Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: http://www.tandfonline.com/loi/pmem20

The effects of bilingual language proficiency on recall accuracy and semantic clustering in free recall output: evidence for shared semantic associations across languages

Wendy S. Francis, Randolph S. Taylor, Marisela Gutiérrez, Mary K. Liaño, Diana G. Manzanera & Renee M. Penalver

To cite this article: Wendy S. Francis, Randolph S. Taylor, Marisela Gutiérrez, Mary K. Liaño, Diana G. Manzanera & Renee M. Penalver (2018) The effects of bilingual language proficiency on recall accuracy and semantic clustering in free recall output: evidence for shared semantic associations across languages, Memory, 26:10, 1364-1378, DOI: 10.1080/09658211.2018.1476551

To link to this article: https://doi.org/10.1080/09658211.2018.1476551

Published online: 19 May 2018.

MEMORY
2018, VOL. 26, NO. 10, 1364–1378
https://doi.org/10.1080/09658211.2018.1476551

Routledge
Taylor & Francis Group

Check for updates

The effects of bilingual language proficiency on recall accuracy and semantic clustering in free recall output: evidence for shared semantic associations across languages

Wendy S. Francis, Randolph S. Taylor, Marisela Gutiérrez, Mary K. Liaño, Diana G. Manzanera and Renee M. Penalver

Department of Psychology, University of Texas at El Paso, El Paso, TX, USA

ABSTRACT
Two experiments investigated how well bilinguals utilise long-standing semantic associations to encode and retrieve semantic clusters in verbal episodic memory. In Experiment 1, Spanish-English bilinguals ($N = 128$) studied and recalled word and picture sets. Word recall was equivalent in L1 and L2, picture recall was better in L1 than in L2, and the picture superiority effect was stronger in L1 than in L2. Semantic clustering in word and picture recall was equivalent in L1 and L2. In Experiment 2, Spanish-English bilinguals ($N = 128$) and English-speaking monolinguals ($N = 128$) studied and recalled word sequences that contained semantically related pairs. Data were analyzed using a multinomial processing tree approach, the pair-clustering model. Cluster formation was more likely for semantically organised than for randomly ordered word sequences. Probabilities of cluster formation, cluster retrieval, and retrieval of unclustered items did not differ across languages or language groups. Language proficiency has little if any impact on the utilisation of long-standing semantic associations, which are language-general.

ARTICLE HISTORY
Received 17 February 2018
Accepted 8 May 2018

KEYWORDS
Free recall; semantic clustering; bilingualism

Free recall output tends to be organised by semantic category, thus exhibiting *semantic clustering* (Bousfield, 1953). According to the *pair-clustering model* (Batchelder & Riefer, 1980; Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002), semantic clusters are formed at encoding and retrieved as units at test. According to the *context maintenance and retrieval model* (Polyn, Norman, & Kahana, 2009), semantic clustering in recall output derives from the long-standing associative relations among the items. Each successful attempt to retrieve an item from episodic memory increases the activation of semantically associated and temporally contiguous items, which makes such items more likely to be retrieved next. Thus, the pair-clustering model focuses more on processes that occur at encoding, and the context-maintenance-and-retrieval model focuses more on processes that occur at retrieval.

It is unknown whether bilingual language proficiency affects the utilisation of semantically organised encoding or retrieval strategies. In the present study we investigated the extent to which bilinguals use semantically organised encoding and retrieval strategies in their more proficient language (L1) and their less proficient language (L2) by comparing semantic clustering in free recall output for bilingual L1, bilingual L2, and monolingual L1. Note that our use of L1 and L2 refers to the dominance or relative proficiency of the two languages, which is not necessarily the order of acquisition. We also investigated how the planned retrieval language impacts picture encoding and whether bilingual language proficiency alters the benefit derived from semantic organisation in a study sequence.

Recall and semantic clustering in single-language studies

Recall performance and the degree of semantic clustering in recall output depend on characteristics of the words to be learned, the conditions under which the words are studied, the presentation modality, and the organisation of the study sequence. Word frequency is associated with both free recall performance and semantic clustering. Specifically, high-frequency words are better recalled than low-frequency words when learned in frequency-pure lists (e.g., Balota & Neely, 1980; Kinsbourne & George, 1974). Nevertheless, semantic clustering in recall output is stronger for low-frequency words than for high-frequency words when taxonomic frequency or association strength is held constant (Jordan & Swartz, 1976; Matthews, 1966).

Dividing attention by imposing the cognitive load of a secondary task at study or test impairs free recall performance (Craik, Naveh-Benjamin, Ishaik, & Anderson, 2000; Naveh-Benjamin, Craik, Perretta, & Tonev, 2000; Whiting,

CONTACT Wendy S. Francis wfrancis@utep.edu Department of Psychology, University of Texas at El Paso, 500 W. University Ave., El Paso, TX 79968, USA

2003) and reduces the degree of semantic clustering in recall output (Park, Smith, Dudley, & Lafronza, 1989). One explanation is that taking cognitive resources away from the encoding task makes it less likely that items will be encoded in an associative or semantic manner. Processing of associations among items elicits stronger clustering than semantic processing of individual items (Hunt & Einstein, 1981). Another possibility is that taking cognitive resources away from the retrieval task leads to a less organised retrieval strategy, which in turn impairs recall performance. That is, the poor recall performance and the reduction in clustering are both indicators of a disrupted encoding or retrieval strategy.

Pictures of objects are better recalled than the words that name them (e.g., Davies, Milne, & Glennie, 1973; Paivio & Csapo, 1973; Paivio, Rogers, & Smythe, 1968). Recall of pictures improves when they are named either overtly or covertly at encoding (Davies et al., 1973; Horowitz, 1969), presumably because accessing the name provides two routes for later retrieval, a verbal route and an image route (Paivio et al., 1968). Previous research has shown that picture recall exhibits stronger semantic clustering than word recall (Cole, Frankel, & Sharp, 1971; Horowitz, 1969). One explanation for this effect is that pictures elicit faster access to conceptual information than words, as evidenced by the finding that pictures are semantically categorised faster than words (e.g., Durso & Johnson, 1980; Potter & Faulconer, 1975). Note also that accessing a picture's name requires access to the concept, whereas reading a word does not (e.g., Potter & Faulconer, 1975). Therefore, naming pictures overtly or covertly at encoding will elicit access to the corresponding concepts, which will make pictures both more memorable and more likely to be clustered than words.

Having an organised study sequence improves recall performance (e.g., Bower, Clark, Lesgold, & Winzenz, 1969; Cofer, 1966; Hunt & Einstein, 1981; Nott & Lambert, 1968), as does having an organised retrieval strategy (e.g., Keniston & Flavell, 1979; Santa, Ruskin, Snuttjer, & Baker, 1975). Specifically, when study sequences were blocked by semantic category, free recall performance was better and semantic clustering was stronger than when the same words were randomly sequenced at study (e.g., Cofer, 1966).

Recall and semantic clustering in bilingual research

Bilingual free recall performance was worse for words in the non-dominant language (L2) than for words in the dominant language (L1) in several early studies where semantic category membership of the word stimuli was not controlled (e.g., Durgunoglu & Roediger, 1987; Glanzer & Duarte, 1971; López & Young, 1974). Note that we consider the dominant language (L1) to be the language in which an individual bilingual has relatively higher proficiency, which does not necessarily reflect the order of acquisition. When the words in a studied list came from a limited number of semantic categories but were randomly sequenced, free recall performance was better in L1 than in L2 (Champagnol, 1973; Harris, Cullum, & Puente, 1995; Lambert, Ignatow, & Krauthamer, 1968; Nott & Lambert, 1968), but when each word came from a different semantic category, the L1 advantage was eliminated (Lambert et al., 1968; Nott & Lambert, 1968). Overall, word lists with semantic categories were recalled much better than lists with no categories (Lambert et al., 1968; Nott & Lambert, 1968), so it appears that bilinguals were better able to utilise the implicit semantic categories in L1 than in L2. When study sequences were blocked by semantic category, bilingual free recall performance was better than when the same words were randomly sequenced at study (Nott & Lambert, 1968). The advantage of blocked over random sequences was equivalent for L1 and L2 (Nott & Lambert, 1968).

Several studies have compared monolingual and bilingual free recall performance. One study showed equivalent English free recall performance for monolinguals and balanced bilinguals, but better performance for monolinguals than for Spanish-dominant bilinguals completing the task in L2 (Harris et al., 1995). Another showed equivalent free recall for monolinguals and highly proficient bilinguals performing the task in L1 and L2 (Francis & Baca, 2014). In contrast, when children free recalled food items presented in a story context, balanced bilinguals performed *better* than monolinguals (Haritos, 2002). In free recall of short imperative sentences like "read the book," bilingual children performed at an equivalent level or higher level relative to monolingual children depending on the scoring system used (Kormi-Nouri, Moniri, & Nilsson, 2003). It is difficult to know what to make of these discrepant results across studies, because it is not clear whether monolingual and bilingual groups were comparable on non-linguistic factors that might impact memory performance. Two other studies did include measures of such factors. In one study, bilingual younger and older adults performed worse than monolingual adults (matched on age and non-verbal intelligence scores) in free recall of English word lists in which all words came from the same semantic category (Fernandes, Craik, Bialystok, & Kreuger, 2007). In the other, highly proficient bilingual adults performed worse in L2 but not in L1 relative to monolinguals who were matched on age, non-verbal cognitive ability, and socio-economic status in free recall of word lists with no categorical structure (Francis, Liaño, & Taylor, 2018).

The effect of language proficiency on semantic clustering in free recall output has been investigated in a small number of studies, but the results are mixed. In one study, semantic clustering was equivalent for L1 and L2 (Lambert et al., 1968), but in two other studies, semantic clustering was greater for L1 than for L2 (Champagnol, 1973; Nott & Lambert, 1968). Not surprisingly, balanced bilinguals showed equivalent clustering in L1 and L2 (Harris et al., 1995). In the one study that compared monolingual and bilingual semantic clustering, balanced bilinguals exhibited stronger semantic clustering than monolinguals and unbalanced bilinguals, but monolinguals and unbalanced bilinguals did not differ (Harris et al., 1995).

The preceding review of the small literature on bilingual free recall and semantic clustering shows mixed results. It cannot be determined whether the effects of language that were found were caused by proficiency differences or were merely an artifact of language differences, because with two exceptions, all bilingual participants had the same dominant language. One exception (Nott & Lambert, 1968) included only 9 French-dominant and 9 English-dominant participants, which is a small sample from which to draw generalisations. The other (Francis et al., 2018) used lists with no categorical structure, so clustering could not be measured. Thus, previous research on how recall performance and semantic clustering compare for L1 and L2 or for monolinguals and bilinguals has not produced definitive answers.

Conceptualizations of bilingual proficiency and memory

Bilingual memory can be conceptualised in at least three ways. L2 memory may be thought of as having lower item familiarity, as making greater demands on cognitive resources, or as connected to a system of semantic associations that is shared with L1. These conceptualizations lead to different predictions about bilingual free recall and semantic clustering performance.

First, L2 words are less familiar and typically have occurred less often in a person's lifetime than L1 words, and may therefore function as if they were lower frequency words in L1 (Ardila, 2003; Gollan, Montoya, Cera, & Sandoval, 2008). The lower level of recall performance for low-frequency words relative to high-frequency words leads to the prediction that bilingual participants will have lower item recall in L2 than in L1. The higher degree of clustering observed for low-frequency words relative to high-frequency words in monolingual participants leads to the prediction that clustering in bilingual participants will be stronger in L2 than in L1. By a similar logic, bilinguals are exposed to each word in their vocabulary less often than age-matched monolinguals, because the experience is divided between two languages. Therefore, under this conceptualisation, bilinguals would be expected to have lower item recall in both of their languages than monolinguals and stronger clustering in both of their languages than monolinguals.

This perspective can be more theoretical if we consider L2 words to be more weakly associated than L1 words with the concepts that they represent in semantic memory. A *frequency-lag hypothesis* has been proposed as a single mechanism to account for monolingual/bilingual differences, bilingual proficiency effects, and word-frequency effects in word production (Gollan et al., 2008). This hypothesis is based on two principles. First, the strength of associations between words and their concepts depends on experience, or the number of prior exposures, with the consequence being that speed (and accuracy) of lexical access increases with experience. Second, the learning function for lexical access is negatively accelerated, meaning that as the number of exposures increases, the beneficial effect of an additional exposure decreases.

A second conceptualisation of bilingual memory is based on the idea that processing information in L2 requires more attention, or makes a greater demand on cognitive resources (Abu-Rabia, 2003; Ransdell, Arecco, & Levy, 2001; Takano & Noda, 1993), and reduces the amount of information that can be held in working memory (e.g., da Costa Pinto, 1991; Service, Simola, Metsaenheimo, & Maury, 2002). Thus, attempting to encode information in L2 may be analogous to trying to encode information under a cognitive load. The impairment in recall performance following encoding under conditions of divided attention leads to the prediction that bilingual participants will have lower item recall in L2 than in L1. The reductions in clustering after encoding under conditions of divided attention lead to the prediction that clustering will be weaker in L2 than in L1. Similarly, we would expect that processing language stimuli would make greater demands on cognitive resources for bilinguals than for monolinguals, which would lead to lower item recall and weaker clustering for bilinguals relative to monolinguals.

A third perspective arises from evidence that translation equivalents share conceptual representations and that the two languages of a bilingual access a common amodal semantic or conceptual system (for reviews, see Francis, 1999, 2005). Most relevant to reasoning about long-standing semantic associations among category exemplars, studies of conceptual repetition priming between languages have shown evidence that three types of semantic associations (i.e., category-exemplar, noun-verb action, and antonym relationships) have shared representations across languages in bilinguals (de la Riva López, Francis, & García, 2012; Francis, Fernandez, & Bjork, 2010; Seger, Rabin, Desmond, & Gabrieli, 1999; Taylor & Francis, 2017). If semantic associations are in fact language-general, that is, independent of any particular language, then there should be no effects of language on their utilisation and therefore no difference between L1 and L2 in the degree of semantic clustering in recall output. Similarly, the semantic associations formed in the bilingual mind should be the same as those formed in the monolingual mind, so bilingual and monolingual speakers would be expected to show equivalent degrees of semantic clustering in recall output. Note, however, that this perspective does not necessarily lead to a prediction of equivalent item recall across languages or language groups, because item recall also depends on non-conceptual factors that are by their nature language-specific, and retrieval may therefore be more difficult in L2 than in L1 and more difficult for bilingual than for monolingual speakers.

The present study

The main goal of the present study was to better understand the impact of bilingual proficiency on explicit memory by comparing the degree of semantic clustering in free recall output across languages in bilinguals and across bilingual and monolingual samples matched on age, education, and socio-economic status. We addressed four primary questions. First, what conceptualisation of bilingual memory best explains free recall accuracy and semantic clustering in bilinguals? This question was addressed in Experiments 1 and 2 by comparing free recall accuracy and semantic clustering for bilingual L1 and L2 word lists and for monolingual and bilingual participants. Second, in free recall of pictures, what is the role of name access in recall accuracy and semantic clustering? This question was addressed in Experiment 1 by comparing free recall accuracy and semantic clustering for picture sets to be recalled in L1 or L2. Third, how do the benefits of semantic organisation at study compare for bilingual L1 and L2 words and for monolingual and bilingual speakers? This question was addressed in Experiment 2 by comparing free recall accuracy and semantic clustering for semantically organised and random word lists.

Experiment 1 examined free recall of word and picture sets and semantic clustering in recall output in Spanish-English bilinguals, using the traditional method of including a limited number of categories in each studied item set and measuring the tendency to recall items from the same category consecutively at test. Experiment 2 examined free recall of semantically organised and randomly ordered word lists in Spanish-English bilinguals and English-speaking monolinguals, using the pair-clustering method of including related pairs of words in the study sequence, and measuring the tendency to recall items from the same pair consecutively at test. Experiment 2 also improved upon the methodology of Experiment 1 by using objective standardised assessments of language proficiency to determine language dominance.

Experiment 1

In Experiment 1, Spanish-English bilingual participants learned four sets of 24 words or pictures in each language. Each set included items from four semantic categories, and the categories were not repeated across lists. After study, participants performed an immediate free recall test. We expected word recall rates to be lower in L2 than in L1, as seen in previous studies with this type of list composition (Champagnol, 1973; Harris et al., 1995; Lambert et al., 1968; Nott & Lambert, 1968). Such a result would be consistent with all three conceptualizations of L2 memory. We predicted that picture recall rates would also be lower in L2 than in L1, because in bilinguals, the names of pictures are more likely to be accessed in L1 than in L2, making them more likely to have two retrieval routes in L1 than in L2. However, an early study comparing bilingual picture recall across languages was inconclusive (Ervin, 1961). The greater likelihood of name access at encoding in L1 than in L2 should also lead to a greater advantage for pictures over words, or a greater picture-superiority effect, in L1 than in L2.

The three conceptualizations of L2 memory lead to different predictions about semantic clustering in free recall of word sets in L1 and L2. The frequency-lag hypothesis leads to the prediction that clustering will be stronger in L2 than in L1. In contrast, consideration of the additional cognitive load imposed by having to work in L2 leads to the prediction of weaker clustering in L2 relative to L1, as found in two previous studies (Champagnol, 1973; Nott & Lambert, 1968). The idea that semantic associations are language general and their utilisation for explicit memory purposes does not differ across languages leads to the prediction of equivalent clustering for L2 and L1, as seen in one previous study (Lambert et al., 1968).

Semantic clustering in picture recall was expected to be stronger than in word recall as in previous research (Cole et al., 1971; Horowitz, 1969). The new question was whether picture recall would differ depending on the language of recall. If semantic clustering depends on name access, then clustering in picture recall will be weaker in L2 than in L1, because name access is more likely in L1. However, previous monolingual studies showed that labelling pictures overtly or covertly did not increase clustering (Davies et al., 1973; Horowitz, 1969), so it seems unlikely that clustering in bilinguals would be affected by name access. If semantic clustering instead depends only on conceptual access from the picture, then semantic clustering will be equivalent for L1 and L2 picture recall, because conceptual access from a picture does not depend on the language of covert naming.

Method

Power and sample size analysis

A power analysis showed that 34 participants would be sufficient to detect a medium sized effect ($f = .25$) with 80% power, and 90 participants would be sufficient to detect a small-to-medium sized effect ($f = .15$) with 80% power. Because the counterbalancing and alternate item orders in Experiment 1 required a multiple of 64 participants, we tested 128 participants. This sample size would yield over 99% power to detect a medium sized ($f = .25$) recall difference between L1 and L2 and over 92% power to detect a small-to-medium sized ($f = .15$) recall difference between L1 and L2.

Participants

Participants were 128 self-reported Spanish-English bilingual students (73 women, 55 men) at the University of Texas at El Paso who participated as part of a research requirement for introductory psychology classes. Participant characteristics are summarised in Table 1. Language dominance was determined by self report using a measure that correlates with asymmetries in productive vocabulary (Francis, Regalado, Sáenz, & Durán, 2006). First, participants were asked to indicate their stronger language overall; the language indicated was considered the dominant language, but if they responded both or a mixture, then rating scale information was used. Participants rated their relative proficiency in English and Spanish on a 7-point Likert-type scale from 1 (strong advantage for English) to 7 (strong advantage for Spanish) for each of 8 aspects of language (speaking, listening, writing, reading, vocabulary, grammar, pronunciation, and spelling). Participants with average ratings below 4 (equal skill in both languages) were considered English dominant, and those with ratings above 4 were considered Spanish dominant; those with an average rating of 4 were classified as English dominant.

Most participants in the English-dominant group began learning Spanish first (49%) or learned both languages simultaneously (31%) but later became more proficient in English. Nearly all participants in the Spanish-dominant group learned Spanish first (98%). Twenty-seven additional participants completed the protocol but were excluded for reporting on the self-report questionnaire that they were not proficient in Spanish (5 participants) or failure to follow instructions (22 participants). Specifically, these participants skipped lines during recall and filled them in later in the recall sequence, such that the written recall sheet did not reflect the retrieval order.

Design

Experiment 1 had a 2 (stimulus format) × 2 (task language) within-subjects design. At study, items were presented as English words, Spanish words, pictures to be recalled in English, or pictures to be recalled in Spanish. The dependent variables were the proportion of items correctly recalled and the degree of semantic clustering in recall output as measured by the Adjusted Ratio of Clustering (ARC; Roenker, Thompson, & Brown, 1971).

Table 1. Mean (SD) characteristics of participants in Experiment 1.

	English Dominant ($N = 77$)	Spanish Dominant ($N = 51$)
Age	20 (2.7)	20 (3.9)
Age of Acquisition – English	3.8 (2.5)	9.7 (4.9)
Age of Acquisition – Spanish	3.5 (4.3)	1.3 (1.9)
Percent Usage of English	64 (20)	34 (16)
Percent Usage of Spanish	25 (15)	61 (17)
Percent Usage of Mixture	11 (19)	5 (7)
Relative Proficiency Rating – Vocabulary[a]	1.8 (1.2)	6.0 (1.4)
Relative Proficiency Rating – Speaking[a]	2.2 (1.2)	5.9 (1.1)

[a]Self ratings of relative proficiency on a labelled 7-point Likert-type scale from 1 (strong advantage for English) to 7 (strong advantage for Spanish); Midpoint of scale is 4 (equal skill in both languages).

Materials

Stimuli were pictures and their names in English and Spanish. The exemplars were chosen from a set of line drawings (including Snodgrass & Vanderwart, 1980 and Pictures Please, Abbate, 1984) on the basis of falling into identifiable semantic categories. Thus, all of the words (picture names) were concrete nouns. All but two of these categories (school supplies; aquatic animals) were categories represented in a set of English-language category exemplar generation norms (Van Overschelde, Rawson, & Dunlosky, 2004). Of those items belonging to the normed categories, the median rank was 5. In a set of Spanish-language category exemplar generation norms (Fernández, Díez, & Ángeles Alonso, 2006), the median rank of the Spanish words was 6. The names had median frequencies of 11 per million in English (SUBTLEX; Brysbaert & New, 2009) and 11 per million in Spanish (SUBTLEX-ESP, Cuetos, Glez-Nosti, Barbón, & Brysbaert, 2011) and mean lengths of 5.4 in English and 6.4 in Spanish. (Note that because no norms were available for Mexican Spanish, Castilian Spanish norms were used to compute ranks and frequencies. For the purpose of approximating these these variables, we looked up ranks and frequencies for the Castilian Spanish equivalents of 7 Mexican Spanish words.) The words had median normative ages of acquisition of 4.4 in English (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) and 3.7 in Spanish (Ángeles Alonso, Fernández, & Díez, 2015). These items were chosen from 16 semantic categories, with 6 items from each category (see Appendix A). The 16 categories were assigned to 4 sets of 4 categories, and these sets were rotated through the experimental conditions across participants using a Latin square to control for specific-item effects. The order of the 24 items in each study sequence was randomised (4 different random orders were used for each set). Each random sequence of words or pictures was printed on a single page.

Procedure

Participants were tested by a bilingual experimenter in sessions lasting approximately 30 min. There were four study-test cycles corresponding to the four cells of the design. The order of languages and formats (word vs. picture) was counterbalanced across participants using a balanced Latin square to control for order and carryover effects. In each block, participants were given 1 min to commit the items to memory under intentional encoding instructions. Participants were told the language of the recall test to follow. Immediately following each study phase, a two-minute free recall test followed with instructions to write responses in the order that they came to mind. Participants also completed a language background questionnaire.

Results

Individual recall sequences were scored for the number of items recalled. The degree of clustering was measured by the ARC (see Roenker et al., 1971, for calculation instructions, formulas, and examples). This measure was chosen because it does not depend upon the number of items recalled, or the length of the recall sequence. A score of zero represents a random sequence, or chance levels of clustering. A score of one represents a perfectly clustered sequence (i.e., blocked by category). Recall and clustering scores are displayed in Table 2.

Free recall scores were submitted to a 2 (stimulus format) × 2 (task language) repeated-measures ANOVA. Recall scores were higher for pictures than for words, $F(1, 127) = 141.77$, $MSE = .012$, $p < .001$, partial $\eta^2 = .53$. Recall scores were higher overall in L1 than in L2, $F(1, 127) = 37.34$, $MSE = .019$, $p < .001$, partial $\eta^2 = .23$, However, this effect was qualified by a significant interaction, $F(1, 127) = 31.70$, $MSE = .010$, $p < .001$, partial $\eta^2 = .20$, showing that the language effect was reliable for pictures, $t(127) = 7.64$, $p < .001$, partial $\eta^2 = .31$ but not for words, $t(127) = 1.74$, $p = .084$, partial $\eta^2 = .02$. This interaction also indicated a stronger picture superiority effect in L1 than in L2.

Clustering scores were submitted to a 2 (stimulus format) × 2 (task language) repeated-measures ANOVA. Clustering was stronger for pictures than for words, $F(1, 127) = 24.49$, $MSE = .183$, $p < .001$, partial $\eta^2 = .16$. The degree of clustering did not differ reliably across languages, $F(1, 127) = 1.71$, $MSE = .099$, $p = .194$, partial $\eta^2 = .01$, for either words, $t(127) = 1.86$, $p = .065$, partial $\eta^2 = .03$, or pictures, $t(127) = .002$, $p = .999$, partial $\eta^2 = .00$. The effects of stimulus format and language did not interact, $F(1, 127) = 2.34$, $MSE = .072$, $p = .128$, partial $\eta^2 = .02$.

Discussion

Free recall accuracy was equivalent for L1 and L2 words, contrary to previous research with similar list structures (e.g., Harris et al., 1995; Lambert et al., 1968). As expected, pictures were recalled more accurately than words, showing the classic picture superiority effect (e.g., Paivio & Csapo, 1973) in both recall languages. The picture superiority effect was stronger in L1 than in L2. This finding suggests that bilinguals covertly named more pictures when studying for an L1 test than when studying for an L2 test, thus creating alternative retrieval routes for more L1 pictures and a greater recall advantage for pictures in L1 recall. This result is consistent with the finding that labelling pictures at study led to better recall (Davies et al., 1973; Horowitz, 1969).

Clustering did not differ reliably for L1 and L2 words, with a small effect size (partial $\eta^2 = .03$, or with a measure that better approximates the population effect size, $d = .14$), consistent with one previous study (Lambert et al., 1968). However, semantic clustering for pictures was stronger than semantic clustering for words as in previous research (Cole et al., 1971; Horowitz, 1969), perhaps because the concepts are accessed more quickly from pictures than from words (Potter & Faulconer, 1975). In the picture encoding conditions, semantic clustering did not depend on the recall language. Therefore, name access does not appear to impact clustering; instead, semantic clustering is based primarily on conceptual access at encoding, which does not depend on the covert naming language or eventual recall language.

Experiment 1 had two important limitations that may have led us to underestimate the effect of language on recall and semantic clustering. First, language dominance was classified according to self-report instead of objective assessments of proficiency, so to the extent that some participants were misclassified, the effects of language may have been underestimated. Second, for this initial study, in the interest of replicating methodology used in the early studies on this topic, we may not have used the most sensitive method for detecting clustering differences between experimental conditions. These limitations are addressed in Experiment 2.

Experiment 2

Experiment 2 incorporated a newer and possibly more sensitive procedure for capturing semantic clustering, which was to have participants memorise lists that included pairs of semantically associated words. Clustering in free recall was measured using a multinomial modelling procedure that allowed for separate estimates of the probability of forming a cluster at encoding and the probability of retrieving the cluster at test (Batchelder & Riefer, 1980; Riefer et al., 2002). Other improvements were made over Experiment 1. First, objective assessments of language proficiency were used to verify bilingual status and classify language dominance, and a monolingual comparison group was included. Second, equal numbers of English-dominant and Spanish-dominant participants were tested so that any language-specific effects associated with English or Spanish would be counterbalanced across L1 and L2.

We also examined whether the effects of study sequence organisation on recall rates and semantic clustering would vary with language proficiency. In the *clustered list* condition, the two words in each pair appeared consecutively in the study sequence. In the *random list* condition, the two words in each pair were separated by several items. Based on previous research (e.g., Bower et al., 1969; Hunt & Einstein, 1981), we expected more items to be recalled in the clustered list condition. Because each successful attempt to retrieve an item from episodic memory increases the activation of semantically associated and temporally contiguous items, the context

Table 2. Mean free recall and clustering scores (SE) in Experiment 1 as a function of language and stimulus format.

Condition	Recall	ARC
Words in L1	.518 (.013)	.313 (.037)
Words in L2	.495 (.012)	.240 (.038)
Pictures with L1 test	.684 (.012)	.464 (.032)
Pictures with L2 test	.560 (.014)	.464 (.033)

maintenance and retrieval model (Polyn et al., 2009) predicts that semantic clustering will be stronger when words at study are both semantically related *and* temporally contiguous, as in our clustered list conditions.

As in Experiment 1, we expected lower recall rates in L2 than in L1 and higher recall rates for monolinguals than for bilinguals. Experiment 2 also addressed the question of whether bilinguals would benefit more or less from the organisation of clustered lists in L2 than in L1 and whether they would benefit more or less than monolinguals. In the one previous study to address this question, bilinguals benefitted from organisation to the same degree in L1 and L2 (Nott & Lambert, 1968).

Predictions for semantic clustering in bilinguals were the same as in Experiment 1. The frequency-lag hypothesis leads to a prediction of stronger clustering in L2 than in L1 and stronger clustering in bilinguals than in monolinguals; the cognitive load conceptualisation leads to a prediction of weaker clustering in L2 than in L1 and weaker clustering in bilinguals than in monolinguals; and the language-general conceptualisation leads to a prediction of equivalent clustering for L2 and L1 and equivalent clustering for bilinguals and monolinguals.

Method

Power and sample size analysis

To be consistent with Experiment 1, Experiment 2 also had 128 bilingual participants, which would yield over 99% power to detect a medium sized ($f = .25$) recall difference between L1 and L2 and over 92% power to detect a small-to-medium sized ($f = .15$) recall difference between L1 and L2. To detect monolingual-bilingual differences (in Experiment 2), 49 participants per group would be sufficient to detect a medium sized effect with 80% power. With 128 participants per group, there was over 99% power to detect a medium-sized group difference and 79% power to detect a small-to-medium sized group difference.

Participants

Participants were 64 English-dominant and 64 Spanish-dominant bilinguals and 128 English-speaking monolinguals. Characteristics of the participants are given in Table 3. Language status classification was based on scores obtained on the Picture Vocabulary subtest of the Woodcock-Muñoz Language Survey-Revised (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005), a standardised objective language assessment normed with English- and Spanish-speaking children and adults in the Americas. The Picture Vocabulary subtest measures productive vocabulary, and the scoring programme provides age-equivalency scores. To qualify as bilingual, a participant needed to demonstrate at least an eight-year-old productive vocabulary level in the less proficient language and be able to converse with the experimenter in both languages. The language in which a participant obtained a higher age-equivalency score was considered to be the dominant language. To qualify as a monolingual English speaker, a participant had to report that not speaking Spanish or any other language proficiently and score at less than an eight-year-old productive vocabulary level in Spanish. Monolingual and bilingual groups were equivalent with respect to age, education level, and socio-economic status, as indicated by parent education levels (Galobardes, Shaw, Lawlar, Lynch, & Davey Smith, 2006).

Fifteen additional participants completed the protocol but were excluded and replaced, 10 because their language assessment scores disqualified them (bilingual scores in English that were too low or monolingual scores in Spanish that were too high), 4 because their language status was accidentally misclassified by the experimenter, and one participant who was an age outlier.

Design

The experimental conditions for bilingual participants formed a 2 (list language) × 2 (list organisation) within-subjects design. Two lists were presented in English and two in Spanish, with one of each language being a clustered list and one being a random list. For monolingual participants, the experimental conditions formed a two-level one-factor (list organisation) within-subjects design. Four lists were presented in English, with two being clustered lists and two being random lists. Recall performance and clustering in recall output were measured.

Materials

The critical stimuli were 40 pairs of nouns chosen from 20 semantic categories (2 pairs per category, see Appendix B). All were among the top 16 associates of their respective categories in English. The median category rank was 3.0 in English (Van Overschelde et al., 2004) and 4.0 in Spanish (Fernández et al., 2000). Because we did not want to assume that the ranks or inter-item associations would be equivalent in English and Spanish, we instead counterbalanced for possible differences by testing half English-dominant and half Spanish-dominant participants. With this manipulation, English (and Spanish) would be L1 for half of the participants and L2 for the other half. The words had a median frequency of 40 per million in English (SUBTLEX; Brysbaert & New, 2009) and 36 per million in Spanish (SUBTLEX-ESP; Cuetos et al., 2011) and mean letter lengths of 5.5 in English and 6.3 in Spanish. The words had median normative ages of acquisition of 4.3 in English (Kuperman et al., 2012) and 3.7 in Spanish (Ángeles Alonso et al., 2015). The four words in each category were assigned to pairs that maximised similarity between the two members of each pair (e.g., *cat-dog* and *horse-cow* rather than *cat-horse* and *dog-cow*.) The pairs were randomly assigned to four sets of 10 pairs, with the restriction that all pairs in any set were from different categories. The four lists were rotated through the four language and list type conditions across participants using a Latin square. In clustered lists, the words of each

Table 3. Mean (SD) characteristics of participants in Experiment 2.

	English Monolingual ($N = 128$)	English Dominant ($N = 64$)	Spanish Dominant ($N = 64$)
Age	21 (4.6)	22 (5.1)	21 (5.0)
Age of Acquisition – English	2.0 (1.5)	4.9 (3.1)	7.2 (3.2)
Age of Acquisition – Spanish	–	2.0 (3.0)	1.5 (1.4)
Percent Usage of English	89 (13)	62 (20)	38 (20)
Percent Usage of Spanish	7 (9)	27 (15)	44 (21)
Percent Usage of Mixture	4 (6)	11 (19)	18 (29)
Age Equivalency – English Picture Vocabulary[a]	16.1 (4.7)	14.2 (3.3)	10.4 (1.5)
Age Equivalency – Spanish Picture Vocabulary[a]	3.5 (1.8)	10.3 (1.7)	13.3 (2.9)
Relative Proficiency Rating – Vocabulary	1.2 (0.6)	2.4 (1.2)	4.2 (1.5)
Relative Proficiency Rating – Speaking	1.2 (0.6)	2.8 (1.3)	4.5 (1.4)
Median Education Level	some college	some college	some college
Median Highest Parent Education Level	graduated college	graduated college	graduated college

[a]Score obtained from the Woodcock-Muñoz Language Survey-Revised (Woodcock et al., 2005).

pair appeared consecutively, and 5 filler words were presented at the beginning and end to control for primacy and recency effects. Thus, the list had 30 words in all. In non-clustered lists, the two words in each pair were separated by at least five intervening items, with one word appearing in the first half and the other appearing in the second half of the sequence. Again, 5 filler words were presented at the beginning and end of the list to control for primacy and recency effects.

Procedure

The Picture Vocabulary subtest of the Woodcock-Muñoz Language Survey-Revised (Woodcock et al., 2005) was administered in English and Spanish to each participant. They also completed demographic and language background questionnaires.

In the main experiment, words to be committed to memory were presented on the monitor of an iMac desktop computer, and the sequence and timing of stimulus presentation was controlled using PsyScope X software (Cohen, MacWhinney, Flatt, & Provost, 1993). There were four study-test cycles, one for each combination of language and list type, with the order counterbalanced across participants. At study, word stimuli were presented at a rate of 2 sec per item. Participants wrote the words that they recalled on a worksheet. Monolinguals performed the same memory tasks in English.

Results

Recall scores are displayed in Table 4. An initial analysis of bilingual recall performance used the standard method, in which the proportions of items recalled in each condition were submitted to a 2 (list language) × 2 (list organisation) repeated measures ANOVA. Item recall did not differ across languages, $F < 1$, partial $\eta^2 = .00$. A main effect of list organisation, $F(1, 127) = 46.774$, $MSE = .016$, $p < .001$, partial $\eta^2 = .27$, indicated that clustered lists were better recalled than random lists. The effects of list language and list organisation did not interact, $F < 1$, partial $\eta^2 = .01$. Monolingual participants also showed higher item recall for clustered lists relative to random lists, $t(127) = 7.124$, $p < .001$, partial $\eta^2 = .29$. Separate 2 (language group) × 2 (list organisation) mixed ANOVAs were performed to compare recall performance for monolinguals and bilinguals. Item recall did not differ for monolinguals and bilinguals performing the task in L1 or L2, $F\text{s} < 1$, partial $\eta^2\text{s} = .00$, and the effects of language group and list organisation did not interact, $F\text{s} < 1$, partial $\eta^2\text{s} = .00$.

Table 4. Recall proportions (SE) in Experiment 2 as a function of language and list organisation.

	Monolingual	Bilingual L1	Bilingual L2
Clustered List	.397 (.012)	.385 (.016)	.390 (.016)
Random List	.312 (.013)	.321 (.015)	.303 (.016)

To evaluate whether language proficiency scores might be associated with recall performance if considered quantitatively, two ANCOVAs were conducted. In the first ANCOVA on English recall scores, English language proficiency scores were entered as a covariate, but the effect of this covariate was not reliable ($F < 1$). In the second ANCOVA on Spanish recall scores, Spanish language proficiency scores were entered as a covariate, but the effect of this covariate was not reliable ($p = .09$).

Recall accuracy and clustering were analyzed using a multinomial processing tree approach (Batchelder & Riefer, 1980; Riefer et al., 2002), and the specific model is illustrated in Figure 1. Recall responses for each semantically related word pair were classified into one of 4 categories: both items recalled consecutively, both items recalled but not consecutively, one item recalled, or neither item recalled. The probability of each outcome is determined by adding the products of probabilities of the cognitive processes on each path to that outcome. With 4 possible outcomes, 4 equations are generated in this manner. The 3 free data points allow the estimation of 3 parameters: C = probability of forming a cluster, R = probability that a cluster is retrieved, and U = probability that an un-clustered item is retrieved. These three parameters were estimated separately for clustered and random lists using multiTree (Moshagen, 2010). The parameters were first estimated for monolinguals, bilingual L1, and bilingual L2 separately and these estimated

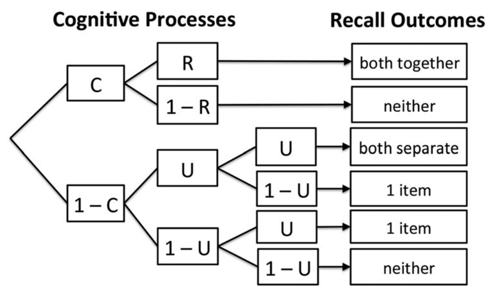

Figure 1. The Pair Clustering Model (adapted from Riefer et al., 2002). Cognitive processes involved include C = probability of forming a cluster at study; R = probability of recalling a cluster at test; and U = probability of recalling an item that was not part of a cluster. Recall outcomes indicate the recall status of the two items in a pair selected from the same semantic category. The two items can be recalled consecutively (both together) or non-consecutively (both separate), or only one item or neither item is recalled.

values from the saturated model are given in Table 5 along with their standard errors.

We then used multiTree to fit a hierarchical lattice of models that restricted different parameters to be equivalent across list types, languages, and language groups. Restricting the C parameter to be equal across *list types* resulted in a significant reduction in model fit for monolinguals, $G^2(1) = 37.2$, $p < .001$, $w = .09$, for bilinguals performing recall in L1, $G^2(1) = 12.5$, $p < .001$, $w = .06$, and for bilinguals performing recall in L2, $G^2(1) = 29.6$, $p < .001$, $w = .07$. These reductions in fit indicate that the probability of forming a cluster at encoding is greater for semantically organised than for random lists. Restrictions of the R or U parameters to be equal across list types did not result in a significant reduction in model fit for any language or group ($ps > .05$, $ws < .02$), indicating that the probabilities of retrieving clusters and unclustered items did not differ across list types.

In bilinguals, restricting the C, R, or U parameters to be equal across languages did not result in significant reductions in model fit ($ps > .10$; $ws < .02$), indicating that the probabilities of forming a cluster, retrieving a cluster, and retrieving an unclustered item did not differ across languages. Similarly, restricting the C, R, or U parameters to be equal for monolinguals and bilinguals performing the task in L1 or L2 had no effect on model fit ($ps > .20$, $ws < .025$), indicating that the probabilities of forming a cluster, retrieving a cluster, and retrieving an unclustered item did not differ for monolinguals and bilinguals.

Table 5. Multinomial model parameter estimates (SE) in Experiment 2 as a function of language and list organisation.

List Organisation Language	Parameter *C*	*R*	*U*
Clustered List			
Monolingual	.658 (.030)	.452 (.023)	.290 (.026)
Bilingual L1	.575 (.044)	.443 (.037)	.303 (.033)
Bilingual L2	.641 (.035)	.419 (.027)	.344 (.035)
Random List			
Monolingual	.298 (.042)	.374 (.054)	.288 (.018)
Bilingual L1	.315 (.048)	.290 (.049)	.332 (.025)
Bilingual L2	.234 (.066)	.404 (.116)	.274 (.025)

Note: C = probability of forming a cluster; R = probability of retrieving a cluster; U = probability of retrieving an unclustered item.

Discussion

The standard analysis of recall rates showed a large effect of list organisation (partial $\eta^2 = .27$) but no differences in performance across languages (partial $\eta^2 = .00$) or language groups (partial η^2s = .00) and no interactions. Similarly, the multinomial model analysis showed no differences across languages or language groups in the probabilities of retrieving clusters or un-clustered items.

According to the multinomial model analysis, the probability of forming a semantic cluster was higher for clustered than for random lists. However, the probability of forming a semantic cluster did not differ across languages or groups. The absence of differences in semantic clustering across languages and groups is most consistent with the shared semantic association approach to L2 memory. In both recall rates and semantic clustering, the benefit of a semantically organised study sequence over a random study sequence was equivalent for L1 and L2 and equivalent for bilinguals and monolinguals, consistent with the one previous study to make this comparison (Nott & Lambert, 1968).

General discussion

As explained in the introduction, most previous studies did not cross language dominance with task language, and the only study of semantic clustering that did cross these factors had insufficient sample size, so definitive conclusions about the effects of language proficiency on recall performance and semantic clustering could not be drawn. In the present study, both English-dominant and Spanish-dominant participants (as determined through self reports in Experiment 1 and objective standardised assessments in Experiment 2) were tested in both English and Spanish, and the sample sizes were relatively large. Note also that most previous studies did not include objective assessments of language proficiency or show that socio-economic status was comparable across comparison groups as done in Experiment 2. In the following sections we discuss bilingual and monolingual semantic clustering, free recall accuracy and implications for conceptualizations of bilingual memory.

Semantic clustering

The degree of semantic clustering in recall output was compared across languages and language groups. The present study is the first to make these comparisons in a relatively large sample. In Experiments 1 and 2, semantic clustering in free recall did not differ reliably for L1 and L2 word lists, consistent with the results of one previous study (Lambert et al., 1968) but contrary to the results of two other studies (Champagnol, 1973; Nott & Lambert, 1968). In Experiment 2, the degree of semantic clustering in free recall output did not differ for monolinguals and bilinguals, which is consistent with a previous result showing equivalent semantic clustering for unbalanced bilinguals and monolinguals (Harris et al., 1995), but inconsistent with a previous result showing stronger semantic clustering for balanced bilinguals relative to monolinguals (Harris et al., 1995). The patterns of performance observed are inconsistent with the predictions of the familiarity and cognitive load conceptualizations of bilingual memory but consistent with the shared semantic association conceptualisation.

Experiment 1 showed for the first time that semantic clustering in picture recall did not depend on the recall language, which is consistent with the conclusion that semantic clustering depends on semantic access but not name access. Semantic clustering in free recall of pictures was stronger than in free recall of words as in previous research (Cole et al., 1971; Horowitz, 1969).

In Experiment 2, clustering in free recall output was stronger when the study sequence was semantically organised than when it was random. This effect was observed in monolingual participants (as in Cofer, 1966) and in bilingual participants in both languages (as in Nott & Lambert, 1968). The context maintenance and retrieval model (Polyn et al., 2009) correctly predicted that clustering would be stronger when words at study were both semantically related *and* temporally contiguous, as in our clustered list conditions. Indeed, the present data showed this pattern for both monolingual and bilingual participants. The effect of list organisation on output clustering did not differ across languages in bilinguals or between monolingual and bilingual participants.

Free recall accuracy

In Experiments 1 and 2, bilingual free recall performance was equivalent for L1 words and L2 words, contrary to past studies that showed better item recall for L1 than L2 words when study sequences contained a limited number of semantic categories (e.g., Champagnol, 1973; Harris et al., 1995; Lambert et al., 1968; Nott & Lambert, 1968). Based on the familiarity or cognitive load conceptualizations of bilingual memory described in the introduction, we would have expected recall performance to be worse for L2 than for L1, but the shared semantic association perspective would allow for equivalent L1 and L2 performance.

We considered whether treating relative proficiency dichotomously in terms of the more and less proficient language weakened power to detect differences. However, in alternative analyses of Experiment 2 that treated English and Spanish proficiency as continuous variables, proficiency in a language did not reliably predict recall performance.

We also considered the possibility that the early bilingual population from which we recruited our sample might have had such a small proficiency difference between L1 and L2 that the L2 disadvantage was not strong enough to impact intentional encoding and free recall. However, language dominance was assessed objectively in Experiment 2, using a standardised testing instrument, and the L2 proficiency criterion was not strict. Also, with 256 bilingual participants, the present study was better powered than previous studies of L1 and L2 free recall. Detecting free recall performance differences across languages may require having participants memorise a larger number of lists. In previous studies with the same participant population, we found no reliable difference between L1 and L2 free recall accuracy when there was only one list per language (Francis & Baca, 2014), but there was a clear advantage for L1 over L2 recall when participants memorised sixteen lists in each language (Francis et al., 2018).

We found no difference in recall accuracy between monolingual and bilingual participants in Experiment 2. Based on the familiarity or cognitive load perspectives on bilingual memory described in the introduction, we would have expected recall performance to be worse for bilinguals than for monolinguals, but the shared semantic association perspective would allow for equivalent monolingual and bilingual performance. It is important to note that previous results on this matter were mixed. Previous research suggested that monolingual performance would be equivalent to L1 but superior to L2 performance in unbalanced bilinguals when lists contained items chosen from a limited set of categories (Harris et al., 1995). With balanced bilinguals, past research results were mixed (Fernandes et al., 2007; Haritos, 2002; Harris et al., 1995).

Pictures were recalled at a higher rate than words, showing the classic picture superiority effect (e.g., Paivio & Csapo, 1973) in both recall languages. The picture superiority effect was stronger in L1 than in L2, which suggests that bilinguals covertly named more pictures when studying for a test in L1 than when studying for a test in L2. Access to the names created alternative retrieval routes for more L1 pictures, thus leading to a greater recall advantage for pictures when retrieval was expected to be in L1. This result is consistent with monolingual studies showing that when participants labelled the pictures that they studied, they were remembered better (Davies et al., 1973; Horowitz, 1969).

An alternative explanation for the larger picture superiority effect in L1 relative to L2 is that bilingual participants covertly named pictures in L1 when they were to later recall in L2 and this created a mismatch between the language of study and the language of test that hurt recall performance. We argue that this explanation is unlikely for three reasons. First, given that the participants knew at the time of study what language they would be tested in, it seems unlikely that they would use covert naming in the non-target language as a memory strategy.

Second, although there is some evidence that the non-target language is activated during picture naming (e.g., Costa, Miozzo, & Caramazza, 1999), there is also evidence that bilinguals do not covertly name pictures in the non-target language or translate the names to the non-target language. Specifically, when the picture-naming language changed from encoding to test, priming was reduced substantially, and priming from different-language naming and word translation to the target language had additive effects in facilitating later picture naming, whether final naming was in L1 or L2 (Francis, Corral, Jones, & Sáenz, 2008). Third, if bilinguals were to covertly name in L1 in conditions where they would be tested in L2, we would not expect their recall to be less accurate than in L1 conditions. In a classic study, Ervin (1961) showed that when bilinguals named pictures aloud and later recalled them, performance did not depend on whether the naming language and test language matched or not. Also, if they covertly named in the wrong language, they could have simply translated the recalled words at test, which would lead to equivalent performance in L1 and L2. Alternatively, they could have translated the words at study, which likely would have led to superior performance in L2 conditions, because translation of words at study makes them more memorable (e.g., Paivio & Lambert, 1981), presumably because of the conceptual access involved and having an additional retrieval route.

When words were presented in a semantically organised sequence at study, monolingual participants exhibited better item recall than when the words were presented in a random sequence, consistent with the predictions of the context maintenance and retrieval model (Polyn et al., 2009) and past research (e.g., Bower et al., 1969; Hunt & Einstein, 1981). The present study replicated this result in bilinguals (as in Nott & Lambert, 1968). The benefit of semantic organisation did not differ for L1 and L2 or between monolingual and bilingual participants.

Implications for conceptualizations of L2 memory

According to the low-familiarity conceptualisation, L2 words should function in a manner similar to that of lower frequency words in L1. Many studies have shown that low-frequency words are not recalled as well as high-frequency words (e.g., Balota & Neely, 1980; Kinsbourne & George, 1974), which led to the prediction that recall performance would be worse for L2 words than for L1 words. This prediction was not supported, in that recall performance was equivalent for L1 and L2 words and equivalent for monolinguals and bilinguals. Two previous studies showed that low-frequency words exhibited a higher degree of clustering in recall output than high-frequency words (Jordan & Swartz, 1976; Matthews, 1966), which led to the prediction that L2 words should show stronger semantic clustering than L1 words. Contrary to this prediction, the degree of clustering in recall output did not differ across languages or between monolinguals and bilinguals.

Although low-frequency words in L1 and high-frequency words in L2 both have weaker associations with their concepts than high-frequency words in L1, there are important differences between them. Most importantly, high-frequency L2 words are associated with high-frequency concepts that have been accessed many times through L1. This means that the concepts associated with L2 words are no less familiar than concepts associated with L1 words, because conceptual representations for translation-equivalents are shared across languages (for reviews, see Francis, 1999, 2005). In contrast, low-frequency L1 words are generally associated with less familiar concepts. This difference may lead to dissociative effects of language proficiency and word frequency for tasks that rely heavily on semantic processing, in that we might expect to see frequency effects but not proficiency effects, as appears to be the pattern for previous findings of word frequency effects on free recall and semantic clustering and the absence of language effects on free recall and semantic clustering in the present study.

The present results therefore indicate limitations on the generality of the low-frequency word analogy for L2 word processing. Nevertheless, the dissociation of the effects of word frequency and language proficiency on recall performance and semantic clustering in recall output do not necessarily falsify the frequency-lag hypothesis. The frequency-lag hypothesis deals with associations between words and concepts (Gollan et al., 2008), whereas recall performance and semantic clustering depend heavily on associations among related concepts.

According to the cognitive-load conceptualisation, L2 words should function in a manner similar to L1 words processed under a cognitive load. Several studies have shown that imposing a cognitive load at encoding impairs recall performance (e.g., Craik et al., 2000), which led to the prediction that recall performance would be worse in L2 than in L1. This prediction was not supported, in that free recall of words was equivalent for L1 and L2 and equivalent for monolinguals and bilinguals. One previous study showed that imposing a cognitive load at encoding reduced clustering in recall output (Park et al., 1989), which led to the prediction that L2 words would show weaker semantic clustering than L1 words. However, the degree of clustering in recall output did not differ across languages or between monolinguals and bilinguals. Thus, although processing L2 words requires more attention than processing L1 words, this additional cognitive load has little, if any, impact on the utilisation of pre-existing semantic associations.

According to the perspective that long-standing semantic associations are shared across languages, no effects of language or language group on semantic clustering in recall output would be expected. In Experiments 1 and 2, clustering was equivalent across languages and language groups. These findings indicate that the semantic associations used, or the semantic clusters formed at encoding were not tied to one specific language. This finding converges with previous results using repetition-priming methodology that also provide evidence that associations at the semantic level are shared across languages. First, repetition priming in category exemplar generation transferred across languages in bilinguals (Francis et al., 2010), indicating that category-exemplar relationships are shared across languages. Second, repetition priming in verb generation (generating an appropriate verb to a noun cue) transferred across languages in bilinguals (de la Riva López et al., 2012; Seger et al., 1999), indicating that object-action associations are shared across languages. Third, semantic processing of adjectives in one language at encoding elicited priming when generating the words as antonyms in the other language at test (Taylor & Francis, 2017), indicating that antonym relationships are shared across languages. These results taken together with the present results provide evidence that long-standing semantic associations are stored in a language-general form in the language-general conceptual system of semantic memory. However, we would not expect representations of *non-semantic* associations to be language-general. For example, lexical associations among words that merely co-occur frequently in a particular language are likely represented in language-specific networks.

Conclusions

Bilingual free recall accuracy and semantic clustering did not differ across more and less proficient languages or across monolingual and bilingual speakers. The finding that the effect of language proficiency on semantic clustering is small to non-existent indicates that language proficiency does not impact the utilisation of long-standing semantic associations while performing an explicit memory task. This finding converges with previous evidence indicating that long-standing semantic associations are language general.

Acknowledgements

We thank Daniel Cruz and Danielle Glover for assistance with data collection and Michelle Martinez for assistance with data entry. Results of Experiment 2 were presented at the 55th Annual Conference of the Psychonomic Society.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Experiment 1 was supported by National Science Foundation grant BCS-0924905 and a Mini-Grant to the first author funded by UTEP Teachers for a New Era, Carnegie Corporation of New York. Experiment 2 was supported by National Institutes of Health Grant R15HD07892 to the first author.

References

Abbate, M. S. (1984). *Pictures please: An articulation supplement*. Tucson: Communication Skill Builders.

Abu-Rabia, S. (2003). The influence of working memory on reading and creative writing processes in a second language. *Educational Psychology*, *23*, 209–222.

Ángeles Alonso, M., Fernández, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7039 spanish words. *Behavior Research Methods*, *47*, 268–274.

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, *36*, 233–240.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576–587.

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review*, *87*, 375–397.

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

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Champagnol, R. (1973). Organisation sémantique et linguistique dans le rappel libre bilingue. [Semantic and linguistic organization in bilingual free recall.] *L'Année Psychologique*, *73*, 115–134.

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

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271.

Cole, M., Frankel, F., & Sharp, D. (1971). Development of free recall learning in children. *Developmental Psychology*, *4*, 109–123.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365–397.

Craik, F. I. M., Naveh-Benjamin, M., Ishaik, G., & Anderson, N. D. (2000). Divided attention during encoding and retrieval: Differential control effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1744–1749.

Cuetos, F., Glez-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica*, *32*, 133–143.

da Costa Pinto, A. (1991). Reading rates and digit span in bilinguals: The superiority of mother tongue. *International Journal of Psychology*, *26*, 471–483.

Davies, G. M., Milne, J. E., & Glennie, B. J. (1973). On the significance of "double encoding" for the superior recall of pictures to names. *Quarterly Journal of Experimental Psychology*, *25*, 413–423.

de la Riva López, E. M., Francis, W. S., & García, J. (2012). Repetition priming within and between languages in verb generation: Evidence for shared verb concepts. *Memory*, *20*, 358–373.

Durgunoglu, A. Y., & Roediger, H. L. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language*, *26*, 377–391.

Durso, F. T., & Johnson, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, *19*, 416–429.

Ervin, S. M. (1961). Learning and recall in bilinguals. *The American Journal of Psychology*, *74*, 446–451.

Fernandes, M. A., Craik, F. I. M., Bialystok, E., & Kreuger, S. (2007). Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. *Canadian Journal of Experimental Psychology*, *61*, 128–141.

Fernández, A., Díez, E., & Ángeles Alonso, M. (2006). Elaboración de normas de asociación libre en castellano y desarrollo de una base de datos integrada. [based on data from Soto, P., Sebastián, M. V., García, E. & del Amo, T. (1982). Categorización y datos normativos en España. Madrid: Universidad Autónoma. (Reedited in 1994 as Las categorías y sus normas en castellano. Madrid: Visor)] Retrieved from http://inico.usal.es/usuarios/emid/nipejcyl/consulta_norma.asp?id=13

Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin*, *125*, 193–222.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 251–267). New York, NY: Oxford University Press.

Francis, W. S., & Baca, Y. (2014). Effects of language dominance on item and order memory in free recall, serial recall, and order reconstruction. *Memory*, *22*, 1060–1069.

Francis, W. S., Corral, N. I., Jones, M. L., & Sáenz, S. P. (2008). Decomposition of repetition priming components in picture naming. *Journal of Experimental Psychology: General*, *137*, 566–590.

Francis, W. S., Fernandez, N. P., & Bjork, R. A. (2010). Conceptual and non-conceptual repetition priming in category exemplar generation: Evidence from bilinguals. *Memory*, *18*, 787–798.

Francis, W. S., Liaño, M. K., & Taylor, R. S. (2018). *Effects of word frequency and bilingual proficiency on the temporal dynamics of free recall*. Manuscript submitted for publication.

Francis, W. S., Regalado, P., Sáenz, S. P., & Durán, G. (2006). Asymmetries in speed and accuracy of bilingual comprehension and production processes. *Abstracts of the Psychonomic Society*, *11*, 89.

Galobardes, B., Shaw, M., Lawlar, D. A., Lynch, J. W., Davey Smith, G. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, *60*, 7–12.

Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 625–630.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787–814.

Haritos, C. (2002). A developmental examination of memory strategies in bilingual six, eight, and ten year olds. *International Journal of Bilingual Education and Bilingualism*, *5*, 197–220.

Harris, J. G., Cullum, C. M., & Puente, A. E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults. *Journal of the International Neuropsychological Society*, *1*, 10–16.

Horowitz, A. B. (1969). Effect of stimulus presentation modes on children's recall and clustering. *Psychonomic Science*, *14*, 297–298.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 497–514.

Jordan, D. E., & Swartz, P. E. (1976). Word frequency as a determinant of clustering when taxonomic frequency is held constant. *Canadian Journal of Psychology*, *30*(4), 187–192.

Keniston, A. H., & Flavell, J. H. (1979). A developmental study of intelligent retrieval. *Child Development*, *50*, 1144–1152.

Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 63–69.

Kormi-Nouri, R., Moniri, S. & Nilsson, L.-G. (2003). Episodic and semantic memory in bilingual and monolingual children. *Scandinavian Journal of Psychology*, *44*, 47–54.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Lambert, W. E., Ignatow, M., & Krauthamer, M. (1968). Bilingual organization in free recall. *Journal of Verbal Learning and Verbal Behavior*, *7*, 207–214.

López, M., & Young, R. K. (1974). The linguistic interdependence of bilinguals. *Journal of Experimental Psychology*, *102*, 981–983.

Matthews, W. A. (1966). Continued word associations and free recall . *Quarterly Journal of Experimental Psychology*, *18*, 31–38.

Moshagen, M. (2010). Multitree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54.

Naveh-Benjamin, M., Craik, F. I. M., Perretta, J. G., & Tonev, S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53A*, 609–625.

Nott, C. R., & Lambert, W. E. (1968). Free recall of bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *7*, 1065–1071.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*, 176–206.

Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 532–539.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*, 137–138.

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1185–1191.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational process in free recall. *Psychological Review*, *116*(1), 129–156.

Potter, M. C., & Faulconer, B. A. (1975, Feb 6). Time to understand pictures and words. *Nature*, *253*, 437–438.

Ransdell, S., Arecco, M. R., Levy, C. M. (2001). Bilingual long-term working memory: The effects of working memory loads on writing quality and fluency. *Applied Psycholinguistics*, *22*, 113–128.

Riefer, D. M., Knapp B. R., Batchelder, W. H., Bamber D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184–201.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparisons of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45–48.

Santa, J. L., Ruskin, A. B., Snuttjer, D., & Baker, L. (1975). Retrieval in cued recall. *Memory & Cognition*, *3*, 341–348.

Seger, C. A., Rabin, L. A., Desmond, J. E., & Gabrieli, J. D. E. (1999). Verb generation priming involves conceptual implicit memory. *Brain and Cognition*, *41*, 150–177.

Service, E., Simola, M., Metsaenheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*, 383–408.

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

Takano, Y., & Noda, A. (1993). A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology*, *24*, 445–462.

Taylor, R. S., & Francis, W. S. (2017). Between-language repetition priming in antonym generation: Evidence that translation-equivalent adjectives have shared conceptual representations across languages. *Memory*, *25*, 344–349.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335.

Whiting, W. L. (2003). Adult age differences in divided attention: Effects of elaboration during memory encoding. *Aging, Neuropsychology, & Cognition*, *10*, 141–157.

Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). *Woodcock-Muñoz language survey–revised*. Rolling Meadows, IL: Riverside Publishing.

Appendix A: English and Spanish Stimuli used in Experiment 1.

Animals		Footwear	
dog	perro	shoe	zapato
cat	gato	boot	bota
horse	caballo	sock	calcetín
cow	vaca	skis	esquís
elephant	elefante	slippers	pantunflas
bear	oso	sandals	sandalias
Tools		Instruments	
hammer	martillo	drum	tambor
screwdriver	desarmador	trumpet	trompeta
ladder	escalera	guitar	guitarra
nail	clavo	bell	campana
saw	sierra	accordion	acordeón
wrench	llave	saxophone	saxofón
Insects/Bugs		Furniture	
fly	mosca	chair	silla
ant	hormiga	table	mesa
bee	abeja	bed	cama
spider	araña	desk	escritorio
butterfly	mariposa	lamp	lámpara
worm	gusano	mirror	espejo
School Supplies		Vegetables	
pencil	lápiz	carrot	zanahoria
paper	hojas	corn	elote
book	libro	tomato	tomate
backpack	mochila	lettuce	lechuga
eraser	borrador	onion	cebolla
scissors	tijeras	peas	chícharos
Aquatic Animals		Clothing	
dolphin	delfín	shirt	camisa
octopus	pulpo	pants	pantalón
seahorse	caballo de mar	skirt	falda
whale	ballena	coat	abrigo
fish	pescado	dress	vestido
seal	foca	sweater	suéter
Toys		Fruit	
ball	pelota	apple	manzana
doll	muñeca	orange	naranja
kite	papalote	pear	pera
puzzle	rompecabezas	grapes	uvas
swing	columpio	strawberry	fresa
skateboard	patineta	watermelon	sandía
Birds		Parts of the Body	
penguin	pingüino	leg	pierna
owl	búho	arm	brazo
duck	pato	eye	ojo
ostrich	avestruz	nose	nariz
rooster	gallo	ear	oreja
eagle	águila	hand	mano
Kitchen Utensils		Transportation	
knife	cuchillo	car	carro
spoon	cuchara	bus	camión
fork	tenedor	airplane	avión
cup	taza	train	tren
plate	plato	bicycle	bicicleta
glass	vaso	motorcycle	moto

Note: Participants were never given the category labels.

Appendix B: English and Spanish Stimuli used in Experiment 2.

Category Set A.

Pair 1		Pair 2	
Vegetables			
onion	cebolla	carrot	zanahoria
spinach	espinacas	lettuce	lechuga
Parts of the Body			
head	cabeza	leg	pierna
nose	nariz	arm	brazo
Human Dwellings			
apartment	departamento	house	casa
dorm	dormitorio	cabin	cabaña
Animals			
horse	caballo	dog	perro
cow	vaca	cat	gato
Occupations/Professions			
lawyer	abogado	doctor	médico
teacher	maestro	nurse	enfermera
Carpenters' Tools			
hammer	martillo	saw	sierra
nail	clavo	screwdriver	desarmador
Non-alcoholic Drinks			
coffee	café	milk	leche
lemonade	limonada	juice	jugo
Units of Time			
week	semana	hour	hora
month	mes	minute	minuto
Furniture			
chair	silla	bed	cama
table	mesa	desk	escritorio
Musical Instruments			
flute	flauta	clarinet	clarinete
trumpet	trompeta	saxophone	saxofón

Category Set B.

Pair 1		Pair 2	
Fruit			
grape	uva	apple	manzana
strawberry	fresa	banana	plátano
Clothing			
dress	vestido	shirt	camisa
skirt	falda	pants	pantalones
Parts of a Building			
stairs	escaleras	window	ventana
elevator	elevador	door	puerta
Insects			
bee	abeja	ant	hormiga
butterfly	mariposa	spider	araña
Relatives			
father	padre	mother	madre
grandfather	abuelo	daughter	hija
Kitchen Utensils			
knife	cuchillo	spoon	cuchara
fork	tenedor	plate	plato
Substances for Flavouring Food			
sugar	azucar	salt	sal
cinnamon	canela	pepper	pimienta
Weather			
snow	nieve	hurricane	huracán
rain	lluvia	lightning	relámpago
Colors			
green	verde	blue	azul
yellow	amarillo	red	rojo
Transportation			
car	carro	bus	camión
airplane	avión	train	tren