are learned (c.f. Craik & Lockhart, 1972; Craik & Tulving, 1975, for a discussion of the effect of "depth of processing" on retrieval). Conversely, it is possible that whatever makes semantically grouped words more difficult to learn also makes them more difficult to retrieve during later language use.

There is, in fact, some evidence that L2 learners are slower to process words presented in semantic clusters. Kroll and Stewart (1994), for example, found that their Dutch–English bilingual subjects (for whom Dutch was the L1) were slower to translate Dutch words into English when the words were grouped by semantic category than when they were not. This result only held for L1–L2 translation, not for L2–L1 translation. Kroll and Stewart argued that this asymmetry is due to the existence of a direct link between L2 words and L1 words and that translation from L2 to L1 simply does not invoke conceptual representations. On Kroll and Stewart's view, no conceptual effects (such as semantic grouping) are expected for L2–L1 translation. This view has not been without challenge. Both Altarriba and Mathis (1997) and La Heij, Hooglander, Kerling, and van der Velden (1996) have reported semantic interference effects in L2–L1 translation tasks. In the Altarriba and Mathis (1997) study, subjects learned a small number of L2 words in association with L1 translation equivalent forms and were then asked to perform a translation verification task, in which pairs of L1 and L2 words were presented. The subjects were to decide whether the pairs were translation equivalents. Among the "no" cases were pairs that were semantically related but not equivalent. Reaction times to these semantically related foils were longer than for unrelated words, indicating the involvement of conceptual representations. This suggests that direct L2 form–meaning connections had been established even after one training session. Overall, then, there is evidence that even novice L2 learners may exhibit semantic interference effects (presumably the result of direct associations between newly learned L2 forms and meaning representations). Furthermore, it appears that presenting new labels by semantic category makes the labels harder to learn and, once learned, harder to retrieve.

In the present study, we tested these ideas directly by having a training phase in which participants learned new labels for familiar concepts and then a test phase in which they were required to retrieve those labels in a translation task. We manipulated semantic grouping both during training and during test. In addition, our test phase included both L1–L2 translation and L2–L1 translation, thereby providing an opportunity to extend the findings of Altarriba and Mathis (1997) to a language production task and to test whether Kroll and Stewart's (1994) asymmetrical pattern of translation results extends to novice language learners.

METHOD

Participants

Forty-seven undergraduates (29 female, 18 male) at the University of Arizona participated for course credit. All of the participants were monolingual English speakers.

Materials

Thirty-two novel words were created and each was paired with a picture of a familiar concept (see Appendix A for a complete list). The new L2 words were created to conform to English phonotactic constraints in order to reduce memory load (Ellis & Beaton, 1993; Gathercole & Baddeley, 1989, 1990; Service, 1992). For variety, half the words for each category were one syllable in length (e.g., *birk, plap, floop*) and the other half were two syllables in length (e.g., *walloon, dopal, fonteen*). The pictures used were adopted from Snodgrass and Vanderwart (1980). The eight most prototypical category exemplars available from the Snodgrass and Vanderwart (1980) set of drawings were selected from each of four categories, animals, kitchen utensils, furniture, and body parts.

Procedure

Individuals participated on 2 separate days within a 5-day period. Each session spanned approximately 45 minutes and consisted of vocabulary training (participants were told that they were learning a new "alien" language), followed by a recognition task, and then two blocks of translation in each direction (L1–L2 and L2–L1), for a total of four blocks of translation. The recognition task and translation tasks at the end of the first session were included because pilot experiments revealed that participants benefited from being "tested" partway through training. Pilot testing also indicated that performance on these first session tests was generally poor, however, so only translation times from the second session were recorded for analysis.[1] Participants were seated individually in sound-resistant computer booths during the entirety of each session. At the beginning of the first session, participants were shown the pictures used in the experiment on flashcards and asked to name them in English. After they had named each of the pictures correctly, they were asked to write the names of the items down before beginning their training on the computer. This was done to ensure name agreement on all of the items because, although L1 labels were never present during training, the translation task did use them. During training, participants first heard a recording of the L2 word over headphones, then saw the L2 word and its corresponding picture for 500 ms on the monitor, and then heard a second recording of the L2 word. Participants were asked to repeat the L2 word twice into a microphone placed in front of them for recording purposes. The purpose of the repetitions was simply to facilitate learning. A post hoc assessment of the recordings revealed that all of our participants repeated the L2 words at least twice during training.

Items were displayed on the computer using the DMDX system developed by J. C. Forster at the University of Arizona. In the "related" training condition, semantically related items were blocked into groups of eight. Each block of eight was presented four times during training, in pseudorandom order such that no block appeared twice in a row. In the "unrelated" training condition, the 32 items were scrambled within a block and each block was presented four times (with the order of items in different random order each time). The vocabulary training was followed by a recognition task, which consisted of the presentation of a picture followed by one of the L2 labels. Of the 64 picture–label pairs, half were correct (the picture was paired with its new label) and half were incorrect (the picture was paired with the wrong label). Participants were instructed to press a "yes" button if the picture and L2 word matched and a "no" button if they did not. Participants were given feedback for each item, including whether they were correct, as well as their reaction times.[2] After the recognition task, participants were given the translation task. In this task, participants were shown a row of hash marks (#######) for 1 s (for orienting purposes only), followed by the word to be translated. For example, in the L1–L2 blocks, an English word appeared and participants were asked to speak the "L2" translation equivalent into the microphone as quickly as possible. Their vocal response triggered a voice key, stopping the computer's timer. Latencies were measured from the time the word to be translated was presented until the voice key was triggered. All responses were recorded on tape so that they could be checked for errors. As per standard procedure in timed language production studies (Damian, Vigliocco, & Levelt, 2001; Levelt, Roelofs, & Meyer, 1999; Schriefers, Meyer, & Levelt, 1990), we counted as errors all incorrect responses, as well as fluency errors like stutterings and "um," because they would lead to an inaccurate measure of timing (triggering the voice key before the onset of the word was uttered). In order to control for any differences that may arise due to order of translation direction, translation direction was counterbalanced such that half the participants in each design cell performed forward translation (L1–L2) first and the other half performed backward translation (L2–L1) first.

Table 1. *Translation means (standard deviations) by training and testing condition*

	Testing (ms)		
Testing	Related	Unrelated	Mean (ms)
Related	1272 (181)	1148 (113)	1210
Unrelated	1323 (213)	1062 (155)	1193
Mean	1298	1105	

Table 2. *Forward translation (L1 > L2) means (standard deviations) by training and testing condition*

	Training (ms)		
Testing	Related	Unrelated	Mean (ms)
Related	1380 (180)	1196 (112)	1288
Unrelated	1415 (215)	1121 (164)	1268
Mean	1397	1159	

RESULTS

As is typical in learning experiments of this type (in which there is no concrete incentive for participants to perform well), many participants failed to reach the preset learning criterion (see, e.g., Altarriba & Mathis, 1997). Data from 23 participants were excluded from analysis for not reaching the predetermined accuracy criterion of 80%.[3] Mean reaction times (RTs) for correct responses per experimental condition and participant were calculated for the remaining 24 participants (14 female, 10 male). The mean translation times for the four treatment conditions appear in Table 1. The average error rate for the experimental conditions was 9% and did not differ across conditions.

Separate analyses of variance with participants (F_1) and items (F_2) as random variables reveal significant main effects of translation direction, F_1 (1, 20 = 41.81, $p < .0001$; F_2 (1, 124) = 1687, $p < .0001$, and training conditions, F_1 (1, 20) = 11.55, $p = .003$; F_2 (1, 124) = 126.56, $p < .0001$, but no effect of testing condition, F_1 (1, 20) < 1; F_2 (1, 124) < 1. The interaction between translation direction and training condition was significant in analyses by items, F_2 (1, 124) = 8.01, $p = .005$, and approached significance by participants, F_1 (1, 20) = 3.00, $p = 0.09$. There were no other significant interactions.

Now let us consider the results for each translation direction separately. The results for L1–L2 translation can be seen in Table 2, and the results for L2–L1 translation can be seen in Table 3. Translation in the forward direction was 238 ms longer for those participants who learned semantically grouped L2 words compared to those who learned a random grouping of the same L2 words. Planned comparisons showed that this difference was highly significant, F_1 (1,

Table 3. *Backward translation (L2 > L1) means (standard deviations) by training and testing condition*

	Training (ms)		
Testing	Related	Unrelated	Mean (ms)
Related	1164 (127)	1100 (98)	1132
Unrelated	1230 (182)	1002 (125)	1116
Mean	1197	1051	

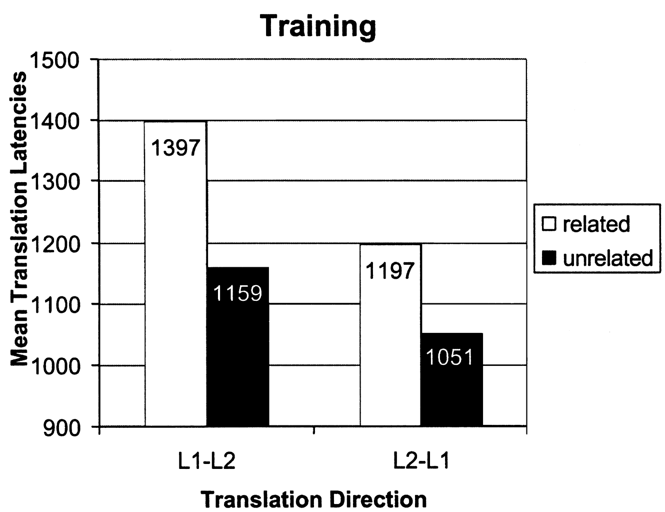

Figure 1. Response latencies by training condition (related vs. unrelated) and translation direction.

20) = 12.60, $p = 0.002$; F_2 (1, 63) = 71.60, $p < .0001$. Translation in the backward direction was 146 ms slower for participants who learned semantically related words compared to participants who learned the same words in unrelated sets. Planned comparisons showed that this difference was significant in both the participants analysis, F_1 (1, 20) = 7.46, $p = .013$, and the items analysis, F_2 (1, 63) = 57.78, $p < .0001$. It is quite clear from these results that there was a semantic category effect in training and that this effect was negative. Participants translated L2 labels learned in semantic sets significantly more slowly than they did L2 labels learned in random order. This was the case in both translation directions.

Mean translation latencies as a function of training condition (related vs. unrelated items) and translation direction are shown in Figure 1. As can be seen, when translation latencies are grouped according to training condition, there is a clear difference between participants who learned L2 words in semantic groups and those who learned the same L2 words in random order.

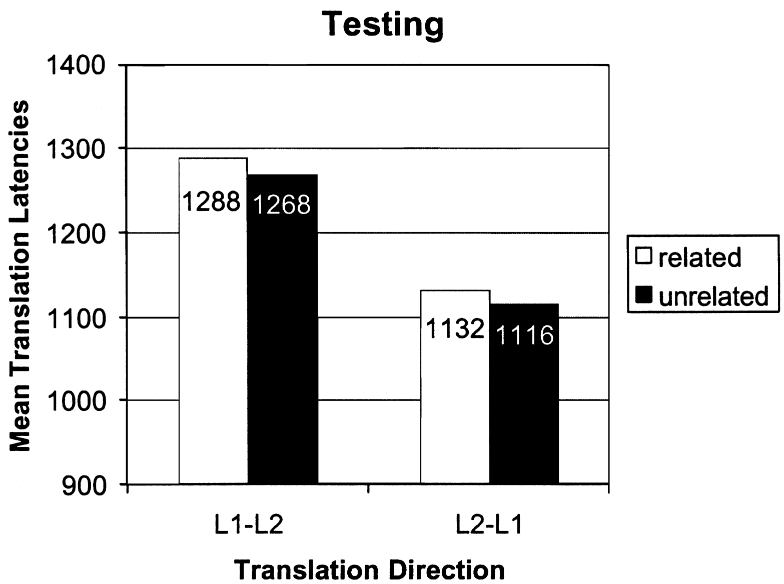

Figure 2. Response latencies by testing condition (related vs. unrelated) and translation direction.

Figure 2 shows mean translation latencies as a function of testing condition (related vs. unrelated items) and translation direction, revealing that participants who translated items that were grouped into semantic sets were not reliably slower than participants who translated random sets of those same items.

These results also suggest that (assuming that our four test groups are otherwise equal) similar conditions at training and test had a small facilitatory effect: note that the fastest translation times appeared in the condition in which unrelated words were presented at both training and at test, as expected, but the slowest condition was not the condition in which related words were presented at training and at test; rather, items that were semantically grouped during training but not at test elicited the slowest translation times.

DISCUSSION

The main purpose of this study was to explore whether grouping words into semantic sets during training affected learners' performance on a translation task once the words were learned. We found that translation times were significantly slower for words learned in semantic sets versus in random order. Further, we found that translation performance was adversely affected by presenting the words to be translated in semantic categories, though this effect was not significant. Overall, then, presenting semantically grouped L2 words to learners has a deleterious effect on learning.

Why should this be? We suggest that simultaneous activation of semantically related lexical items is at the root of the effect. Several different theories of lexical representation (e.g., de Groot, 1992; Levelt et al., 1999; Schreuder & Flores D'Arcais, 1989) assume that when a concept is activated, there occurs spreading activation among representations that are similar in meaning. This spreading activation is responsible for the facilitatory effects in categorization tasks (Glaser & Düengelhoff, 1984) and semantic priming tasks (Meyer & Schvaneveldt, 1971; Neely, 1991), as well as inhibitory effects in production tasks (Levelt et al., 1999). Let us consider some of this research in more detail.

In a typical semantic priming experiment (cf. Neely, 1991, for a review), participants are presented with two words in succession. The first word is referred to as the "prime," and typically participants do not respond to it; the second is the "target" item to which participants respond, typically making a decision about lexical status (word vs. nonword) or by saying the word out loud. Participants usually respond faster to a target word (e.g., *boy*) following a semantically related prime (e.g., *girl*) than following an unrelated prime (e.g., *desk*). Because there is no overlap in orthographic form between the prime and target and because priming effects do not depend on established co-occurrence patterns of the prime and target (Thompson–Schill, Kurtz, & Gabrieli, 1998), the locus of the facilitation may be best characterized as due to the "preactivation" of semantic features of the target due to the activation of the shared semantic features of the prime.

In production, spreading activation can create interference. One task that has been used extensively to explore aspects of the language production system is the *picture–word interference task*. In this task, participants are asked to name a picture that appears more or less coincident with a written or spoken "distractor" word. A semantic relationship between a picture and a distractor word slows picture naming relative to an unrelated condition. Interference in this task is argued to arise as a result of coactivated lexical entries that are semantically related. The picture of an object activates a cohort of related, mutually activating lexical concepts, which in turn activate their corresponding lexical representations, or *lemmas* (the term lemma refers to a sort of lexical hub, a representation that connects the three major aspects of a lexical item—form, meaning and syntax). Coactivated lemmas compete with each other (Damian et al., 2001) and affect the speed with which a lexical form is retrieved and output. The strength of activation of competitors is increased by the presentation of a semantically related distractor word. Thus, selection of a particular target lemma is delayed by the simultaneous activation of related lemmas (Levelt et al., 1999).

How do such results bear on our word learning task? Let us assume that learning a new label requires representing the form information (either orthographic, phonological, or both) and linking this representation to a concept or L1 lexical representation (or both). In the present study, participants saw a picture paired with L2 orthographic and phonological forms. We assume that viewing the picture served to activate its corresponding concept, which, in turn, activated its L1 lemma as well as related concepts. In both training conditions, the L2 form is linked to a concept and corresponding L1 lemma. In the unrelated

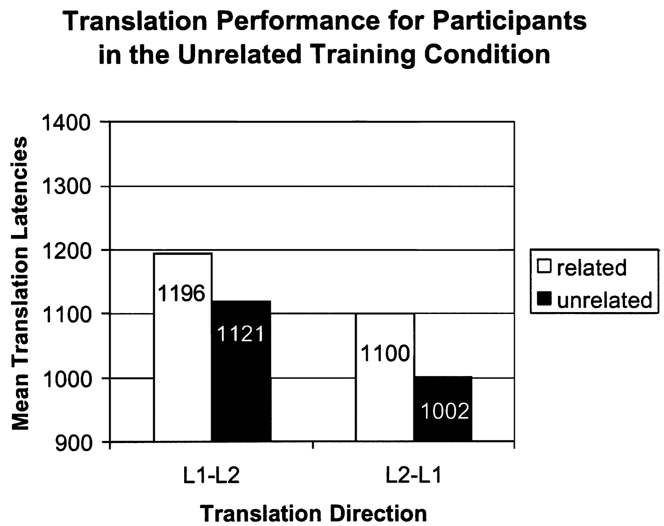

Figure 3. Response latencies by testing condition (related vs. unrelated items) and translation direction for participants from the unrelated training condition only.

training condition, interference from related lemmas is present but minimal. In the related condition, however, interference from related lemmas is heightened due to repetitive and residual activation of concepts and lemmas within the same category. This results in a much weaker connection between the new L2 lexical entry and the corresponding concept and L1 lemma.

This account is compatible with the results of a complementary study conducted in our lab. Finkbeiner and Nicol (2000) reasoned that if activation of L1 lemmas could be "heightened" without creating interference from related lemmas, then facilitation, instead of inhibition, should be observed in subsequent performance tasks. They found that if they enhanced the perceptual elements (cf. Schreuder & Flores D'Arcais, 1989) of target pictures during learning, while controlling for semantic interference from related items (by not presenting related items together), participants performed faster in translation and picture naming tasks relative to participants who learned novel words for perceptually unremarkable pictures. Taken together, these studies suggest that "heightening the level of activation" of target concepts during learning can lead to facilitation, but only if interference from competing items is minimized.

The present study also showed that translation direction did not affect the results: the semantic interference effect was present for both conditions in which participants learned words presented in random order (Figure 3). This result is compatible with the findings of Altarriba and Mathis (1997), who used translation verification. Both studies suggest that both L1–L2 translation and L2–L1 translation involve semantic representations: there is no indication that L2–L1 translation bypasses the conceptual store.

These results are problematic for the revised hierarchical model (RHM). Recall that the RHM proposes that, while forward translation is conceptually mediated, backward translation is lexically mediated (i.e., L2 words are directly linked to L1 words). The evidence for this model comes largely from the study of proficient language learners (bilinguals), but the asymmetry has been claimed to be strongest in novice learners: L2 words are initially linked directly to L1 words; the connections between L2 words and their corresponding concepts develops over time, becoming stronger with increasing proficiency (Kroll & Curley, 1988; Kroll, Michael, & Sankaranarayanan, 1998). Why then, do our novice learners (along with those studied by Altaribba & Mathis, 1997) show no asymmetry? One reason for our symmetric pattern might be that our training paradigm involved associating new labels with pictures rather than with L1 words. It may well be that the strong lexical link between L2 words and L1 words arises when L2 vocabulary is taught via translation (but see Altaribba & Mathis, 1997). Much more research along these lines is clearly necessary.

The implications that this study has for vocabulary instruction and curriculum development are not trivial. As pointed out in the introductory section, several authors in the teaching methodology literature have argued that vocabulary should be taught in semantic groups. The results of the present study converge with those of Tinkham (1993, 1997) and Waring (1997) to suggest that teaching words in semantic sets creates competition between items, which in turn increases difficulty during learning and during memory retrieval in language production.

APPENDIX A

The Stimuli used in the experiment are presented along with the novel "L2" forms and their rank by frequency of the concept in each category (according to Battig & Matague, 1969).

Items	Ranking	L2 Word
	Animals	
Cat	2	Birk
Cow	4	Gorp
Dog	1	Floop
Elephant	7	Glip
Horse	3	Larpell
Lion	5	Treffim
Pig	8	Ploozette
Tiger	6	Walloon
	Kitchen Utensils	
Bowl	9	Fonteen
Cup	11	Blikeet
Fork	3	Dopal
Frying pan	15	Tilkoon

APPENDIX A (*cont.*)

Items	Ranking	L2 Word
Knife	1	Dax
Pot	5	Plap
Spoon	2	Fremm
Stove	8	Cald
	Furniture	
Bed	3	Carm
Chair	1	Lig
Couch	7	Vid
Desk	5	Zek
Dresser	8	Fozzeel
Lamp	6	Soudleen
Table	2	Beleem
Television	9	Chabouk
	Body Part	
Ear	8	Detalle
Eye	4	Ecrus
Foot	5	Fossay
Hair	13	Jacoll
Hand	9	Milg
Leg	1	Pelf
Nose	6	Roop
Toe	10	Yume

ACKNOWLEDGMENTS

The authors gratefully acknowledge support for this research from a number of sources, including the Faculty Small Grants Program (Office of the Vice President for Research and Graduate Studies, in conjunction with the University of Arizona Foundation), grant NIDCD DC-01409 (as part of the National Center for Neurogenic Communication Disorders, University of Arizona), and the Cognitive Science Program, University of Arizona.

Both authors contributed equally to this study.

NOTES

1. Because of this design, learning performance was not assessed until the language production test on Day 2. For this reason, all participants were tested, regardless of whether they had sufficiently learned the new labels.
2. Note that the recognition task was not designed to elicit analyzable data in the sense that the method of display (the picture appeared for 500 ms, followed by one of the L2 words) could encourage some subjects to guess what label would follow (which would facilitate responses half the time and inhibit responses the other half of the time).

3. The accuracy criterion of 80% is the standard employed in our lab; its purpose is to exclude participants who are not engaged in the experiment. We should point out, however, that the overall pattern of reaction times when all participants are included is similar to that when just those who met the accuracy criterion are included.

REFERENCES

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

Battig, W. F., & Montague, W. E. Category norms of verbal items in 56 categories. A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80*(3, pt. 2), 1–46.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49,* 229–240.

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory and Cognition, 3,* 216–220.

Brown, H. D. (1991). *Vistas.* Englewood Cliffs, NJ: Prentice-Hall.

Cofer, C. N. (1966). Some evidence for coding processes derived from clustering in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 188–192.

Cohen, B. H. (1963). Recall of categorized word lists. *Journal of Experimental Psychology, 66,* 227–234.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychology Review, 82,* 407–428.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671–684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition, 81,* 77–86.

de Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 389–412). Amsterdam: North–Holland.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Journal of Experimental Psychology: Human Experimental Psychology, 46,* 533–558.

Finkbeiner, M., & Nicol, J. (2000). *Early L2 lexical representation: An episodic memory explanation*. Paper presented at the Second Language Research Forum, Madison, WI.

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

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28,* 200–213.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81,* 439–454.

Glaser, W. R., & Düengelhoff, F. (1984). The time course of picture–word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 640–654.

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

Kintsch, W., & Kintsch, E. (1969). Interlingual interference and memory processes. *Journal of Verbal Learning and Verbal Behavior, 8,* 16–19.

Kolers, P. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior, 2,* 291–300.

Kolers, P. (1966). Interlingual facilitation of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 5,* 314–319.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second languages words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (Vol. 2). London: Wiley.

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

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. Borne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 365–395). Mahwah, NJ: Erlbaum.

La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language, 35,* 68–665.

Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–75.

Lewis, M. Q. (1971). Categorized lists and cued recall. *Journal of Experimental Psychology, 87,* 129–131.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Molinsky, S. J., & Bliss, B. (1996). *ExpressWays*. Englewood Cliffs, NJ: Prentice–Hall.

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal, 9,* 6–10.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23–38.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199–227). Cambridge: Cambridge University Press.

Schreuder, R., & Flores D'Arcais, G. B. (1989). In W. Marslen–Wilson (Ed.), *Lexical representation and process* (pp. 409–436). Cambridge, MA: MIT Press.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture–word interference studies. *Journal of Memory and Language, 29,* 86–102.

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.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 45,* 21–50.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Experimental Psychology, 6,* 174–215.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2,* 45–58.

Thompson–Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. E. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory & Language, 38,* 440–458.

Tinkham, T. (1993). The effects of semantic clustering on the learning of second language vocabulary. *System, 21,* 371–380.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13,* 138–163.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381–391.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology, 87,* 1–8.

Underwood, B. J., Ekstrand, B. R., & Keppel, G. (1965). An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. *Journal of Verbal Learning and Verbal Behavior, 4,* 447–462.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 25,* 261–274.