JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **15**, 347–364 (1976)

Bilingual Episodic Memory: Acquisition and Forgetting

COLIN M. MACLEOD

University of Washington

Two experiments investigated the representation of meaning and input language in bilingual memory. Experiment I used the savings method to investigate the kind of information remaining in a nonrecallable memory trace 5 weeks after original learning. The results showed significant savings for same versus changed meaning (across translation equivalents such as HORSE and CHEVAL) but not for same versus changed language. This result, in conjunction with previous research on savings, suggests that translation equivalents do not function as synonyms but, instead, are mediated by an underlying supralinguistic concept. Experiment II used the depth-of-processing incidental-learning paradigm to investigate the kind of information acquired when words are classified in terms of either meaning (LIVING vs. NONLIVING) or language (FRENCH vs. ENGLISH). The results showed that meaning classifications produce better memory than linguistic classifications, both in terms of recall for meaning and in terms of recognition for the language of presentation; this suggests that memory processing is not directed to a single level but, instead, passes through the linguistic level on the way to the deeper semantic level.

The representation problem is currently a focal issue in theorizing about human memory. Whether at the level of specific events or general knowledge, the concern is with how knowledge is represented in memory. To assist in classification, Tulving (1972) has created a useful dichotomy between two types of memory he calls episodic memory and semantic memory; within this framework, he distinguishes the specific knowledge from the general knowledge aspects of memory. Thus, *episodic* memory contains the records of unique events (episodes) which occurred at particular times (e.g., that I first saw a penguin at the Granby Zoo near Montréal when I was 6 years old). In contrast, *semantic* memory does not contain temporally coded information; rather, it is a kind of mental lexicon containing all of the attributes of event-free knowledge (e.g., that a penguin cannot fly, although it is a bird and almost all birds can fly). The classification is helpful, but a central question remains: How is information represented in each of these types of memory?

The present paper investigates the representation problem in episodic memory using bilingual subjects. Bilinguals are an interesting special case because they have *two* language systems in which to represent their knowledge. Also, in the context of a list-learning type of episodic memory experiment, retention of input language is generally required in addition to retention of meaning. Thus, it is possible to examine the representation of two components of the same information in bilingual memory. These are the aims of the two experiments

This research was partially supported by U.S. Public Health Service Grant MH-21037 to Thomas O. Nelson. For their assistance in locating subjects, I thank Chris Adjemian (Experiment I) and Rona Kornblum (Experiment II). I am grateful to Edwin R. Anderson, Robert C. Bolles, Earl Hunt and Geoffrey R. Loftus for their valuable advice and criticism. Most of all, I wish to thank my committee chairman, Thomas O. Nelson. This paper is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Washington. Requests for reprints may be sent to Colin M. MacLeod, Department of Psychology, NI-25, University of Washington, Seattle, WA 98195.

reported here—to examine acquisition and forgetting in bilingual episodic memory and to relate these findings to notions of how information is represented in episodic memory.

Because the two experiments approach the study of bilingual episodic memory from quite different directions, they will be discussed individually first, followed by a more general discussion of representation in episodic memory. However, the rationale underlying the two experiments will be outlined before introducing Experiment I.

Experiment I examines forgetting of mixed-language information from episodic memory. One of the most glaring gaps in our knowledge of bilingual episodic memory concerns long-term retention—no laboratory study has examined the retention of information in two languages beyond a single session. For this reason, Experiment I uses a 5-week retention interval, in part to demonstrate the feasibility of research on bilingual long-term retention. This experiment uses the savings method (cf. Nelson, 1971b), perhaps the most sensitive measure of long-term retention (cf. Nelson, 1971a), to examine the possibility that translation equivalents are represented in episodic memory by the same supralinguistic semantic concept (i.e., they share an abstraction at a level more primitive than words). Opposed to this "concept" notion is the notion that translation equivalents are synonyms of each other (i.e., they have different supralinguistic abstractions). The focus of the experiment is on the information remaining in episodic memory for those instances where items are nonrecalled (i.e., forgotten) after a 5-week retention interval.

Experiment II examines acquisition of mixed-language information into episodic memory. Of major concern in Experiment II is the fact that language retention seems to have been remarkably good in previous research on bilingual memory. Because this might have been due to the demand characteristics of the intentional learning situations in prior studies, Experiment II uses an incidental paradigm. Also of interest in Experiment II is the question of whether processing time and depth of processing can be distinguished experimentally. Pertaining to depth of processing, data are obtained relating to the manner in which different levels of processing are accessed during acquisition. The focus of the experiment is on an immediate retention test for items processed with respect to their input language versus with respect to their meaning.

EXPERIMENT I

The way in which translation equivalents are represented in episodic memory remains a problem for theories of bilingual memory. Dalrymple-Alford and Aamiry (1970) have argued that the relationship in memory between two words such as HORSE and CHEVAL is synonymic, implying nonidentity at the supralinguistic level. Kolers (1966a, b) and others (e.g., Liepmann & Saegert, 1974) have countered that the two words represent the same abstract underlying concept, implying identity at the supralinguistic level. To understand representation in bilingual memory, a critical experiment is required posing these two views against each other.

Before introducing the methodology of Experiment I, the notion of supralinguistic concepts must be elaborated. By "supralinguistic concept" is meant an abstraction of meaning at a level more primitive than the word itself. Philosophers have often discussed similar notions of underlying representation (e.g., Langer, 1953, has introduced the idea of a "conception" which is not unlike the idea presented here). Recent computer simulations of sentence memory by psychologists (e.g., Anderson & Bower, 1973; Kintsch, 1974; Rumelhart, Lindsay, & Norman, 1972) have begun using similar constructs in theories of memory representation (cf. "types" in Anderson & Bower, 1973).

Unfortunately, psychologists have avoided defining synonymy (other than by use of specific examples), perhaps because the term is quite elusive. Philosophers, on the other hand, have been very concerned with synonymy in attempting to avoid circularity in definition. Generally, philosophers emphasize the context (Alston, 1964) or occasion (Quine, 1960) of usage of synonymous terms. Thus, Alston (1964, p. 45) argues that "even if we restrict ourselves to those contexts within which a pair of terms seem to have exactly the same meaning, . . . there are various differences that attach to the use of terms". Furthermore, the argument is frequently made regarding the meaning of synonyms that two words are nominally synonymous "without having the same meaning in any acceptably defined sense of 'meaning'" (Quine, 1960, p. 46). In terms of supralinguistic concepts, then, a pair of synonyms would not have the identical representation. Stern (1964, p. 226) is particularly clear on this issue: "Synonyms may be defined as words with identical or partly identical referential range, but different semantic ranges. That is to say, they denote the same referents, but each word denotes it in an aspect that somehow differs from the others".

With respect to translation equivalents, then, two possible representations can be suggested. The first is that two words such as HORSE and CHEVAL are synonymous and thus do not have the identical supralinguistic concept. The second is that the two translation equivalents are identical at a more primitive level and therefore do have the same underlying representation. Experiment I addresses the representation problem by posing these alternatives against each other.

In a series of studies using the savings paradigm, Nelson and his co-workers (Nelson, 1971a, b; Nelson & Rothbart, 1972; Nelson & Fehling, Note 1) have examined the residual information in the memory traces of items which cannot be recalled (Nelson, 1971b) or even recognized (Nelson, 1971a) on retention tests 4–7 weeks after original learning. Generally, however, Nelson has used the operational definition of failure to recall as his criterion for isolating forgotten items; this definition is also adopted in the present experiment. The subject in a savings study originally learns a list of 16–20 number–word paired associates (e.g., 56–CAR) in the first session of the experiment. In the second session several weeks later, the subject is tested for retention of the words by presenting only the numbers and allowing the subject as long as he wants to search his memory for the correct words. For those words that are forgotten (i.e., not recalled), new words are substituted and the resulting new list is studied and tested on one relearning trial. Savings is evidenced to the extent that the residual information in the forgotten word facilitates relearning of a related new word relative to an unrelated new word.

The savings studies have revealed a clear pattern concerning the information which remains in a nominally forgotten item. Figure 1 exemplifies this pattern for the target 56–CAR. Given that CAR cannot be recalled in the presence of 56, the most easily relearned word is CAR (Nelson, 1971b, Experiment 1). This demonstrates that there is *some* residual information in the memory trace; subsequent studies have focused on the nature of this residual information. The Nelson and Fehling (Note 1) results show that relearning is also facilitated for *subordinates* and *superordinates*

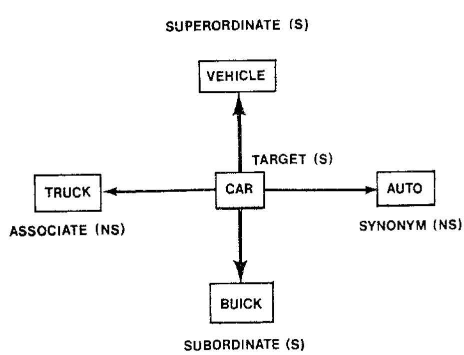

FIG. 1. The pattern of savings for substituted words in the case of the forgotten target, 56–CAR. S indicates savings; NS indicates no savings.

(e.g., BUICK and VEHICLE), when compared to unrelated words. However, in repeated attempts, no reliable savings has been obtained for *antonyms* (Nelson, 1971b, Experiment 2; Nelson & Fehling, Note 1, Experiment 1), for *synonyms* (Nelson, 1971b, Experiment 2; Nelson & Fehling, Note 1, Experiments 1 and 3), or for *associates* (Nelson & Fehling, Note 1, Experiment 4) of the target item.

Because synonymic and associationistic savings are absent in long-term retention, the savings paradigm is appropriate for examining the nature of the representation of translation equivalents in episodic memory. If there is no savings for translation equivalents, then the idea of conceptual identity is disconfirmed, while the synonym notion remains quite plausible. On the other hand, if there is savings for translation equivalents, then the synonym view is disconfirmed and the conceptual identity view is favored. To test these predictions, Experiment I examines long-term retention of bilingual lists using the savings paradigm.

Method

Subjects. Of the 24 subjects in original learning, 23 returned for retention tests and relearning. All were graduate and undergraduate students in French at the University of Washington. Six of the subjects had spoken both languages from early childhood; of the remaining 17 subjects, all had studied French, their second language, for a minimum of 2 years at the university level. Subjects rated their knowledge of each language on the scale shown in Figure 2. All subjects rated themselves "fluent" in English and at least "competent" in French. This self-evaluation procedure

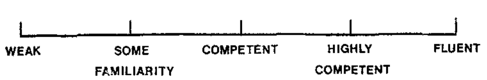

FIG. 2. The scale used for rating fluency in each language.

was chosen to estimate degree of language balance because retrospective reports of language ability have been found to correlate quite highly with proficiency tests (Fishman & Cooper, 1969).

Subjects participated individually in both sessions of the experiment and were paid $2.50 after the second session was completed.

Stimuli. Each subject learned a 20-item paired-associate list. The stimuli were two-digit numbers of low association value (1.36–1.99) in the Battig and Spera (1962) norms; these were the same numbers as those used by Nelson in his studies. The responses were 40 English nouns (and their translations) selected from the A and AA nouns in the Paivio, Yuille, and Madigan (1968) norms. These nouns were selected such that their most preferred French translation equivalents were both orthographically and acoustically dissimilar to the English words (e.g., FLAG–DRAPEAU). Because presentation was visual, this precaution helped to ensure that subjects encoded each word in the correct language.

Each number was randomly paired with two English nouns and their French translations with the restriction that no obvious semantic similarity existed between the two translated word pairs associated with each number. Table 1 presents the 20 × 4 matrix of nouns and indicates the four possible relationships between a relearned item and its corresponding (same number) originally learned item: Original language–original meaning (OL–OM), different language–original meaning (DL–OM), original language–different meaning (OL–DM), and different language–different meaning (DL–DM). For a sample subject in Table 1, the underlined items were originally learned and the all-English list was relearned. Thus, five items in the originally learned list fell into each of the four relationships in terms of the relearned list. Of course, from the subject's point of view there were simply 10 words in each of his languages during original learning.

Because two relearning lists were used, an

TABLE 1

EXPERIMENT 1: ORIGINALLY LEARNED AND RELEARNED WORDS FOR A SAMPLE SUBJECT

	Original learning				Relearning English
	OL–OM	DL–OM	OL–DM	DL–DM	
17	health	santé	nail	*clou*	*health*
26	bird	*oiseau*	vegetable	légume	*bird*
28	dog	chien	*father*	père	*dog*
30	*queen*	reine	glass	verre	*queen*
32	world	monde	butter	*beurre*	*world*
35	gift	*cadeau*	month	mois	*gift*
42	window	fenêtre	*city*	ville	*window*
44	*tree*	arbre	door	porte	*tree*
48	flag	drapeau	shoe	*soulier*	*flag*
55	meat	*viande*	ticket	billet	*meat*
56	boy	garçon	*brain*	cerveau	*boy*
70	*house*	maison	winter	hiver	*house*
72	square	carré	horse	*cheval*	*square*
76	clothes	*vêtements*	airplane	avion	*clothes*
77	church	église	*ink*	encre	*church*
80	*night*	nuit	earth	terre	*night*
81	apple	pomme	death	*mort*	*apple*
94	year	*année*	knife	couteau	*year*
95	elbow	coude	*star*	étoile	*elbow*
97	*shadow*	ombre	clock	horloge	*shadow*

English list for half of the subjects and a translated French list for the other half of the subjects, two subjects received each original-learning randomization of the items. In this way, every original randomization of the list was used once with each of the relearned lists, and the two languages were used equally often in relearning.

Procedure: Original learning. Each number–noun paired-associate was typed on a 3 × 5-in. index card, as were the four vowel–consonant practice paired associates. Using the study–test method of paired-associate learning, the pairs were presented for study at an 8-sec rate, paced by an audible click from a metronome. During original learning, pairs were also tested at an 8-sec rate. All retention tests in the second session were self-paced.

Before learning his number–noun list, each subject had one trial on the practice list of four vowel–consonant pairs to familiarize him with the procedure and timing. The subject studied each pair silently for 8 sec. Following the final study pair, the subject saw a three-digit number from which he counted backward by threes as rapidly as possible for 32 sec. This task was interpolated between study and test to ensure that items recalled on the subsequent test were stored for long-term retention during study. By preventing the subject from simply rote repeating the last few items studied (cf. Glanzer & Cunitz, 1966; Postman & Phillips, 1965), the interpolated number-counting task encourages the subject to store all items for long-term retention (Hinrichs & Grunke, 1975). After the interpolated task, each item was tested by presenting only the stimulus (vowel) and allowing the subject 8 sec to say the correct response (consonant).

The same timing, interpolated task, and study–test procedures were used for the main list of 20 number–noun pairs. However, after each test sequence, the study and test cards for all correctly recalled pairs were removed prior to the next trial. This was done to minimize possible overlearning effects (cf. Battig, 1965). After all pairs had been correct once, the card deck was reassembled and the procedure was repeated until the subject attained the criterion of one errorless trial on a single test of the entire list. To minimize study–test order effects, study and test cards were shuffled after each trial.

To ensure that subjects attended to the language as well as to the meaning of the word, they were carefully instructed that a response would be correct only if it occurred in the same language in which it had been presented.

Procedure: Retention tests and relearning.[1] Subjects returned after 5 weeks for the retention tests and relearning in the second session.

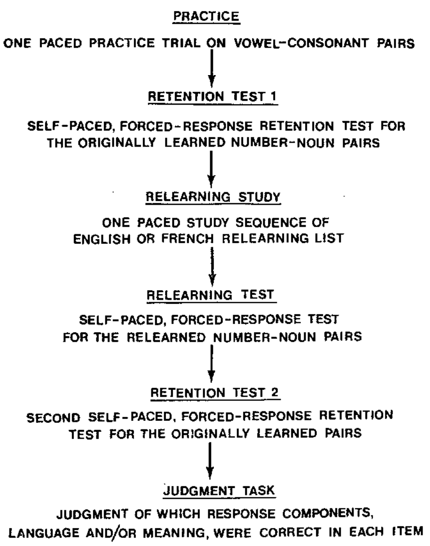

FIG. 3. The procedural sequence of events in the second session of Experiment I.

[1] The use of subjects naive to psychological experiments and the instructions that, after original learning, a second experiment would be conducted in 5 weeks helped to reduce the likelihood that subjects would anticipate the retention tests. These cautions appear to have been at least partially successful because no subject had more than 13 of the 20 items correct on Retention Test 1.

Figure 3 illustrates the procedure during the second session, in which all tests were recall tests. After the practice list, which refamiliarized the subject with the procedure and reduced warm-up effects on the main list, the first retention test of the originally learned pairs was given. All tests were self-paced, forced-response tests. Self-pacing was employed to allow subjects as much time as they needed to retrieve all possible items, even those with long latencies. Omissions were not permitted because they would result in ambiguity in defining a forgotten item.

After Retention Test 1, the subject had one paced relearning study trial on either the all-English list (11 subjects) or the all-French list (12 subjects) shown in Table 1. This was followed by 32 sec of number counting and then by the test of relearned items and Retention Test 2 of the originally learned items, as shown in Figure 3. Retention Test 2 was added to the basic savings procedure for two reasons. First, it is conceivable that some of the originally learned items would be reinstated by the relearning procedure and a second retention test could detect such a trend if it were present. Second, this additional retention test permitted a judgment task to be added in which the subject said for each Retention Test 2 response whether he thought he was correct on meaning only, correct on language only, correct on both, or guessing on both. This task was included in an attempt to determine how certain the subject was of the two components of his response, meaning and language.

Results and Discussion

Original learning. The mean number of trials to criterion can be represented in two ways—the number of *passes* through the entire 20-item study list, and the total number of *trials* on portions of the study list that were incorrect. The mean number of whole-list passes to criterion was 3.56 (range of 2–5, standard deviation of .79); the mean number of part-list trials to criterion was 9.04 (range of 5–15, standard deviation of 2.82).

Retention Test 1. Performance on Retention Test 1 was scored in two ways—a *lenient* scoring method allowing translations, and a *stringent* method disallowing translations. Because the outcomes of the two methods were so similar (probability of correct recall was .36 under the lenient system and .33 under the stringent system), analyses are conducted only on the scores obtained using the lenient method. The lenient scores are preferred because relearning analyses focus only on items initially forgotten; consequently, the lenient method biases the results against a significant savings effect and is, thus, the more conservative scoring method. In the analyses below, all significant statistics are reliable beyond $p < .001$ unless otherwise indicated; all nonsignificant statistics are stated for $p > .10$ unless otherwise indicated.

There was no recall advantage for either language on Retention Test 1: Mean probabilities of correct recall were .37 for French and .34 for English, $t(22) = .72$.

Using the lenient scoring system, the correlation between number of part-list trials to criterion and number correct on Retention Test 1 was −.64; using the stringent scoring system the correlation was −.61. Both of these correlations differ significantly from zero (using Fisher's *r*-to-*Z* transformation, $p < .01$ for the stringent), but do not differ significantly from each other. These correlations indicate that the subjects who learned more rapidly performed better on the delayed retention test and that this performance advantage was not due to the number of translation errors.

Relearning test. Performance was equivalent on the two relearning lists, French ($M = .51$) and English ($M = .57$), $t(21) = .77$, indicating no relearning advantage for either language.

The critical data concern the proportion of items correctly recalled on the relearning test that had been incorrect (in terms of both meaning and language) on Retention Test 1. The four conditional proportions presented in Table 2 represent the extent to which the residual (saved) information in a forgotten item facilitates relearning of items in each of four relationships to the originally learned item: Original language—original meaning (OL–OM), different language—original meaning (DL–OM), original language—different meaning (OL–DM), different language—different meaning (DL–DM). A two-way within subjects analysis of variance showed a significant savings effect for the meaning of the original words, $F(1, 22) = 20.25$, but not for the language in which they appeared, $F < 1$. The Meaning-by-Language interaction was also nonsignificant $F(1, 22) = 2.38$. Thus, the savings residual for nonrecallable items contains semantic information in a relatively language-free form.[2] This contrasts with the fact that, for recallable items on Retention Test 1, the language information is almost always correct.

TABLE 2

EXPERIMENT 1: MEAN PROPORTION OF ITEMS CORRECT ON THE RELEARNING TEST THAT WERE INCORRECT ON RETENTION TEST 1

Language in relearning	Meaning in relearning OM	DM	*M*
OL	.70	.34	.52
DL	.59	.41	.50
M	.65	.38	

Note. OL represents original language, DL represents different language, OM represents original meaning, and DM represents different meaning.

Retention Test 2. Performance improved considerably from Retention Test 1 to Retention Test 2, although subjects also made more

[2] "Relearning" also occurs for items *correctly* recalled on Retention Test 1, corresponding to the bilingual transfer paradigm (e.g., López & Young, 1974). Considering items scored as correct under the lenient system, only meaning significantly influences relearning of items correct on Retention Test 1, $F(1, 22) = 19.22$; the effect of language and the interaction of meaning and language are both nonsignificant (*F*s < 1). The pattern of mean proportions correct in relearning given that they were *correct* on Retention Test 1 (OL–OM, .80; DL–OM, .91; OL–DM, .46; DL–DM, .36) resembles the pattern in savings shown in Table 2.

translation errors on the second test. The mean proportion correct in recall on Retention Test 2 was .50 using the stringent method and .57 using the lenient method. Almost all of the translation errors (31 out of 34) involved incorrectly saying an originally English word in French. In fact, while there was essentially no language response bias on Retention Test 1 (.53 of all responses were French), there was such a bias on Retention Test 2 (.62 of all responses were French). This bias was significantly greater ($p = .005$, binomial test) on Retention Test 2 than on Retention Test 1; the bias on Retention Test 1 did not differ significantly from chance. It is probably the case that there was more confusion in selecting the correct language for a given response on Retention Test 2 than on Retention Test 1 because of the intervening relearning trial. Speculating further, subjects in an experiment on bilingual memory may expect to be tested on their knowledge of their *second* language, and therefore may have made their responses in French when confused by the relearning trial as to the original language of a given item. In fact, this possibility is one reason for looking at *incidental* language learning in Experiment II.

Of the items not recalled on Retention Test 1 (lenient scoring), .32 were reinstated in the correct language, .08 were reinstated as translations, and the remaining .60 were not reinstated on Retention Test 2. Table 3 shows the pattern of reinstatement as a function of relearning condition. As is clear from the table, most of the reinstatement is due to the two instances where original meaning is preserved. (Binomial tests showed each of the original-meaning conditions to be significantly different from each of the different-meaning conditions.) Not surprisingly, most of the translated reinstatements occur when meaning is preserved and language is different in relearning (DL–OM); what is surprising at first is that there is such a large proportion of original-language reinstatements in this case. However, because the two proportions do not differ significantly, it is possible that subjects simply guessed at language in the DL–OM condition.

TABLE 3

EXPERIMENT 1: PROPORTION OF CORRECT-LANGUAGE AND TRANSLATED REINSTATEMENT OF ITEMS ON RETENTION TEST 2 THAT WERE INCORRECT ON RETENTION TEST 1, AS A FUNCTION OF RELEARNING CONDITION

Relearning condition	Correct language	Translated
OL–OM (74)	.64	.05
DL–OM (76)	.30	.20
OL–DM (73)	.19	.01
DL–DM (72)	.14	.04

Note. The values in parentheses indicate the number of observations upon which the proportions in that row are based.

A crucial question with respect to savings is whether the savings effect is dependent upon originally learned items being reinstated during relearning, or whether savings can occur even when items are not reinstated. Thus, the question is whether the relearning item (e.g., DRAPEAU) restores the original item (e.g., FLAG), to which the subject then applies a response rule during the relearning test (e.g., "translate the original item"), or whether savings can occur even without restoration of the original item. This question can be answered using two conditional probabilities as dependent variables: (a) The probability of items showing savings on relearning given reinstatement on Retention Test 2, and (b) the probability of items showing savings on relearning given *no* reinstatement on Retention Test 2. Two separate two-way within-subjects analyses of variance were conducted, one on each of these dependent variables. The results are very similar for the two analyses: Both show significant savings for meaning (with reinstatement, $F(1, 22) =$

[3] The term "reinstatement" is used with respect to all of the four conditions for consistency. Of course, in the OL–OM condition, reinstatement is not distinguishable from new learning since the same item appears in both original learning and relearning.

52.86, $MS_e = .07$; without reinstatement, $F(1, 22) = 6.50$, $MS_e = .08$, $p < .05$) but no other significant effects. These findings indicate that the savings effect in relearning is not due simply to item reinstatement. Because the mean square errors and the degrees of freedom were very similar in these two analyses, it is reasonable to compare the variance accounted for by meaning to estimate the magnitude of the effect of reinstatement on savings. With reinstatement, the meaning variable accounts for more variance ($\omega^2 = .34$) than it does without reinstatement ($\omega^2 = .06$). Thus, although reinstatement increases the savings effect, it is not essential; savings occurs even without reinstatement.

The final data to be examined are the subject's judgments of which response components he had correct (language or meaning) for each item recalled on Retention Test 2. Table 4 presents subjects' judged response accuracy for the language and meaning components separately as a function of their actual response accuracy. The pattern is very similar in the cases where the subject has language and meaning correct and where he has only meaning correct—with a fairly high probability, he judges his response as correct on both components as long as the meaning component is correct. The pattern is also very similar for the cases where the subject has only the language correct—he judges that he has neither component correct about half the time and both components correct about a third of the time (judgments of the latter type usually involve intrusion errors). Thus, the subject appears to be responding only on the basis of correct versus incorrect meaning for a given item; his judgment as opposed to his recall of the correct language is surprisingly poor.

Conclusions

That input language can be retained over long periods when the subject is so instructed is demonstrated by the paucity of translation errors on Retention Test 1. Contrariwise, the relearning results indicate that input language is not a significant part of the savings residual for unrecalled items. Perhaps, as Liepmann and Saegert (1974) have suggested, there are input-language tags stored with the items, and these tags are lost more rapidly than are the items themselves. The finding that the savings residual for nonrecalled items contains meaning but not language information is in accord with the notion of conceptual identity of translation equivalents (cf. Kolers, 1966a, b). Furthermore, on the basis of prior research with the savings paradigm (Nelson, 1971b, Experiment 2; Nelson & Fehling, Note 1, Experiments 1 and 3) which demonstrated negligible synonymic savings, the presence of

TABLE 4

EXPERIMENT 1: JUDGED RESPONSE ACCURACY (MEANING AND/OR LANGUAGE CORRECT) AS A FUNCTION OF ACTUAL RESPONSE ACCURACY ON RETENTION TEST 2

	Judged correct			
Actually correct	Both meaning and language	Meaning only	Language only	Neither meaning nor language
Meaning and Language (228)	.86	.02	.04	.08
Meaning Only (33)	.67	.15	.06	.12
Language Only (103)	.27	.02	.18	.53
Neither meaning nor language (96)	.38	.00	.18	.45

Note. The values in parentheses indicate the number of observations upon which the proportions in that row are based.

savings for translation equivalents disconfirms the idea that translation equivalents are themselves synonyms (cf. Dalrymple-Alford & Aamiry, 1970).

Experiment II

Kintsch (1970) has demonstrated that subjects in an episodic memory experiment can, when so instructed, use either language-specific or language-free response strategies. In a continuous-recognition task, subjects were told to respond "old" for either (a) original meaning–original language only, or (b) original meaning regardless of language. They proved to be capable of either form of response and, in fact, could even respond accurately with whether the language of an old item was the same or changed on its second presentation. It is not surprising that subjects can respond on the basis of language information which has been *intentionally* learned (even after a 5-week interval, as in Experiment I of this paper), but the question remains as to whether *incidentally* learned input-language information is retrievable. The incidental case is of interest because it more closely parallels the everyday situation wherein anecdotal evidence suggests that bilinguals often do not remember the language in which a fact was originally learned (cf. Macnamara & Kushnir, 1971). Of course, these anecdotes often deal with facts which were learned very much earlier and which may have been re-encountered in the other language since original learning. To avoid such contamination, Experiment II examines retention of incidentally learned language and meaning information immediately after acquisition. Because Experiment I employed recall, recall tests also were used in Experiment II to maximize the comparability of the two experiments.

Two sets of incidental instructions were used. One set (Language) oriented the subject to the language of the word (English or French); the other set (Meaning) oriented the subject to the meaning of the word (Living or Nonliving). Consider first the retention of language information under the two instructional sets, Language and Meaning. Within a levels-of-processing framework (cf. Craik & Lockhart, 1972), two mutually exclusive (though not necessarily exhaustive) ways of accessing information are possible. The first is called *direct access*. If, in an incidental-learning task, the subject acquires only that aspect of the presented items that he is instructed to process or attend to (cf. Posner & Warren, 1972), then input language should be retained better in the Language condition than in the Meaning condition (in the limit, input language should not be retained in the Meaning condition). The second way of accessing information involves passage through nonsemantic toward semantic levels—this is called *indirect access*. If the subject proceeds through levels in an increasingly semantic order, then input language is necessarily processed before meaning, and input language might be retained as well, or nearly as well, in the Meaning condition as in the Language condition. This notion that processing proceeds from shallow (nonsemantic) to deep (semantic) has been widely assumed within the levels-of-processing framework, but has received little empirical support thus far. Experiment II permits a direct test of these two alternative access-route notions with respect to input-language retention.

Turning to retention of meaning, the two access-route notions do not appear to make differential predictions. By either account, retention of meaning should be greater in the Meaning condition than in the Language condition; thus, the critical test of how information is accessed for processing is embodied in the retention of input language and not in the retention of meaning. However, the data on retention of meaning are crucial to another aspect of the levels-of-processing framework. If this framework is to be distinguishable from a total-time hypothesis (cf. Cooper & Pantle, 1967), depth of processing and retention cannot always be positively correlated with time spent processing. Specifically, the prediction from the total-time hypothesis is that retention will be a direct function of the time spent processing, independent of the nominal depth to which items are processed. On the other hand, the prediction from the levels-of-processing framework is that retention will be governed primarily by the depth of processing (deeper processing resulting in better retention than shallower processing), not by the time spent processing. For this reason, item-processing times (decision latencies) are collected in Experiment II (cf. Gardiner, 1974).

Method

Subjects. The 32 subjects were faculty, staff, and students from the University of California, Irvine, as well as other adults from the Orange County area. Twelve of the subjects had spoken both English and French since early childhood. All of the remaining 20 subjects had studied their second language (French for 12, English for 8) for a minimum of 2 years at the university level. Using the scale discussed in Experiment I, all subjects rated themselves "highly competent" or "fluent" in English and at least "competent" in French. French–English bilinguals were chosen for maximum comparability to those in Experiment I. The self-ratings of the two sets of subjects in terms of English were very similar, although the subjects in Experiment II did tend to rate themselves as slightly more fluent in French than did the subjects in Experiment I.

Subjects participated individually in Experiment II and were paid $5.00 for the session.

Stimuli and apparatus. The stimuli were 20 English words and their French translation equivalents, two words from each of 10 categories in the Battig and Montague (1969) English category norms. Five of the categories represented living things and five represented nonliving things, as shown in Table 5. All of the category members were chosen from the range 5–10 in the norms, such that the living and nonliving sets had the same overall mean dominance. Translation equivalents were chosen to be orthographically and acoustically dissimilar to reduce the confusion possible during visual presentation. The words were typed in uppercase letters and photographed to produce slides. As is typical when using uppercase in French, accents were omitted—this had the added advantage of preventing subjects in the Language condition from being able to make their judgments for some of the French words solely on the basis of presence or absence of accents (i.e., without attending to the words themselves). Two randomizations of list order were created such that half of the words in each list were French and half were

TABLE 5

EXPERIMENT 2: THE ENGLISH STIMULI AND THEIR FRENCH TRANSLATIONS AS A FUNCTION OF WHETHER THEY ARE LIVING OR NONLIVING

Living		Nonliving	
English	French	English	French
LEMON	CITRON	STOOL	TABOURET
CHERRY	CERISE	DESK	PUPITRE
SPINACH	EPINARDS	GERMANY	ALLEMAGNE
LETTUCE	LAITUE	SPAIN	ESPAGNE
SPIDER	ARAIGNEE	COAT	MANTEAU
GRASSHOPPER	SAUTERELLE	HAT	CHAPEAU
CANARY	SERIN	WEEK	SEMAINE
CROW	CORBEAU	CENTURY	SIECLE
PIG	COCHON	WAGON	CHARRETTE
MOUSE	SOURIS	BOAT	BATEAU

English; language was assigned to individual items at random with the restriction that not more than three consecutive items could be in the same language. To counterbalance the language of the words, each of these lists was then translated, creating a total of four lists. For the eight subjects in each of the English, French, Living, and Nonliving groups, there were two replications per list.

The slides were presented using a Kodak Carousel projector with a tachistoscopic shutter. Once the slide projector was advanced, the experimenter pressed a button which simultaneously opened the shutter to display the slide and started a Lafayette digital stop clock to record the decision latency. The subject responded aloud into a microphone connected to a voice-activated relay; the response simultaneously stopped the clock and closed the shutter.

Procedure. Because the learning task was incidental, the subject was told only that the purpose of the experiment was to examine rapid word classification in response to a simple question; he was not informed of the retention tests that followed the processing task. In a between-subjects design, four groups of eight subjects were asked one of the following four classification questions: "Is the word in English?", "Is the word in French?", "Does the word represent something living?", or "Does the word represent something nonliving?" Thus, there were four question groups, the first two corresponding to the Language condition, and the last two corresponding to the Meaning condition. Subjects were assigned to list and question in random order of appearance.

The subject was instructed to respond aloud as rapidly as possible with a "YES" or "NO" as each word appeared, and his decision latency was recorded. Before beginning the experimental list, three randomized repetitions of the words MAN, HOMME, BOOK, and LIVRE were presented as practice trials to familiarize the subject with the task and to help stabilize response latency. Then after any procedural questions were answered, the experimental list was presented. Each word was shown individually and remained on the screen until the subject responded, after which the experimenter recorded the response and its latency. The resulting lag between a response and the next presentation was approximately five seconds; otherwise, list presentation was uninterrupted.

Following the last word in the list, the subject heard a set of instructions (taking about 30 sec to reduce recency effects) and then was given an oral self-paced free-recall test. After the recall test, a self-paced language-recognition test was administered. The subject was presented with a list containing all of the 20 words with their translation equivalents (side by side) and was instructed to circle the word in the language in which it originally had been presented (two-alternative forced-choice recognition), being certain to make a response for every pair of words. Before the experiment was concluded, the subject was asked to note any of the words on the list with which he was unfamiliar. If there were more than three such words, the subject was replaced; he was also replaced if he made more than three errors in his initial word-classification decisions. In all, seven subjects were replaced, six under the first restriction and one under the second. Finally, each subject was asked whether he had anticipated the retention test; all subjects reported that they had not expected to be tested.

Results and Discussion

The error rate on responses to classification questions over all subjects and question groups was less than 6%. Since the conclusions from subsequent analyses are unaffected by whether the incorrectly classified items are included or omitted, they are included in all cases. This choice was made so that none of the items acquired during classification processing would be overlooked in later analyses of retention. It should be noted that, in the analyses below, all significant statistics are reliable beyond $p < .05$; all nonsignificant statistics are stated for $p > .10$.

Some characteristics of the recall protocols should be mentioned before proceeding to the results of primary interest. First, as in Experiment I, translation errors in recall were very infrequent—only three of the 273 items recalled correctly with respect to meaning were recalled in the wrong language. Thus, even in an incidental task, language retention is excellent (proportion correct of .99). Intrusion errors were also very rare, comprising only 3% of the 291 words actually recalled, half of these 3% being the practice items. There were no repetitions, which is noteworthy in that subjects never said the same word twice, not even once in each language. This is in accord with the conclusion from Experiment I that translation equivalents represent the same underlying semantic concept.

In the second experiment, the mean proportion of French words recalled ($M = .48$) was significantly greater than the mean proportion of English words recalled ($M = .37$), $t(31) = 3.71$.[4] Although there is no *a priori* reason for this difference, it is also reflected in the median decision latencies where responses to French words ($Md = 1.07$ seconds) took significantly longer than did responses to English words ($Md = .97$ seconds), $t(31) = 2.89$. Since the subjects were, in general, less familiar with French than with English, this may have resulted in a longer time spent processing French words and, consequently, better retention of the French words (cf. the processing time arguments in Craik & Tulving, 1975).

The results of primary interest are the mean proportions of items correctly recalled and the median decision latencies for initial classification; these are presented as a function of question group in Table 6. Consider first the data on proportion correct in recall, shown in the top row of Table 6. As predicted, a planned comparison demonstrated that subjects in the Language condition (English and French question groups) recalled significantly fewer words than did subjects in the Meaning condition (Living and Nonliving question groups) $F(1, 28) = 5.03$, accounting for over 30% of the between-subjects variance ($\omega^2 = .33$). Neither of the remaining two orthogonal comparisons (English vs. French and Living vs. Nonliving) was significant (both $F\text{s} < 1$). Clearly, then, items are better retained following a semantic processing task than following one wherein only the language of the item must be processed.

Next, consider the data on the decision latencies for word classification, shown in the bottom row of Table 6. *A priori*, there was ample reason to expect that decisions in re-

[4] It is worth noting that, unlike in Craik and Tulving (1975), words given positive responses during processing were not retained significantly better than words given negative responses in the present experiment, $t(31) = .96$, the respective mean proportions correct being .41 for positive responses and .45 for negative responses. Perhaps subjects classify the items into two sets during acquisition (cf. differential storage in Epstein, 1972) along what they believe to be the most salient dimension—in this experiment, the language in which an item appeared might well have replaced the binary classification response as the basis for such a dichotomy.

TABLE 6

EXPERIMENT 2: MEAN PROPORTION CORRECTLY RECALLED AND MEDIAN DECISION LATENCY DURING CLASSIFICATION AS A FUNCTION OF QUESTION GROUP

	Question Group			
Dependent variable	English	French	Living	Nonliving
Proportion correct	.37	.38	.48	.48
Median latency	.92	.82	.92	1.42

sponse to the question "Does the word represent something nonliving?" would require more time than would decisions in response to the other questions, due to the required additional step of negation (cf. Clark & Chase, 1972). This expectation was borne out in the results of a planned comparison showing that the Nonliving group took significantly longer to decide than did the other three groups together, $F(1, 28) = 42.75$, the comparison accounting for over 80% of the between-subjects variance ($\omega^2 = .81$). Again, neither of the remaining two orthogonal comparisons was significant (both $Fs < 1$).[5] Given this finding, it is now possible to examine the question of whether greater depth of processing can be distinguished from longer time spent processing.

Taken together, the recall and latency data present an interesting picture. Although subjects took considerably longer (about 0.5 sec) to respond in the Nonliving group than in the Living group, they recalled the same proportion of the words classified. Furthermore, although subjects in the Living group took approximately the same amount of time to make their original decisions as did subjects in the English and French groups, those in the Living group recalled a greater proportion of the words classified. Finally, correlations of median decision latency with proportions of items correctly recalled are consistently nonsignificant. For the four question groups, the correlations are rather unreliable since they are based on only eight subjects each; more informative are the correlations over all 32 subjects ($r = .31$) and the correlations (based on 16 subjects) within the Language condition ($r = -.08$) and within the Meaning condition ($r = .25$). These results are in accord with those of Gardiner (1974) and Craik and Tulving (1975, Experiment 5) in demonstrating that the total-time hypothesis cannot encompass as diverse a set of findings as can the levels-of-processing framework.

The meaning recall and decision latency results are quite straightforward; however, a surprising finding emerges in the data on input-language recognition. Contrary to the predictions made above, the proportion of items for which the input language was correctly recognized was significantly *higher* in the Meaning group than in the Language group (shown in the column marginal of Table 7), $z = 2.27$. This finding, true for both recalled and nonrecalled words, directly contradicts the prediction of the direct access notion that input-language recognition should be better in the Language condition than in the Meaning condition. Furthermore, although the indirect access notion can be modified *post hoc* to account for this result, the result remains counterintuitive. However, because this effect is rather small (a difference of only 4%), caution should be exercised in its interpretation.

TABLE 7

EXPERIMENT 2: THE PROBABILITY OF CORRECT LANGUAGE RECOGNITION AS A FUNCTION OF PROCESSING CONDITION AND PREVIOUS RECALL PERFORMANCE

Previous recall performance	Processing condition: Language	Processing condition: Meaning	*M*
Recalled	.96	.99	.97
Nonrecalled	.90	.95	.93
M	.93	.97	

Table 7 also depicts, in the row marginal, another small but reliable effect. Recognition of input language was better for words that were previously recalled than for words that were previously nonrecalled, $z = 2.23$. Once again, this result should be interpreted with caution, but it does suggest that the retention

[5] This negation effect of about 0.5 sec is considerably larger than that observed by Clark and Chase (1972), the reason for which is not obvious. However, it is worth noting that the latencies themselves are considerably longer than those in Clark and Chase and that the negation in the present experiment is not as straightforward as is the simple "not" in their study.

of input language for a word depends to some extent upon whether that word can be retrieved.

Conclusions

Words that are initially evaluated on the basis of their meaning subsequently are recalled better than are the same words evaluated on the basis of the language in which they were presented. This retention difference can be accounted for by the depth of processing achieved in making the initial classification of the words—processing time during acquisition does not provide an adequate account of retention performance. Surprisingly, input-language recognition is also better following the semantic-orienting task than following the language-orienting task. This recognition finding disconfirms a direct access model of levels of processing and, although not irreconcilable with an indirect access model, nonetheless poses some difficulties for such a view. This issue merits further investigation.

Experiment II demonstrates that input-language information is stored during acquisition within an incidental-learning task. Recall following incidental learning (Experiment II) is similar to that following intentional learning (Experiment I) in that correctly recalled items almost always are recalled in the input language. Furthermore, although input-language recognition is high overall, it is somewhat better for previously recalled than for previously nonrecalled words, suggesting that retention of the language of an item is at least partly dependent on retention of its meaning.

General Discussion

A question that is central to understanding bilingualism concerns the representation in memory of a word and its translation equivalent (e.g., HORSE and CHEVAL)—are they synonyms or are they simply different ways of identifying the same underlying concept? Experiment I directly addressed this question. Because input-language information was not contained in the savings residual for non-recalled items but significant savings occurred for the meaning of translation equivalents, these results inductively support the notion of a single underlying concept for translation equivalents (cf. Kolers, 1966a, b). Furthermore, since the relearning of synonyms does not result in significant savings (e.g., Nelson, 1971b, Experiment 2), whereas the relearning of translation equivalents does result in significant savings, the possibility that two words such as HORSE and CHEVAL are retained as synonyms is disconfirmed. Rather, the two words appear to share the same supra-linguistic semantic representation in memory (cf. the type-token distinction in Anderson & Bower, 1973).

The above discussion focuses on the representation of the meaning of translation equivalents, but the representation of their input language must be considered as well. The notion of input language tags (Liepmann & Saegert, 1974; Saegert, Hamayan & Ahmar, 1975) is one possible representation and appears compatible with the results of Experiment I. Basically, the argument is that the meaning of the word is stored in memory and the word's language is affixed to the semantic trace. Consistent with this tagging notion, Experiment I demonstrated that language information is not a significant part of the savings residual for words whose meaning cannot be retrieved. These results support the already prevalent view that input language is stored in the form of a tag on the language-free semantic representation of a word.

If language tags are employed to retain input language, then the degree to which the expected retention test influences the establishing of these tags must be examined. To date, virtually all of the research on bilingual memory has emphasized the acquisition of both language and meaning by using intentional learning paradigms. This emphasis may be made explicit (as in Experiment I); certainly, it is always implicit in that the subjects are bilingual and are learning mixed-language lists (cf. demand characteristics in Orne, 1962). Assuming that input-language information is ordinarily not useful to retain, it might be hypothesized that subjects would not encode input language if they did not expect a retention test and if they did not have to use the language information *per se* during processing. This hypothesis was tested in Experiment II using an incidental-learning paradigm, and the hypothesis was refuted. In a condition in which classification pertained only to meaning, subjects actually retained input language better than did subjects whose classifications pertained to the input language itself. In Experiment I, input-language retention was excellent 5 weeks after intentional learning; in Experiment II, input-language retention was excellent immediately after incidental learning. Because words correctly recalled with respect to meaning are nearly always in the correct language in these two experiments, input language appears to be stored regardless of intention to learn it.[6]

The results of these two experiments have been related to the study of bilingualism and must now be related to the study of memory. Turning first to Experiment I, the savings method (Nelson, 1971a, b) was used to examine the residual information in the memory traces of nonrecalled bilingual items. The findings regarding lack of synonymic savings have been useful in evaluating the notion of a language-free semantic concept, and the generalizability of the savings paradigm has been extended. Furthermore, the finding that savings occurs for items not reinstated by relearning (as well as for items reinstated by relearning) strengthens the argument that the savings method is highly sensitive as a measure of retention (cf. Nelson, 1971a).

The findings of Experiment II supported the levels-of-processing framework (cf. Craik & Lockhart, 1972). As predicted, a language-classification task resulted in poorer retention of meaning than did a meaning-classification task. That this retention difference can be attributed to the depth of processing in the two conditions and not to the time spent processing is demonstrated by the pattern of classification latencies and by the lack of a significant correlation between the items recalled and their initial classification latency. An unexpected finding was that recognition of input language was better after the meaning-classification task than after the language-classification task. This invalidates the direct access notion that only the level to be evaluated is accessed during classification. Instead, the shallower levels seem to be processed en route to the deeper ones. Although the reason for the better retention of input-language following meaning classification still is not obvious, one possibility is that language somehow is mediated by meaning.

In terms of the study of bilingual memory, the two studies reported here have contributed at several levels. Experiment I is the first study of long-term retention in bilinguals and helps to clarify the nature of the representation of meaning in bilingual memory. Experiment II is the first study of incidental acquisition in bilingual memory using the levels-of-processing methodology and suggests that the encoding of language occurs at a shallower level than does the encoding of meaning. Together, the two experiments demonstrate that bilinguals store the input language of words exceedingly well, regardless of their intention to learn input language. There is also the suggestion that language recognition depends somewhat on the retrievability of meaning, although the extent of this dependence remains to be investigated. The principal conclusions to be drawn are that a word is represented in a bilingual's memory as a language-free semantic trace, and that input language is attached, perhaps in the form of some kind of language tag, to that semantic trace.

[6] This is certainly the case at the level of the single word; however, at the level of the sentence, there is evidence to indicate that this remarkable accuracy breaks down somewhat (cf. Macnamara & Kushnir, 1971; Rose, Rose, King & Perez, 1975).

References

ALSTON, W. P. *Philosophy of language*. Englewood Cliffs, New Jersey: Prentice-Hall, 1964.

ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. Washington, D.C.: V. H. Winston, 1973.

BATTIG, W. F. Procedural problems in paired-associate learning research. *Psychonomic Monograph Supplements*, 1965, **1**, No. 1.

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**, (3, Pt. 2).

BATTIG, W. F., & SPERA, A. Rated association values of numbers from 0–100. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 200–202.

CLARK, H. H., & CHASE, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, **3**, 472–517.

COOPER, E. H., & PANTLE, A. J. The total-time hypothesis in verbal learning. *Psychological Bulletin*, 1967, **68**, 221–234.

CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, **104**, 268–294.

DALRYMPLE-ALFORD, E. C., & AAMIRY, A. Word associations of bilinguals. *Psychonomic Science*, 1970, **21**, 319–320.

EPSTEIN, W. Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1972. Vol. 6.

FISHMAN, J. A., & COOPER, R. L. Alternative measures of bilingualism. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 276–282.

GARDINER, J. M. Levels of processing in word recognition and subsequent free recall. *Journal of Experimental Psychology*, 1974, **102**, 101–105.

GLANZER, M., & CUNITZ, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

HINRICHS, J. V., & GRUNKE, M. E. Control processes in short-term memory: Use of retention interval information. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 229–237.

KINTSCH, W. Recognition memory in bilingual subjects. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 405–409.

KINTSCH, W. *The representation of meaning in memory*. Hillsdale, New Jersey: Erlbaum, 1974.

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

KOLERS, P. A. Reading and talking bilingually. *American Journal of Psychology*, 1966, **79**, 357–376. (b)

LANGER, S. *An introduction to symbolic logic*. New York: Dover, 1953.

LIEPMANN, D., & SAEGERT, J. Language tagging in bilingual free recall. *Journal of Experimental Psychology*, 1974, **103**, 1137–1141.

LÓPEZ, M., & YOUNG, R. K. The linguistic interdependence of bilinguals. *Journal of Experimental Psychology*, 1974, **102**, 981–983.

MACNAMARA, J., & KUSHNIR, S. Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 480–487.

NELSON, T. O. Recognition and savings in long-term memory: Related or independent? *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **5**, 15–16. (a)

NELSON, T. O. Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 568–576. (b)

NELSON, T. O., & ROTHBART, R. Acoustic savings for items forgotten from long-term memory. *Journal of Experimental Psychology*, 1972, **93**, 357–360.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values of 925 nouns. *Journal of Experimental Psychology Monograph*, 1968, **76** (1, Pt. 2).

POSNER, M. I., & WARREN, R. E. Traces, concepts, and conscious constructions. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972.

POSTMAN, L., & PHILLIPS, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 132–138.

QUINE, W. V. O. *Word and object*. New York: Wiley, 1960.

ROSE, R. G., ROSE, P. R., KING, N., & PEREZ, A. Bilingual memory for related and unrelated sentences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 599–606.

RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972.

SAEGERT, J., HAMAYAN, E., & AHMAR, H. Memory for language of input in polyglots. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 607–613.

STERN, N. G. *Meaning and change of meaning*. Bloomington: Indiana University Press, 1964.

TULVING, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972.

REFERENCE NOTE

1. NELSON, T. O., & FEHLING, M. R. *Semantic information in long-term episodic memory*. Unpublished mimeo, University of Washington, 1974.

(Received January 9, 1976)