Divergence and Overlap in Bilingual Concept Representations

Thomas B. Ward (tward@bama.ua.edu)
Center for Creative Media, Box 870172, University of Alabama
Tuscaloosa, AL 35487 USA

Amy H. Chu (chu004@bama.ua.edu)
College of Communication and Information Sciences, Box 870172, University of Alabama
Tuscaloosa, AL 35487 USA

Jyotsna Vaid (jxv@psyc.tamu.edu)
Department of Psychology, Texas A&M University
College Station, TX 77843 USA

Roberto R. Heredia (rheredia@tamiu.edu)
Department of Psychology, Texas A&M International University
Laredo, TX 78041 USA

Abstract

Chinese-English bilingual participants listed exemplars of 10 common categories on two occasions, one week apart. Half responded in the same language in both sessions (Chinese or English) and half responded in one language in one session and the other language in the other session. There was substantial overlap in the exemplars listed across the sessions, but those responding in different languages showed less overlap than those responding in the same language. The results indicate differences in graded structure of the categories across the participants' languages; the exemplars associated with translation equivalent category labels in bilinguals' two languages differ in how representative they are of the category.

Keywords: bilingualism; concept representation.

Introduction

To what extent do the conceptual representations associated with translation equivalents in a bilingual person's two languages overlap or diverge? This question relates to, but goes beyond the issue of whether a bilingual person has two separate lexical systems each associated with separate underlying conceptual representations or an integrated, shared system. Francis (1999; in press) has reviewed the extensive literature on this topic and concluded that the evidence supports a single, integrated concept representation view. At the same time, however, cross-language effects on a variety of tasks ranging from priming to Stroop interference typically tend to be weaker than comparable same-language effects, suggesting that although a bilingual person's languages may access a common conceptual system, the mapping of lexical to conceptual referents and perhaps even the nature of the conceptual referents themselves may differ across the two languages (see Paradis, 1979; de Groot, 1992; de Groot, Dannenburg, & Van Hell, 1994).

As a way of gaining insight into the question of potentially differing bilingual concept representations, the present study made use of the well-known fact that exemplars of a given category exhibit a graded structure. That is, within a given category, certain of its exemplars appear to be more representative of the category than others. Variations in exemplar representativeness are evident in several measures, including output dominance, in which some instances are consistently listed as category members earlier and more frequently than others (e.g., Battig & Montague, 1969), and typicality or goodness of example, in which some instances are reliably rated as better examples of the category than others (e.g., Rosch, 1975). Output dominance and typicality are positively correlated; items that are listed early and by many respondents also tend to be rated high in typicality (Mervis, Catlin & Rosch, 1976.)

In the present study, we focused specifically on output dominance as a measure for assessing the similarities and differences in the way the underlying concepts associated with translation-equivalent pairs are represented in an individual's two languages (e.g., the match between the conceptual representations associated with the words FURNITURE in English and MUEBLES in Spanish). To what extent do those corresponding terms access the same or different sets of exemplars, attributes and understandings about the concepts in question? For example, when a Spanish/English bilingual is asked to list exemplars of FURNITURE (in English), do the same instances come to mind for that individual as when asked to list exemplars of MUEBLES (in Spanish)? Although several studies have documented differences in category exemplar listing and typicality across different groups that use different languages, examinations of differences within the same bilinguals across their two languages are rare (e.g., Roberts & LeDorze, 1997), and no published studies have examined such differences for a broad sampling of concept types.

Suggestive evidence that different exemplars come to mind when bilingual individuals access concepts via their two languages comes from a study by Pena, Bedore, and Zlatic-Giunta (2002). Those investigators had Spanish-English bilingual children between the ages of 4;5 and 7;1 name animals, clothing and food in Spanish on one occasion and in English on another occasion. The children generated about the same number of items in each language, but their overall tendency was to list different items in each language (e.g., elephant, lion and dog as the top items in English versus caballo (horse), elefante (elephant) and tigre (tiger) in Spanish). In fact on average, only about one third of the children's responses were *doublets* (i.e., listed in Spanish and English), whereas two thirds were *singlets* (i.e. listed in one language or the other, but not both). Because none of the bilingual children was tested twice in the same language, it is not certain that the proportion of doublets was higher than it would have been for children responding in the same language on two separate occasions, but the results provide at least suggestive evidence for a different pattern of association of exemplars with categories in the children's two languages. Roberts and LeDorze (1997) also had French-English bilinguals list exemplars of foods and animals in each of those languages and found that there was substantial, but less than perfect overlap in the sets of exemplars listed.

Data reported by Jeng, Lai, & Liu (1973) allow a look at a much wider set of categories and provide indirect support for the idea that the exemplars that come to mind in response to category labels will differ across translation equivalents. Jeng et al. obtained listings for categories in the Battig and Montague (1969) norms from Chinese-English bilingual students at two universities in Taiwan. Some of the participants were given labels for the categories in Chinese and listed exemplars in Chinese. Others were given English labels and responded in English. Jeng et al. reported frequencies for listings of exemplars for each of the categories separately for the respondents who listed in Chinese and English, but did not report analyses comparing the patterns of responses by the two groups.

A preliminary examination and comparison we made of the Jeng et al. lists suggests that there is substantial overlap, but there are also striking differences in the exemplars that come to mind depending on the language in which the participants responded. For the category SPORT, for example, we found that the correlation between the output dominance of particular exemplars between the Chinese and English responses was .52. This significant positive correlation indicates, as would be expected, that there is at least some overlap in the exemplars that are strongly associated with the concept SPORT for Chinese bilinguals when accessing the concept in either language. More striking however, when we compared the Jeng et al. data to listings obtained on largely monolingual US college students from our own datasets, the correlation between output dominace for Jeng et al.'s English responses and responses by our US college students was .79. In other words, the sports that came to mind most readily when Chinese participants listed exemplars of the category of SPORTS in English were more similar to those that came to mind for US college students tested about 30 years later than they were to exemplars accessed by other Chinese students who responded in the same year in Chinese.

Because each participant in the Jeng et al. study only responded in one language, there is no way to know for sure whether or not the same individuals would list different items in each of their languages. The fact that the participants were sampled from the same population and yet responded differently in Chinese than in English suggests that translation equivalents of category labels are associated with differing sets of exemplars in each of their two languages, but without having the same individuals respond in both languages, it is not possible to firmly establish that a difference exists. Indeed a complete design requires that some participants respond twice in the same language and some respond once in each of their two languages on separate occasions. This is because output dominance is dynamic and changeable, at least within limits, even when monolinguals respond on two separate occasions in their one language. For example, Bellezza (1984) had monolingual English speakers list exemplars of several categories in two sessions held one week apart and found that correlations between the output dominance of items was strong and positive, though not perfect, and averaged in the .70 range (e.g., Bellezza, 1984). Thus, in general, items that are highly accessible remain so, but there is also considerable variation from one time to the next in the relative dominance of items.

In the present study we examined the category exemplar listing responses of Chinese-English bilinguals in two separate testing sessions, and had some of them respond twice in the same language and other respond once in each language.

Method

Participants

Participants were 36 Chinese-English bilinguals (17 female and 19 male) recruited from social and community groups in the Tuscaloosa area. The majority were students from Mainland China attending the University of Alabama, and the remainder were family members of those students. Participants ranged in age from 20 to 57, with a mean age of 31 years, 1 month.

Materials

All participants listed exemplars for the 10 categories of ANIMALS, SPORTS, VEGETABLES, OCCUPATIONS, TYPES OF MUSIC, FRUIT, SUBJECTS STUDIED IN SCHOOL, RELATIVES, BREAKFAST FOODS, and ENTERTAINMENT in each of two sessions approximately one week apart. Participants were equally divided across four conditions formed by crossing the variables of language used in Session 1 (Chinese or English) and

whether the language used in Session 2 was the same or different from that used in Session 1. For ease of reference, these groups are designated, with respect to the language used in the first and second sessions as Chinese-Chinese (CC), English-English (EE), Chinese-English (CE), and English-Chinese (EC). As nearly as possible, the groups were equivalent in terms of the numbers of males and females and age distribution.

Task instructions and category labels were always in the language being used for listing exemplars, and testing was done by a Mass Communication graduate student who is fluent in Chinese and English. Instructions and category labels were translated from English to Chinese by two fluent Chinese-English bilinguals. For each session, participants were given sheets of paper with category labels and twelve lines beneath each label. Different random orderings of the categories were used so that no particular category would always occur either late or early in the sequence. One constraint on the random order was that the food items (FRUIT, VEGETABLES, and BREAKFAST FOODS) did not immediately precede or follow one another. Participants were allowed to respond at their own pace. Although participants were not told explicitly that they should list twelve items for each category, in practice they typically did so.

Results

Of most interest was the extent of overlap in the specific exemplars listed in the two sessions. The simplest measure of overlap is the number of items listed in common across the two sessions. As shown in Table 1, for all 10 categories, the mean number of overlapping items was greater for individuals who listed in the same language on both occasions than for those who listed in a different language on each of the two occasions. A multivariate analysis of variance (MANOVA) confirmed that the overall effect of language (same versus different) was significant, $F = 2.54$, $p < .05$, and that the differences in overlap were individually significant for the categories of SPORTS, SUBJECTS STUDIED IN SCHOOL, VEGETABLES, TYPES OF MUSIC, BREAKFAST FOODS, and RELATIVES, and marginally significant for OCCUPATIONS.

Table 1: Mean exemplar overlap by condition.

Category	Same	Different	*p*
Sports	6.9	5.2	.011
Fruit	7.1	5.9	.111
Occupation	4.8	3.5	.055
Animals	6.9	5.7	.154
Subjects	7.5	4.9	.002
Entertainment	5.0	4.5	.512
Vegetables	6.3	4.0	.008
Music	5.6	3.9	.050
Breakfast	5.9	4.2	.019
Relatives	7.8	5.7	.012

One concern in using raw overlap as a measure is that it is influenced by the total number of items listed. The more total items listed across the sessions, the more possible overlapping items there are. To adjust for the number of items listed in the sessions, we followed an analysis procedure used by Bellezza (1984) in a study on the reliability of retrieval from semantic memory. Specifically, we computed *common element correlations*, which are defined as the number of exemplars listed in common across the sessions divided by the square root of the total number of exemplars from Session1 times the total from Session 2 (the geometric mean). The measure can range from 1.00 when the exact same exemplars are listed in both sessions to 0 in the case of completely different exemplars across the sessions. The same-language conditions (CC and EE) provide an essential baseline measure of reliability of retrieval from each of the categories when individuals access information in a single language (see e.g., Bellezza, 1984). In other words, because the way in which a category is constructed or instantiated on two different occasions will not be identical, even when an individual accesses information from that category in the same language, common element correlations will not reveal perfect overlap even in the CC and EE conditions. The extent to which CE and EC overlap scores are even lower can then be taken as an indicator of differences in the way category information is represented and accessed in a bilingual individual's two languages beyond that baseline limit on reliability.

As shown in Table 2, for all 10 categories, the mean common element correlations for items listed in Session 1 and Session 2 were smaller for groups that responded in different languages (CE and EC) than for groups that responded in the same language (CC and EE). Using multivariate analysis of variance (MANOVA), it was found that the mean differences between the common element correlations were statistically significant overall and significant individually for 7 of the 10 categories (SPORTS, FRUIT, OCCUPATIONS, SUBJECTS STUDIED IN SCHOOL, VEGETABLES, BREAKFAST FOODS, and RELATIVES), and marginally so for another (TYPES OF MUSIC). Thus the response patterns point to a divergence in the exemplars most associated with corresponding categories in Chinese and English across a range of concept types.

Table 2: Common Element Correlations by condition.

Category	Same	Different	*p*
Sports	.63	.49	.004
Fruit	.68	.58	.010
Occupation	.43	.31	.020
Animals	.61	.50	.128
Subjects	.65	.44	.002
Entertainment	.49	.41	.192
Vegetables	.58	.42	.002
Music	.60	.46	.067
Breakfast	.58	.42	.012
Relatives	.70	.51	.003

Neither the analysis on raw overlap scores nor that on common element correlations revealed an effect of starting language or an interaction of starting language and same versus different language. In effect these participants showed as much overlap in the category exemplars they listed when they responded twice in English as they did when they responded twice in Chinese, and they showed as much divergence in shifting from Chinese to English as they did in shifting from English to Chinese.

Discussion

The present findings are consistent with the idea that exemplars of common categories are differentially representative of those categories when accessed via bilingual individuals' two languages. Overlapping, but nevertheless divergent sets of exemplars come to mind when bilingual individuals are prompted with translation-equivalent category labels in their two languages.

It is important to note that the divergence found across languages was reliably more than that found when individuals responded in the same language on two occasions. Same-language common element correlations provide a kind of test-retest reliability and an indicator of the stability of category structure within a language, whereas cross-language correlations are influenced by both the reliability of the measure and differences in dominance structure in the languages. That is, the cross language divergence reflects more than just the limited test-retest reliability of responding in a category exemplar listing task.

Hierarchical models of bilingual language representation (e.g., Kroll & Stewart, 1994; Kroll & Tokowicz, in press; Potter, So, Eckardt, & Feldman, 1984) typically distinguish between two levels of representation, a lexical level at which the words of the two languages are represented in distinct lexicons and a conceptual level concerned with the representations associated with the bilingual's two lexicons (see e.g., Francis, in press; Heredia & Brown, 2004, Kroll & Stewart, 1994). Attention has focused on the links between corresponding items in the separate lexicons and on whether access to meaning from L2 is direct or mediated through L1. Although the distinction between the lexical and conceptual level has been an extraordinarily useful one, these models do not directly address the form or structure of underlying conceptual representations and address bilingualism largely at the single word level (but see Heredia & Altarriba, 2001). Other bilingual memory models, however, such as the distributed model (e.g., de Groot, 1992; 1993; 2002), and the tripartite model (e.g., Paradis, 1979) may be more easily amenable to a more nuanced account of bilingual conceptual representations, and may help in interpreting the present findings.

Paradis' (1979) model builds on a proposal by Kolers (1968) that bilinguals either store all information centrally in one information store and have access to it equally or store it in two separate conceptual stores, one associated with each language. Paradis suggested that there may be a single conceptual store but that languages differ in how they organize experience and thus differentially access the common conceptual-experiential store.

In de Groot's distributed model, translation equivalents from each of the bilingual's two languages are associated with semantic elements, and different types of concepts may have relatively more or less overlap in the associated elements (de Groot, 1992; de Groot et al., 1994). de Groot has argued that some concepts (e.g., concrete words) have more overlapping characteristics across languages, and others (e.g., abstract words) have less overlap across languages. According to this framework, abstract words that are translation equivalents from two languages would be especially likely to be associated with diverging, though partially overlapping semantic elements. For example, because LOVE in English can be applied to animate as well as inanimate entities, whereas AMOR in Spanish is reserved for people those words in a bilingual's English and Spanish lexicons will be linked to partially overlapping, but not identical sets of semantic elements. Those differences then might be expected to manifest themselves in a variety of conceptual tasks, such as the category exemplar listing task used in the present study. For example, rubbing a pet dog's ears, caressing a favored security blanket or rereading a special poem for the nth time could conceivably be listed as instances of things that exemplify LOVE, but they would not be listed as exemplars of AMOR. Thus, the conceptual representations associated with the otherwise equivalent LOVE and AMOR would be different, and that difference would be revealed in traditional tasks that assess the nature of conceptual structures.

Although divergence in the associated meanings of abstract words is to be expected in de Groot's model, the present findings reveal divergence even for categories that have concrete, tangible exemplars, such as fruit and vegetables. Our results are also consistent with evidence from other types of conceptual tasks for divergence in the meanings of words that refer to concrete objects. Malt and Sloman (2003), for example, showed that adult, nonnative English speakers, who were in various stages of learning English, diverged from native English speakers in the way they applied labels, such as "bottle" and "dish" to a set of presented stimuli, and in their judgments of the typicality of those stimuli within categories. With more experience, particularly years of immersion in an English speaking environment, their English naming patterns and typicality judgments shifted in the direction of those of native English speakers, but did not come to match native patterns exactly, even for individuals who had become quite proficient in English. More importantly for present purposes, the participants' naming patterns in English also differed from the naming patterns they used in their native languages, and the mismatches increased slightly, though not significantly, with increased immersion in an English-speaking environment. Participants' self reports of the strategies they used in determining the appropriate English term also suggested that, when responding in English, these

individuals did not simply try to retrieve the most reasonable label in their native language and then translate it into the English equivalent. The results are consistent with the idea that there are different mappings between words and meanings across the bilingual's languages. This type of finding can be explained by a distributed model such as the one proposed by de Groot (1992) if one extends the idea of non-overlapping elements to concrete concepts as well as abstract concepts, and by the Paradis model (1979).

Based on our study and other findings (e.g., Malt and Sloman; 2003; Pena et al. 2003; Roberts & LeDorze, 1997), it is clear that bilinguals show less overlap in their listings when they respond in their different languages than when they respond in the same language, even for relatively concrete types of concepts. However, it is not clear yet whether systematic studies using paradigms that assess conceptual structure will find *relatively more* cross-language divergence for abstract and social/cultural categories that may share fewer meaning elements across the words from two languages (see e.g., de Groot, 1992). Another interesting case would be goal derived or ad hoc categories (see e.g., Barsalou, 1985), because, in contrast to the other types of concepts that have pre-established structures they are constructed online in response to particular goals or needs. Because those constructions are driven by broader understandings and beliefs about the workings of the world, discrepancies in exemplar listings for those concepts may be expected to be strongly related to differences in the context of acquisition for the person's languages.

References

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 629-654.

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

Bellezza, F. S. (1984). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society*, *22*, 3324-326.

De Groot. A. M. B. (1992). Determinant of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1001-1018.

DeGroot, A. 1993. Word type effects in bilingual processing tasks: Support for a mixed-representational system. In Schreuder, R. and Weltens, B. (Eds.), 1993. *The bilingual lexicon*. Amsterdam: John Benjamins.

De Groot, A.M.B. (2002). Lexical representation and lexical processing in the L2 user. In V. Cook (Ed.), *Portraits of the L2 user*, (pp. 32-63). Clevedon: Multilingual Matters.

De Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language, 33,* 600-629.

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

Francis, W. S. (in press). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. de Groot (Eds.). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science, 10,* 164-168.).

Heridia, R. R., & Brown, J. M. (2004). Bilingual memory. In T.K. Bhatia and W. Ritchie (Eds.), *The Handbook of Bilingualism*. Oxford: Blackwell.

Jeng, Chung-In, Lai, Mei-Wei, & Liu, In-Mao (1973). Category norms in Chinese and English from bilingual subjects. *Acta Psychologica Taiwanica, 15*, 81-153.

Kolers, P. 1968. Bilingualism and information-processing. *Scientific American*, *218*, 78-86.

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

Kroll, J. F., & Tokowicz, N. (in press). Models of bilingual representation and processing. In J. F. Kroll & A. M. B. de Groot (Eds.). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.

Malt, B. & Sloman, S. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition, 6(1),* 47-67.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*, 283-284.

Paradis, M. (1979). Language and thought in bilinguals. In W.McCormack & H. Izzo (Eds.), *The Sixth LACUS Forum* (pp. 420-431). Columbia, SC: Hornbeam Press.

Pena, E., Bedore, L., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language, and Hearing Research, 45*, 938-947.

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

Roberts, P. & LeDorze, G. (1997). Semantic originality, strategy use and productivity in bilingual semantic verbal fluency. *Brain and Language, 59,* 412-449.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192-232.