False memory in bilinguals: Does switching languages increase false memories?

GLORIA MARMOLEJO
Winona State University, Rochester

KRISTEN A. DILIBERTO-MACALUSO
Berry College

JEANETTE ALTARRIBA
University at Albany, State University of New York

People often receive and recount information in different languages. This experiment examined the impact of switching languages on false recall, recognition, and recognition confidence. We presented Spanish–English bilinguals with 10 lists of words associated to a critical nonpresented lure, either in English or in Spanish. Each list was followed by free recall either in English or in Spanish. The final stage was a recognition test in either language. Results showed a higher proportion of veridical and false recall in English, the more dominant language, than in Spanish, the native language. Noncritical intrusions were equivalent in both languages. More importantly, false recall, false recognition, and false recognition confidence were higher across languages than within languages. The results are examined in relation to current research and interpretations of bilingual false memory.

The study of false memory, which is memory of events that did not happen, has received much attention in the media and the scientific community. In the widely used Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995), participants may be presented with a list of words (e.g., *table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking,* and *bench*), which are all related to a nonpresented common associate or critical lure (e.g., *chair*). On later memory tests, participants often claim to remember the nonpresented critical lure word with surprisingly high levels of subjective confidence (Lampinen, Neuschatz, & Payne, 1998; Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger & McDermott, 1995).

Recent investigations using the DRM paradigm have examined the influence of language on false memory. From a practical standpoint, there are important reasons for studying this variable. First, there are thought to be about 3,500 different languages (Steigerwald, 1987), spoken by 5,000 to 8,000 different ethnic groups residing in approximately 160 nation-states (Stavenhagen, 1990). From those, an increasing number of people often receive information in one language (e.g., in English at work) and then recount it in another (e.g., in Spanish at home).

American Journal of Psychology
Spring 2009, Vol. 122, No. 1 pp. 1–16 •

Consequently, it is crucial to learn the mechanisms involved in transferring information from one language at encoding to the other language at retrieval and to understand the degree to which false memories occur with such a transfer in stimulus format.

A second reason relates to eyewitness testimonies from bilingual witnesses in the courtroom (Rubin, 1996; Wright, Loftus, & Hall, 2001). As bilinguals are questioned in a language that represents their nondominant language, they may actually produce fewer or less accurate details and perhaps communicate those details with less intensity of emotion than when the native language is used (see Altarriba, 2003; Altarriba & Santiago-Rivera, 1994; Santiago-Rivera & Altarriba, 2002, for further discussion). Given the importance of a witness's testimony in supporting the verdict of innocence or guilt for some defendants, it would be important to know whether rendering a description of events from memory should be conducted in one or both languages to preserve its accuracy and completeness. In fact, the cognitive interview technique (see Campos & Alonso-Quecuty, 1998; Fisher, Geiselman, & Amador, 1989) suggests that bilinguals be questioned in both of their languages rather than just one.

Research examining the influence of language on false memories has applied the DRM paradigm to monolinguals who speak Spanish from Spain (García-Bajos & Migueles, 1997; Pérez-Mata, Read, & Diges, 2002), Spanish from Mexico (Anastasi, Rhodes, Marquez, & Velino, 2005), Portuguese (Albuquerque & Pimentel, 2005; Stein & Pergher, 2001, cited in Sahlin, Harding, & Seamon, 2005), and Japanese (Kawasaki & Yama, 2006). For example, García-Bajos and Migueles (1997) presented Roediger and McDermott (1995) lists translated into Castilian Spanish to monolingual Spanish speakers from the Basque region of Spain. They replicated both false recall ($M = .38$) and false recognition ($M = .70$) of critical nonpresented lures. Thus, it appears that false recall and recognition are not limited to English.

Of these studies using monolingual word lists, few have explicitly compared the likelihood of triggering false memories to English. Pérez-Mata et al. (2002) found that the proportion of false recall in Castilian Spanish (Experiment 2, $M = .24$) was half of that obtained in English (Experiment 1, $M = .48$). However, because each language was tested in separate experiments (instead of being explicitly manipulated as a factor), and each language had a different set of lists and critical lures, one must interpret this comparison across languages cautiously. One cannot ignore the possibility that the four lists presented in English might have triggered a higher proportion of false recall than the four lists presented in Spanish. In fact, Stadler, Roediger, and McDermott (1999) uncovered a wide variability in the proportions of false memories that different DRM lists trigger, even though all lists consisted of the 15 strongest associates of the critical word, listed in descending order of association. For example, whereas the probability of falsely recognizing the critical nonpresented word *window* was .84, *king* was falsely recognized only with a probability of .27. Similarly, the probability of false recall ranged from .10 to .65 across lists.

Few studies have explored the influence of language on false memory effects in bilinguals, and all but one (Molina, Su, Knight, Holliday, & Weekes, 2006) were behavioral. It is important to investigate false memory in bilinguals in addition to monolinguals because as Grosjean (1989) pointed out, a bilingual is not merely the sum of two monolinguals. He added that bilinguals have a unique language structure and organization. Therefore, one must be cautious when comparing performance across these two groups.

In one study investigating false memory in bilinguals, Miyaji-Kawasaki, Inoue, and Yama (2003) found a higher proportion of critical false alarms when the test language was different from English, in this case Japanese, the native language. Participants were Japanese–English bilinguals from Japan with a wide range of proficiency in English. They viewed DRM lists, half in English and half translated to Japanese. Then, they received a list of words, half in English and half in Japanese. Their task was to recognize the studied items from the new items, and for old items, to indicate their presentation language. An open question is whether the lower false recognition rate in English was due to a low proficiency in English and whether the results generalize to bilinguals with more equivalent proficiency in both languages.

In fact, research with Chinese–English bilinguals from Singapore (Holliday, Kang, & Lee, 2003) and with Spanish–English bilinguals born in Mexico and residing in Arizona (Anastasi, Rhodes, et al., 2005) supports the idea that false recall and recognition are constrained by proficiency in the second and less dominant language. Anastasi et al. presented their Spanish–English bilinguals with word lists in one language and a recognition test in the same language. Then, they presented a set of word lists in the other language and a recognition test in that same language. In an attempt to control for associative power of the lists in the two languages, the authors used six of Roediger and McDermott's (1995) lists in the English condition (*mountain, needle, rough, sleep, soft, sweet*) but did not translate the lists into Spanish for the Spanish condition. Instead, they built different lists in Spanish by asking native Spanish speakers to generate associates to a different set of critical lures (*chair, cold, doctor, bread, fruit,* and *foot*) and choosing the 15 most common responses. Results showed a higher proportion of false alarms to critical lures in English, the more dominant language, than to critical lures in Spanish, the native language (Experiment 2). This effect disappeared with monolingual Spanish speakers (Experiment 3) and monolingual English speakers (Experiment 4).

A question related to this research is whether shifting languages between encoding and retrieval affects the generation of false memories. In the case of veridical bilingual memory, information processing in one language has been found to either assist or hinder the processing and recall of translation equivalents, depending on the task (Altarriba, 2000). It is also possible that changing language from encoding to retrieval might not affect the magnitude of false memories, just as postevent misinformation presented to Spanish–English bilinguals had comparable detrimental effects in recognition accuracy across and within languages (Shaw, García, & Robles, 1997).

Support for a language-independent view using the DRM paradigm comes from Cabeza and Lennartson (2005), who presented some lists in English and some in French to English–French bilinguals from Alberta, Canada. The recognition test included intermingled English and French items. "Old" answers applied only to test items that matched an encoded item in content and language; "new" answers pertained to words that either were not presented at encoding or were presented in a different language. The results showed an equivalent amount of false recognition of critical nonpresented lures in French and English, both within and across languages.

In contrast, support for a language-dependent false memory effect comes from Sahlin et al. (2005). They presented DRM English and Spanish word lists auditorily to proficient Spanish –English bilinguals. After study, participants were given a written recognition test in the same or different language. The same procedure was repeated for five consecutive study–test trials. Sahlin et al. found that during the initial trials, false recognition errors were high for both same- and different-language words because of the use of gist representations; however, during later trials false recognition errors decreased for both same- and different-language critical words because of the use of language-specific lexical representations and decreased reliance on conceptual representations (Sahlin et al., 2005). In addition, the researchers found that false recognition of critical words was greater when the study and test languages matched than when they differed, thereby supporting a language-dependent view.

EXPERIMENT

The research just reviewed suggests that false recall and recognition for monolingual word lists can be replicated in languages other than English and that false recognition in bilinguals may be constrained by language proficiency and task demands. The present research sought to replicate and extend research on bilingual veridical and false memory in the following ways. First, we investigated whether false recall and recognition could be replicated using materials that can apply to Spanish-English bilinguals born in a variety of countries. Second, we re-examined whether language congruence between encoding and retrieval is a necessary condition for the creation of false memories and, if not, whether false memories across languages are as strong as those generated within language when unlike previous studies, item retrieval does not require reporting in presentation language. In our study, Spanish–English participants heard DRM word lists presented in either Spanish or in English. Unlike those in previous studies on bilingual false memory (e.g., Anastasi, Rhodes, et al., 2005; Holliday et al., 2003; Sahlin et al., 2005), our participants received a written recall test after each list, which was in the same or different language as the word list presented during study. After recall, participants completed a final written recognition test on which they indicated whether they heard each word during the study phase and their level of confidence. The recognition test was in either the same or a different language as the language of the word lists during study (see Robinson & Roediger, 1997, and Roediger & McDermott, 1995, who used similar procedures with monolingual participants).

METHOD

Participants

The participants were 119 (22 males, 97 females) Spanish–English bilinguals from the Florida International University community in Miami, Florida, between the ages of 15 and 44, $M = 20.63$, $SD = 4.91$. Participants were born in a variety of locales (the United States, Central America, South America, Caribbean, other) but were permanent residents of South Florida, where both languages are used on a daily basis. Eligibility was determined through a preliminary phone interview and a language history questionnaire (Altarriba, 1992). The language history questionnaire assessed the length of exposure to both languages, the percentage daily usage of each language, the number of years residing in the United States (with a minimum of three years required), the participants' native country (if not the United States), and the number of years attending schools in the United States. All participants were regarded as fluent based on self-report ratings of spoken, written, and vocabulary comprehension ($M = 7.8$ to 9.7 on a scale from 1 to 10) in English and in Spanish, but their current dominant language was English (Table 1). Participation was compensated either with extra academic credit or payment.

TABLE 1. Mean (*SD*) self-report ratings of spoken, written, and vocabulary in English and in Spanish

Comprehension	English	Spanish
Spoken*	9.54 (0.82)	8.92 (1.34)
Written*	9.50 (0.80)	8.34 (1.65)
Vocabulary*	9.02 (1.21)	7.94 (1.87)

*$p < .05$.

Design

We implemented a 2 (encoding language) × 2 (recall language) × 2 (recognition language) mixed design. Encoding and recognition were manipulated between participants by randomly assigning participants to one of four encoding–recognition language combinations: English–English, Spanish–Spanish, English–Spanish, and Spanish–English. Recall language was manipulated within participants: Everybody recalled half of the studied lists in English and half in Spanish, with the order intermingled.

Materials

Thirty lists of 12 words each were developed, 15 in English and 15 in Spanish. Most English lists were adapted from Stadler et al.'s (1999) false memory norms. All the words in each list were associates of a common nonpresented critical lure. The lists combined two properties: They had been shown to produce medium to high levels of false recall (31% or more) and false recognition (58% or more) of the critical lure in English, and their items could be suitably translated into Spanish.

The process of list and item selection involved five stages to tackle the challenges encountered when translating from English to Spanish.[1] First, we checked the eligibility of individual items. Words from Stadler et al.'s (1999) original lists were discarded if their translation to Spanish presented any of these eight problems:

It was identical to that of the critical nonpresented lure (e.g., chilly = frío and COLD = FRIO).

It shared a stem with the translated critical nonpresented lure (e.g., shutter = contraventana and WINDOW = VENTANA).

It was a homonym of the critical nonpresented item's translation (e.g., feather = pluma and PEN = PLUMA).

It was a feminine adjective (e.g.,small = pequeña) that could not be changed into a neutral word in a list with a feminine critical lure (SPIDER = ARAÑA) so the item would not stand out.

It consisted of a prepositional phrase or more than two words (e.g., landfill = vertedero de basura) to avoid word length confounding across languages.

It varied widely for different countries (e.g., cake = pastel, tarta, torta, biscocho, ponqué, quequé).

It did not exist in Spanish (e.g., galoshes = cobertura impermeable colocada sobre los zapatos).

It was a cognate that shared all or all but one character (e.g., piano = piano) to allow the study of cross-language false memories.[2]

Given that only nine list items have been needed to produce reliable false recall and recognition (Robinson & Roediger, 1997), list length was standardized to 12 items. If an original 15-item list had exactly three problematic items, these items were discarded. If a list had fewer than three problematic items, the lowest associates were also eliminated. If three to six items were problematic, three were eliminated and the rest were replaced by other associates and their Spanish translation, taken from Russell and Jenkins's (1954) association norms, or when absent in the norms, by a word that according to the authors' judgment would elicit the critical word.

The second stage was to check the eligibility of the lists. A list was eliminated if one of these situations applied:

There were more than six problematic items (40% difference or more from the original list).

More than one critical word in Spanish could be elicited by the associates of the list (e.g., *SMOKE* = *HUMO* and *FUMAR*).

The translation of the critical word varied across countries (e.g., *RUBBER* = *HULE* in Mexico, *CAUCHO* in Spain).

The critical word was an identical cognate (e.g., *DOCTOR* = *DOCTOR*).

The critical word was a homographic noncognate (e.g., *FOOT* = *PIE*).

Multiple items in English translated into the same word in Spanish (e.g., *trash, garbage, refuse, rubbish, litter* = *basura*).

Most items in the list pertained uniquely to the United States (e.g., for *FLAG, stars, stripes, checkered, freedom*).

After eliminating unsuitable lists, we constructed three new lists using unpublished Castilian Spanish association norms (Algarabel, Sanmartín, García, & Espert, 1986) and translated them into English to complete 15 lists in each language.

The third stage involved making adjustments to the Spanish lists such as choosing between nouns and verbs and choosing adjective gender. English words that are used both as noun and verb were translated into a noun (e.g., *hate* = *odio* instead of *odiar*) because, based on informal observations, Spanish speakers produce nouns as associates more often (97%) than verbs (3%) (Algarabel, Ruiz, & Sanmartín, 1988). Only when the critical word in English could be either a verb or a noun (e.g., *SLEEP*), the verb translation was used (e.g., *yawn* = *bostezar* instead of *bostezo*). Adjectives were always translated to the masculine gender (e.g., *hesitant* = *indeciso* instead of *indecisa*) except in lists with a feminine noun as the critical lure (e.g., *NEEDLE* = *AGUJA*) and adjectives as items (e.g., *sharp*). In these cases, a neutral adjective translation was chosen over the masculine (e.g., *punzante* instead of *picudo*) to keep adjective–noun agreement and thus maximize the likelihood that the critical lure would be falsely remembered. Feminine adjectives that qualified critical lures and that did not have a neutral equivalent were eliminated.

The final stage was to check that our materials were appropriate for speakers from a variety of countries. For items with multiple correct translations to Spanish, we selected the one understandable across regions according to the input obtained from 10 bilingual judges from nine different countries (Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Mexico, Puerto Rico, and Spain). The final lists appear in the Appendix.

For counterbalancing purposes, the 15 lists in English were arbitrarily divided into three sets of 10 lists each (Set 1 contained Lists 1–10, Set 2 contained Lists 11–15 and 1–5, and Set 3 contained Lists 6–15). The 15 Spanish lists were also divided into the same three sets for the same purposes. Each set of 10 lists was recorded onto a CD by an English–Spanish fluent bilingual male voice at a rate of 2 s per word. Participants were randomly assigned to one of the resulting six discs of 10 lists. Five of these lists were followed by an immediate free recall test in English, and the other five were followed by an immediate free recall test in Spanish, with the order of the recall language determined randomly. The remaining five lists were not studied but were instead reserved to get new items for the recognition test. Each list served equally often in each of these conditions.

The recognition test consisted of 60 words arranged in the same random order for all participants, one per row, 30 of which were previously studied (first, fifth, and ninth items of each of the 10 studied lists). The remaining 30 items in the test were not previously studied: the 10 critical nonpresented lures related to the words in the 10 studied lists, 15 nonstudied lures (first, fifth, and ninth items from each of the five nonstudied lists), and the five nonstudied lures related to the words in the nonstudied lists. There were two versions of the test, one in English and one in Spanish.

Procedure

The experiment consisted of three stages. In Stage 1, participants listened to 10 lists of words, all in English or all in Spanish. Each list was followed by the expression "now recall" in the language of presentation and by either a ring or a knock to indicate whether to write down the list words in English or in Spanish for 1.5 min. For half of the participants, the ring indicated recall in English and the knock indicated recall in Spanish. For the rest of the participants, the signals were reversed. The order of the signals was randomized.[3] The language of the instructions matched the language of the lists to encourage participants to think in that language during this portion of the experiment. After 1–2 min of explanation, participants proceeded to Stage 2, consisting of answering a written recognition test either in English or in Spanish, at their own pace. Participants indicated whether they had heard the item on the disc (disregarding the presentation language) by providing a combined "yes/no" and a confidence rating between −3 and +3 for each item, where −3 = *completely sure I did not hear it*, −2 = *fairly sure I did not hear it*, −1 = *somewhat unsure I did not hear it*, +1 = *somewhat unsure I heard it*, +2 = *fairly sure I heard it*, and +3 = *completely sure I heard it*. A *u* (for *unknown*) was used to indicate that the meaning of the word was unknown in the conditions where the encoding language did not match the recognition language. The language of the instructions matched the language of the test to encourage participants to think in that language in that portion of the experiment. Finally, in Stage 3, participants completed the language history questionnaire.

RESULTS

Preliminary analyses

The translation equivalents of the lists were clear to participants despite their varied countries of origin, as judged by the low percentage of items they marked as unknown in the Spanish recognition test of the English–Spanish condition (1.67%). In fact, the percentage of words that were marked as unknown in the English recognition test of the Spanish–English condition was the same (1.67%).

The different sets of lists did not influence veridical recall, false recall, veridical recognition, or false recognition, all Fs < 2.74, ps $> .07$. (The significance level for all analyses was set at .05.) The three-way interactions with the list factor were not significant either, all Fs < 1.52, ps $> .22$. Therefore, subsequent analyses were performed on encoding and retrieval languages collapsing over lists. Central tendency and variability results of free recall, recognition, and recognition confidence are summarized in Table 2.

Veridical recall

A free recall item was scored as correct when it was present in the original list and was reported in the requested language, disregarding number, gender, and lexical category in Spanish. For example, for the item *happy*, the responses *contento, contenta, contentos, contentas,* and *feliz* were all scored as correct in the Spanish recall condition.

There was a main effect of recall language, where the proportion of correct responses was significantly higher in English than in Spanish, $F(1, 117) = 111.75$, $p < .001$, $MSE = .004$. This was not surprising because English was the participants' dominant language. More importantly, there was a significant encoding × recall interaction, $F(1, 117) = 353.68$, $p < .001$, $MSE = .004$, such that within-language recall was more accurate than across-language recall, $t(118) = 13.70$, $p < .001$. We also compared encoding and recall in English with encoding in English and recall in Spanish and found greater accuracy in the within-language condition, $t(60) = 22.05$, $p < .001$. Similar results were obtained when we compared the Spanish–Spanish with the Spanish–English condition, $t(57) = 5.50$, $p < .001$.

False recall of critical nonpresented lures

False recall was defined as the proportion of critical nonpresented lures from the presented lists reported in the requested language. Just as in Spanish veridical recall, number, gender, and lexical category were disregarded.

Not surprisingly, false recall was comparable to that obtained in previous research when only English was used (37%). In addition, critical lures were

TABLE 2. Mean proportions as a function of encoding–retrieval language congruence

	ENGLISH–ENGLISH		ENGLISH–SPANISH		SPANISH–ENGLISH		SPANISH–SPANISH	
	M	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
				Free recall				
Correct	.56	.01	.31	.01	.42	.01	.49	.01
Critical intrusions	.37	.04	.46	.03	.44	.04	.19	.03
Noncritical intrusions[a]	.32	.04	.66	.06	.67	.04	.21	.06
				Recognition				
Hits	.84	.02	.77	.02	.76	.02	.82	.02
Critical false alarms	.80	.03	.87	.03	.80	.03	.73	.03
Noncritical false alarms	.13	.03	.22	.03	.12	.03	.20	.03
				Recognition confidence				
Hits	2.80	.03	2.82	.04	2.81	.03	2.89	.04
Critical false alarms	2.68	.05	2.78	.06	2.71	.05	2.60	.06
Noncritical false alarms	1.93	.14	2.38	.13	2.09	.13	1.85	.14

[a]*This row represents means, not proportions.*

reported in all combinations of encoding and recall languages in various degrees as follows. A significant encoding language main effect indicated that the proportion of false recall was higher with English than with Spanish encoding, $F(1, 117) = 6.45$, $MSE = .09$, $p = .01$. There was also a recall language main effect. The proportion of false recall was higher when items were reported in English than in Spanish, $F(1, 117) = 7.49$, $MSE = .05$, $p = .007$. Finally, there was an encoding × recall interaction, $F(1, 117) = 34.73$, $MSE = .05$, $p < .001$. Critical nonpresented lures appeared more frequently across languages than within languages, $t(118) = -5.67$, $p < .001$. Also, we compared encoding and false recall in the English–English condition with that in the English–Spanish condition and found that false recall was greater in the cross-language condition, $t(60) = -2.12$, $p = .038$. Similar results were obtained when we compared the Spanish–Spanish with the Spanish–English condition, $t(57) = -6.50$, $p < .001$.

Recall of noncritical intrusions

Noncritical intrusions were defined as the proportion of noncritical nonpresented words mistakenly reported. These included presented words reported in the wrong language. Language mixing accounted for a negligible percentage of the intrusions (0.02%). Results showed that noncritical intrusions did not vary as a function of encoding language, $F(1, 117) = .72$, $MSE = .14$, $p = .40$, or recall language, $F(1, 117) = 1.79$, $MSE = .18$, $p = .18$. Just as with critical lures, however, there was an encoding × recall interaction, $F(1, 117) = 93.27$, $MSE = 9.50$, $p < .001$, such that there were fewer noncritical intrusions within languages than across languages, $t(118) = -9.60$, $p < .001$. Furthermore, we found fewer noncritical intrusions when encoding and recall were in English than when encoding was in English and recall was in Spanish, $t(60) = -5.18$, $p < .001$. We found similar results when comparing the Spanish–Spanish and Spanish–English conditions, $t(57) = -9.45$, $p < .001$.

Veridical recognition

Veridical recognition was measured as the proportion of hits. Hits were computed by collapsing "old" responses of +1 (*somewhat unsure I heard it*), +2 (*fairly sure I heard it*), and +3 (*completely sure I heard it*) for words presented in the original list.

This and other measures derived from the recognition ratings must be interpreted cautiously because they may have been influenced by the preceding recall test. However, because being previously asked to free recall the word in English or Spanish did not influence the proportion of hits, as indicated by a nonsignificant main effect of recall, $F(1, 115) = 1.04$, $MSE = .01$, $p = .31$, and nonsignificant interactions with recall, $F(1, 115) < 2.14$, $p = .15$, the scores were collapsed over recall language.

Although neither of the main effects was significant, Fs $< .62$, ps $> .43$, there was a significant encoding × recognition interaction, $F(1, 115) = 12.83$, $MSE = .01$, $p = .001$. Just as in free recall, within-language recognition was more accurate than across-language recognition, $t(117) = 3.66$, $p < .001$. Also, recognition accuracy was higher for lists that were encoded and recognized in English than for those encoded in English but recognized in Spanish, $t(59) = 2.58$, $p = .01$. The same result was found when we compared the Spanish–Spanish with the Spanish–English condition, $t(56) = 2.50$, $p = .02$.

False recognition of critical nonpresented lures

False recognition was measured as the proportion of false alarms to critical nonpresented lures from presented lists. Critical false alarms were present in all combinations of encoding and recognition as follows. An encoding main effect indicated that the proportion of false recognition was higher with English than with Spanish encoding, $F(1, 115) = 6.00$, $MSE = .02$, $p = .02$. The influence of recognition language by itself was not reliable, $F(1, 115) = .00$, $MSE = .02$, $p = .96$, yet there was an encoding × recognition interaction, $F(1, 115) = 5.49$, $MSE = .03$, $p = .02$. Critical nonpresented lures appeared more frequently across languages than within languages, $t(117) = -2.17$, $p = .03$. However, there was not a significant difference in false recognition for lists that were encoded and recognized in English as compared with those encoded in English but recognized in Spanish, $t(59) = -1.55$, $p = .13$. The same result was found when we compared the Spanish–Spanish and Spanish–English conditions, $t(56) = -1.79$, $p = .08$.

Noncritical false alarms

"Old" responses to new items different from the critical lures showed a main effect of recognition language, where the proportion of noncritical intrusions in Spanish was significantly higher than in English, $F(1, 115) = 8.21$, $MSE = .02$, $p = .005$. Neither the main effect of encoding language nor the encoding × recognition interaction was significant, Fs $< .34$, ps $> .56$.

To test the possibility that the high rate of critical false alarms represented merely a high tendency to guess, being higher when languages shifted, the proportion of false alarms was analyzed in two different 2 (false alarm type) × 2 (encoding language) × 2 (retrieval language) mixed-factorial analyses of variance (ANOVAs). As expected, the proportion of false alarms to critical lures was much higher than the proportion of false alarms to noncritical new words, $F(1, 115) = 1{,}482.82$, $MSE = .02$, $p < .001$. The proportion of false alarms to critical lures of presented lists was also clearly higher than the proportion of false alarms to the critical lures of nonpresented lists, $F(1, 115) = 842.84$, $MSE = .03$, $p < .001$. Thus, although pure guessing was low ($M = .17$ and $M = .19$, respectively), the probability of falsely retrieving the critical lures due to the connectedness of the lists was high ($M = .80$). This pattern was unambiguous in all language combinations.

Recognition confidence

Participants' confidence that the item was or was not presented was indicated with a rating from −3 (*completely sure that I did not hear it*) to +3 (*completely sure that I heard it*). Mean (absolute) confidence ratings were computed for both veridical recognition and false recognition of critical nonpresented lures.

There were neither significant main effects nor interactions in confidence for hits or for critical false alarms, all Fs < 3.56, ps $> .06$. Only the encoding × recognition language interaction for noncritical false alarms confidence was significant, $F(1, 96) = 6.66$, $MSE = .44$, $p = .01$. A careful inspection of this interaction indicated that participants were mistakenly more confident of the correctness of their noncritical false alarms across languages than within languages, $t(98) = -2.16$, $p < .001$. Finally, there was a weak positive correlation between recognition confidence and hits, $r(117) = .27$, $p = .005$.

DISCUSSION

The present research examined veridical and false recall and recognition in Spanish–English bilinguals from a variety of countries by explicitly manipulating encoding and retrieval languages, discouraging shallow processing of the items based on language, and using materials that apply to a wider variety of Spanish–English bilinguals. One purpose was to investigate false recall and recognition could be replicated in Spanish. We found both false recall ($M = .49$) and false recognition ($M = .73$) in Spanish, thus extending the results obtained in Castilian Spanish by García-Bajos and Migueles (1997) ($M = .38$ false recall and $M = .70$ false recognition) and Perez-Mata et al. (2002, Experiment 2) ($M = .24$ false recall).

Although our direct manipulation of encoding and retrieval languages revealed false recall in all language combinations, false recall in Spanish, the native language of our participants, was less pronounced than false recall in English, the second but slightly more dominant language. In fact, the lowest false recall occurred when both encoding and recall were in Spanish.

The amount of false recognition was equivalent in both languages. These recognition results are in contrast to those of Sahlin et al. (2005), who found more false recognition in their English–English condition than their Spanish–Spanish condition. Although our study differs from that of Sahlin et al. in a number of ways (e.g., they used multiple encoding–recognition test trials, manipulated encoding language within participants, and instructed participants to recognize only the items presented in the same language during encoding), it appears that the results of both studies were influenced to some degree by small differences in participants' language proficiency. Apparently, our participants' slightly lower proficiency in Spanish decreased the amount of false recall but was not low enough to decrease false recognition.

Our false recognition results also differ from those of Anastasi, Rhodes, et al. (2005), who obtained higher false recognition in English than in Spanish with native Spanish speakers. There are several possible reasons for this discrepancy. First, Anastasi et al.'s lists were presented visually instead of auditorily. Second, as mentioned earlier, instead of translating the English lists or using items generated in Spanish for the same critical lures, they used word lists in Spanish that were different from the lists in English. Because of the high variability in the false recognition that different lists generate despite inclusion of the items with strongest association to the lures in all (Stadler et al., 1999), the particular lists used in their English condition might have been simply more successful in generating false recognition than those used in the Spanish condition. Third, encoding and recognition were manipulated within participants in their experiment (see also Sahlin et al., 2005), as opposed to between participants in ours. It is possible that our manipulation missed a tendency for higher accuracy in the more dominant language that their design allowed. Fourth, our participants were engaged in a recall task before recognition, but their participants answered an unrelated filler task after encoding in each language. This experience with the more dominant language on half of the lists could have smoothed out the differences. However, our data do not support this interpretation because recall language did not reliably interact with the other factors. Finally, although in both experiments the participant's native language was Spanish but the dominant language at the time of the experiments was English, all Anastasi, Rhodes, et al.'s (2005) participants reported Mexico as their country of origin, as opposed to a variety of Latin American countries in ours. Although proficiency indexes are unavailable in the Anastasi et al. article, it is possible that the gap in proficiency between English and Spanish was greater in Anastasi's participants than in ours, materializing in a reliable difference even in false recognition.

Support for this latter account comes from Miyaji-Kawasaki et al.'s (2003) study on false recognition in Japanese–English bilinguals. Participants had never lived in an English-speaking country and had a wide range of English dominance, including a low level of proficiency. Their results showed a higher proportion of false recognition in Japanese (the native and more dominant language) than in English (the second and less dominant language). It is possible that a markedly lower proficiency in English elicited markedly lower false recognition in the less dominant language.

A possible interpretation of our data is that the language terms had a different associative structure in English and Spanish. This might have resulted from our use of English norms for the English conditions and our translation of English norms into Spanish for the Spanish conditions. If so, it would have been preferable to build the Spanish lists from Spanish word association norms. This way, the list items would be the ones generating the intended critical lure. In fact, since the current research was completed, two false memory norms for Spanish have been compiled. Anastasi, De Leon, and Rhodes (2005) developed a set of 24 Spanish word lists, and Alonso, Fernández, Diez, and Beato (2004) developed 55 Spanish word lists based on Fernández, Diez, Alonso, and Beato's (2004) Spanish association norms. Unfortunately, the associative structure underlying such lists does not make them suitable for a population of Spanish–English bilingual speakers born in a wide variety of regions, like ours. Fernández et al.'s norms were obtained from monolingual college students from Spain who speak Castilian Spanish. Anastasi's norms were obtained from bilinguals born in Mexico, most of whom had no college background, whereas our population consisted of bilingual college students born in a variety of countries and attending college. There were no monolingual Spanish speakers and no bilinguals who spoke the Castilian variety of Spanish. Although syntax and morphology stay nearly constant across different versions of Spanish, vocabulary changes dramatically across regions. Therefore, it is unlikely that their lists possess the intended associative structure across participants from different backgrounds.

The second purpose of our study was to examine whether language congruence between encoding and retrieval is a necessary condition for the creation of false memories of nonpresented lures in Spanish–English bilinguals and, if not, whether false memories across languages are as strong as those generated within language. Results showed that false recall, false recognition, and false recognition confidence all increased when the language of encoding did not match the language of retrieval. This result stresses the relevance of encoding and retrieval contexts (Gallo, McDermott, Percer, & Roediger, 2001), in this case linguistics, as a factor in the creation of false memories.

Our false recognition data differ from those of Cabeza and Lennartson (2005), who found a comparable proportion of false recognition in English and French within languages and across languages, and Sahlin et al. (2005), who found greater false recognition for same language than for different language. A crucial difference is that in their experiments, participants were asked to provide a positive "old" response only when both the meaning and the language of the test item matched. Requiring a language match for positive recognition might have encouraged a shallow level of processing and increased the distinctiveness of the list items, in turn decreasing the frequency of false recognition of the critical nonpresented items. Alternatively, paying special attention to language might have helped reject critical nonpresented words that were sensed as familiar but were not remembered to be encoded in that language. In our experiment, with only a conceptual match between encoding and testing being required, the amount of false memories increased with language incongruence. The act of shifting languages might make bilinguals somewhat more prone not only to critical but also to noncritical intrusions. This was true in recall where both critical and noncritical intrusions increased in the cross-language condition but not in recognition.

Further research is needed to clarify the circumstances under which each of these three cross-language false recognition patterns is obtained. The instructions to recognize words as presented could be explicitly manipulated by asking half of the participants to recognize only words presented in the same language and the remaining half to recognize presented words regardless of the encoding language. Any possible language advantage could be controlled by constructing two full sets of DRM lists: one from Spanish norms and one from English norms. List sets should be balanced for factors that are known to influence false memory (e.g., backward associative strength), translated into the opposite language, and assessed across the different language conditions.[4]

Theoretical implications

Although the present study was not designed to test particular theories, the results stress the importance of considering language as a factor in formulating theories of false memory. Many of the results observed in our study can be explained using fuzzy trace theory (FTT) (Brainerd & Reyna, 1998, 2003; Reyna & Lloyd, 1997). According to FTT, verbatim (exact surface form) and gist (general meaning) representations of events are activated simultaneously and stored in parallel. In studies of false memory, false gist-consistent events often are confused with actual events and may cue verbatim traces, resulting in false memories. However, according to Brainerd and Reyna (2002, 2003; see also Brainerd, Reyna, Wright, & Mojardin, 2003), people may be able to use an editing operation known as recollection rejection to uncover mismatches in actual versus gist-based verbatim traces. If the verbatim information is accessible, this operation can override the familiarity that gist-based traces produce, thereby diminishing false memory. Sahlin et al. (2005) demonstrated that when participants are explicitly instructed to use the encoding language as a cue over multiple study–test trials, participants used this heuristic to reduce false recognition.

Applying FTT to our study, performance on immediate free recall tests is dominated by accessibility of verbatim traces and the simple readout of surface information that is represented there. Although gist memory makes contributions, recall of list items is influenced disproportionately by factors that impair initial storage of verbatim traces (such as language). Thus, verbatim traces would be stronger when encoding and test language match because those memory traces are more readily available. Furthermore, adult models of bilingual memory (e.g., Kroll & Stewart, 1994) seem to suggest that greater reliance on verbatim storage in the participant's more dominant language produces more accurate veridical recall in the English–English than in the Spanish–Spanish condition. In addition, if the storage of verbatim traces in the cross-language condition is less reliable, then those traces cannot be used equally effectively to help reject critical lures related to the list items. This results in a greater reliance on gist information and greater false recall (see Fazendeiro, Winkielman, Luo, & Lorah, 2005).[5]

FTT cannot equally be applied to explain the results observed in veridical and false recognition in the present study. Unlike in recall, it is not necessary for participants to reconstruct the verbatim memory traces, and therefore performance when the encoding and test languages match is expected to be equivalent. In addition, because the verbatim traces are equally accessible in recognition, false memory is expected to be reduced by the process of recollection rejection when the encoding and test languages do not match.

The source monitoring framework (SMF) is an alternative explanation of the false memory phenomenon. According to this view, thoughts, images, and feelings that are experienced as memories are attributed rapidly to a source, with little conscious thought (Lindsay & Johnson, 2000). The SMF attributes false memory to an inability to indicate whether activated associates appeared in the study list. Furthermore, according to this view, different sources are more likely to result in memory errors if the sources match than if they mismatch. Applying this framework to our data, SMF would predict greater veridical and false memory when the encoding and test language match than when they mismatch. However, the present study found that whereas veridical memory was greater when encoding and test language matched, false memory was greater when they mismatched. Thus, although it does not appear that SMF can account for our results, Lindsay and Johnson did note that there may be a trade-off between source similarity and other factors, which may affect the likelihood of memory errors due to source misattributions. If that is true, then further research is needed to examine whether language produces such a trade-off.

NOTES

This research was partially supported by a professional improvement grant and a summer research stipend awarded to Gloria Marmolejo by Winona State University.

Portions of the free recall data were reported at the biennial meeting of the Society for Applied Research in Memory and Cognition in Aberdeen, Scotland, July 2003. Recognition results were reported at the annual meeting of the Psychonomic Society in Vancouver, Canada, November 2003.

We thank Cynthia Armstrong, Vanessa Currier, Matthew Ellis, Jennifer French, Debbie Fuehrer, Leigh Oglesby, and the late Mariella Griffiths for assisting with data coding and proofreading. We are also grateful to Bennett Schwartz for making possible the recruitment of the bilingual participants; to Ramón M. Riba for Spanish linguistic consulting, computer programming, and list voiceover; to the 10 judges who reviewed the Spanish lists; and to the participants in the pilot study conducted in Rochester. Finally, we would like to thank Jason Arndt, Nancy Franklin, Henry L. Roediger III, C. J. Brainerd, and an anonymous reviewer for their helpful comments and suggestions on a previous version of this manuscript.

Address correspondence about this article to Gloria Marmolejo, Department of Psychology, Winona State University at Rochester, 859 Thirtieth Avenue SE & Highway 14 E, Rochester, MN 55904 (e-mail: gmarmolejo@winona.edu).
Received for publication January 30, 2007; revision received August 8, 2007. Action Editor: Donelson E. Dulany.

APPENDIX. Word lists

ANGER	ENOJO	BREAD	PAN	CHAIR	SILLA
mad	furioso	butter	mantequilla	table	mesa
fear	miedo	food	alimento	sit	sentarse
hate	odio	eat	comer	legs	patas
rage	rabia	rye	centeno	seat	asiento
temper	genio	milk	leche	desk	escritorio
fury	furia	flour	harina	wood	madera
wrath	cólera	jelly	mermelada	cushion	cojín
happy	contento	dough	masa	swivel	girar
fight	pelea	crust	corteza	stool	banquito
mean	malo	slice	rebanada	rocker	mecedora
calm	calmado	wine	vino	bench	banca
enrage	enfurecerse	loaf	hogaza	relax	relajarse

CITY	CIUDAD	COLD	FRIO	CUP	TAZA
town	pueblo	hot	caliente	mug	tarro
state	estado	snow	nieve	saucer	plato
streets	calles	warm	tibio	measuring	medir
country	país	winter	invierno	coaster	posavasos
New York	Nueva York	ice	hielo	lid	tapa
village	aldea	wet	húmedo	handle	asa
big	grande	heat	calor	coffee	café
suburb	afueras	weather	clima	goblet	copa
county	municipio	freeze	congelar	soup	sopa
people	gente	shiver	tiritar	stain	mancha
building	edificio	frost	escarcha	drink	bebida
noise	ruido	dark	obscuro	sip	sorbo

DANCE[a]	BAILE (BAILAR)	LOVE[a]	AMOR (AMAR)	MOUNTAIN	MONTAÑA
party	fiesta	affection	afecto	hill	loma
fun	diversión	kiss	beso	climb	escalar
joy	alegría	pain	dolor	top	cima
waltz	vals	life	vida	peak	pico
discoteque	discoteca	friendship	amistad	plain	plano
movement	movimiento	everything	todo	goat	chivo
shoe	zapato	happiness	felicidad	bike	bicicleta
step	paso	feeling	sentimiento	climber	alpinista
partner	pareja	heart	corazón	range	cordillera
jump	saltar	tenderness	ternura	ski	esquiar
song	canción	pleasure	placer	cave	cueva
costume	disfraz	desire	deseo	rock	piedra

(cont.)

APPENDIX. Word lists (cont.)

NEEDLE	AGUJA	RIVER	RIO	SLEEP	DORMIR
thread	hilo	water	agua	bed	cama
pin	alfiler	stream	corriente	rest	descansar
sewing	costura	lake	lago	awake	despierto
sharp	punzante	boat	bote	tired	cansado
point	punta	tide	marea	dream	soñar
prick	pinchazo	swim	nadir	wake	despertar
thimble	dedal	run	correr	snore	ronquido
thorn	espina	creek	riachuelo	nap	siesta
hurt	lastimar	fish	pez	peace	paz
syringe	jeringa	bridge	puente	yawn	bostezar
cloth	tela	winding	tortuoso	drowsy	soñoliento
knitting	tejido	deep	profundo	night	noche

SLOW	DESPACIO	SOFT	BLANDO	TIME[a]	TIEMPO
fast	rápido	hard	duro	hour	hora
stop	detener	light	ligero	clock	reloj
listless	apático	pillow	almohada	years	años
snail	caracol	plush	afelpado	past	pasado
cautious	cauteloso	cotton	algodón	short	corto
delay	retraso	touch	tocar	age	edad
turtle	tortuga	fluff	pelusa	space	espacio
hesitant	indeciso	furry	peludo	eternal	eterno
speed	velocidad	kitten	gatito	epoque	época
wait	esperar	skin	piel	eternity	eternidad
move	moverse	tender	tierno	century	siglo
lazy	perezoso	silk	seda	second	segundo

[a]*These lists were initially built in Spanish using Spanish association norms (Algarabel et al., 1986) and then translated into English.*

1. The first challenge occurs in languages that share common origins and therefore contain cognates and foreign words. This makes entire lists, such as that of *music,* unsuitable because of the presence of several identical cognates (e.g., *piano* in English is translated as *piano* in Spanish). In contrast, the second challenge is the noteworthy vocabulary variations in languages that are spoken in many different countries. For instance, whereas the item *cake* in the *sweet* list is translated as *tarta* in Castilian Spanish (Spain), the correct translation is *pastel* in Mexico, *torta* in Argentina, *ponqué* in Colombia, and *bizcocho* and *quequé* in other Latin American countries. Unless such words are identified and avoided in research with heterogeneous populations, participants are unfamiliar with their meaning, thus underestimating the proportion of veridical and false memories.

A third challenge is that in English, the same word can be used as a noun, a verb, or an adjective, but this does not happen in Spanish. Thus, in DRM lists such as that of *anger,* it is not clear what lexical category is the best associate and the best translation equivalent for the item *hate:* the noun *odio* or the verb *odiar.* Furthermore, although in English nouns, adjectives, and verbs are often compounded (e.g., *landfill*), in Spanish prepositional phrases are often used instead (e.g., *vertedero de basura*). This often results in Spanish lists with different composition properties and item lengths. Because word length exerts a significant effect on veridical recall (Glanzer & Razel, 1974) and critical lure length influences false memory, this creates a confound when false memories are compared in different languages.

The final aspect to address is adjective gender. Whereas English adjectives and common nouns are neutral, Spanish common nouns may be feminine, masculine, or neutral, and adjectives agree in gender and number with the noun they qualify (Steigerwald, 1987). Therefore, in DRM lists such as that of *spider,* where the translation is the feminine noun *araña,* it is not clear which gender should be selected for the translation of adjective list items such as *small*—the feminine *pequeña* or the masculine *pequeño*—and whether this choice influences the rate of false reporting of the critical lure.

2. There were three exceptions (1.6% of items), where cognates differed in only one character, all from the *music* list. Cognates differing by more than one letter were not replaced because that would have entailed eliminating many more items, possibly decreasing the associative strength of the lists and thus decreasing their capability to generate false memories. It is also important to add that Anastasi, De Leon, et al.'s (2005) and Sahlin et al.'s (2005) word lists contained several cognates and noncognate homophones. Although it is uncertain whether the inclusion of these types of items affected their results, it remains possible that memorability of certain list items and nonpresented critical lures could have been differentially affected by the presence of a cognate or homophone.

3. A separate experiment was conducted to ensure that the phrase "now recall" presented at the end of each list in false recall did not alter the magnitude of false recall. In the veridical recall literature, the inclusion of a single irrelevant item or recall instruction such as "recall" or "now" at the end of a verbal list has been found to reduce the recallability of the last few items of the list even when the instruction itself does not have to be recalled (the suffix effect; see Roediger & Crowder, 1976). The test was conducted with 54 monolingual native English undergraduate students from Berry College, Mount Berry, Georgia. The materials and procedure were essentially the same as those used in Experiment 1 with two exceptions. All stages were conducted in English, and participants heard one of two possible signals at the end of each list to indicate the start of the free recall period: either the expression "now recall" followed by a ring or a knock, or just the ring or the knock alone. Recall accuracy with a signal was significantly lower with than without a signal, $F(1, 48) = 9.47$, $MSE = .01$, $p = .003$. This was due to a less pronounced recency effect when the "now recall" signal was presented at the end of the list. In contrast, the signal manipulation left the amount of false recall intact, $F(1, 48) = 1.92$, $MSE = .04$, $p = .17$. In fact, the proportion of false recall was comparable to that obtained in past research and was very close to that of items presented in the middle of the list.

4. We would like to thank H. L. Roediger for this useful suggestion to eliminate disadvantaged recall in Spanish while not introducing a confound by using different list sets in English and Spanish.

5. We thank C. J. Brainerd for his helpful suggestions with respect to the implications of our results for fuzzy trace theory and the source monitoring framework.

REFERENCES

Albuquerque, P. B., & Pimentel, E. (2005). Recency effect inhibition and the false memory production: A study with Portuguese word lists. *Psicologia, Educação e Cultura, 9,* 69–87.

Algarabel, S., Ruiz, J. C., & Sanmartín, J. (1988). The University of Valencia's computerized word pool. *Behavior Research Methods, Instruments, & Computers, 20,* 398–403.

Algarabel, S., Sanmartín, J., García, J., & Espert, R. (1986). *Normas de asociación libre para investigación experimental: Informes del Laboratorio de Aprendizaje y Memoria* [Free association norms for experimental research: Learning and Memory Laboratory report], No. 1. Universidad de Valencia. Unpublished manuscript.

Alonso, M. A., Fernández, A., Diez, E., & Beato, M. S. (2004). Indices de produccion de falso recuerdo y falso reconocimiento para 55 listas de palabras en castellano [False recall and false recognition norms for 55 word lists in Castilian]. *Psicothema, 16*(3), 357–362.

Altarriba, J. (1992). The representation of translation equivalents in bilingual memory. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 157–174). New York: Elsevier.

Altarriba, J. (2000). Language processing and memory retrieval in Spanish-English bilinguals. *Spanish Applied Linguistics, 4,* 215–245.

Altarriba, J. (2003). Does *cariño* equal "liking"? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism, 7,* 305–322.

Altarriba, J., & Santiago-Rivera, A. L. (1994). Current perspectives on using linguistic and cultural factors in counseling the Hispanic client. *Professional Psychology: Research and Practice, 25,* 388–397.

Anastasi, J. S., De Leon, A., & Rhodes, M. G. (2005). Normative data for semantically associated Spanish word lists that create false memories. *Behavior Research Methods, 37,* 631–637.

Anastasi, J. S., Rhodes, M. G., Marquez, S., & Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory, 13,* 815–828.

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science, 9,* 484–489.

Brainerd, C. J., & Reyna, V. F. (2002). Recollection rejection: How children edit their false memories. *Developmental Psychology, 38,* 156–172.

Brainerd, C. J., & Reyna, V. F. (2003). Recollection rejection: False memory editing in children and adults. *Psychological Review, 110,* 762–784.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110,* 762–784.

Cabeza, R., & Lennartson, E. R. (2005). False memory across languages: Illusory recognition in English-French bilinguals. *Memory, 13,* 1–5.

Campos, L., & Alonso-Quecuty, M. L. (1998). Knowledge of the crime context: Improving the understanding of why the cognitive interview works. *Memory, 6,* 103–112.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Fazendeiro, T., Winkielman, P., Luo, C., & Lorah, C. (2005). False recognition across meaning, language, and stimulus format: Conceptual relatedness and the feeling of familiarity. *Memory & Cognition, 33*(2), 249–260.

Fernández, A., Diez, E., Alonso, M. A., & Beato, M. S. (2004). Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behavior Research, Methods, Instruments, & Computers, 36*(3), 577–583.

Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of the actual victims and witnesses of crime. *Journal of Applied Psychology, 74,* 722–727.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. III. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 27(2), 339–353.

García-Bajos, E., & Migueles, M. (1997). Falsas memorias en el recuerdo y reconocimiento de palabras [False memories in the recall and recognition of words]. *Estudios de Psicología, 58,* 3–14.

Glanzer, M., & Razel, M. (1974). The size of the unit of short-term storage. *Journal of Verbal Learning and Verbal Behavior, 13,* 114–131.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36,* 3–15.

Holliday, R., Kang, S., & Lee, K. (2003, November). *False memory effects on spoken word recognition among bilingual speakers.* Paper presented at the meeting of the Psychonomic Society, Vancouver, Canada.

Kawasaki, Y., & Yama, H. (2006). The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory, 14*(1), 68–78.

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

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1998). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Directions in Psychological Science, 16,* 181–224.

Lindsay, D. S., & Johnson, M. K. (2000). False memories and the source monitoring framework: A reply to Reyna and Lloyd (1997). *Learning and Individual Differences, 12,* 145–161.

Miyaji-Kawasaki, Y., Inoue, T., & Yama, H. (2003). Cross-linguistic false recognition: How do Japanese-dominant bilinguals process two languages: Japanese and English? *Psychologia, 46*(4), 255–267.

Molina, A. I., Su, I. F., Knight, C., Holliday, R., & Weekes, B. S. (2006). Bilingual false recollection: An EEG study. *Brain and Language, 99,* 180–182.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35,* 261–285.

Pérez-Mata, M. N., Read, J. D., & Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory, 10*(3), 161–177.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9,* 95–123.

Robinson, K. J., & Roediger, H. L. III. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*(3), 231–237.

Roediger, H. L., & Crowder, R. G. (1976). Recall instructions and the suffix effect. *American Journal of Psychology, 89,* 115–125.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Rubin, D. C. (Ed.). (1996). *Remembering our past: Studies in autobiographical memory.* New York: Cambridge University Press.

Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test.* Minneapolis: University of Minnesota.

Sahlin, B. H., Harding, M. G., & Seamon, J. G. (2005). When do false memories cross language boundaries in Spanish-English bilinguals? *Memory & Cognition, 33,* 1414–1421.

Santiago-Rivera, A. L., & Altarriba, J. A. (2002). The role of language in therapy with the Spanish-English bilingual client. *Professional Psychology: Research and Practice, 33,* 30–38.

Shaw, J. S. III, García, L. A., & Robles, B. E. (1997). Cross-language postevent misinformation effects in Spanish-English bilingual witnesses. *Journal of Applied Psychology, 82*(6), 889–899.

Stadler, M., Roediger, H. L. III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494–500.

Stavenhagen, R. (1990). The ethnic question: Conflicts, development and human rights. In G. Valdés, *Multilingualism.* Retrieved April 30, 2002, from http://www.lsadc.org/fields/index.php?aaa=multiling.htm

Steigerwald, J. (1987). *Exploring French, German, and Spanish.* Winona, MN: WSU Campus Press.

Wright, D. B., Loftus, E. F., & Hall, M. (2001). Now you see it, now you don't: Inhibiting recall and recognition of scenes. *Applied Cognitive Psychology, 15,* 471–482.