Draft – Do Not Quote Without Permission

Do Numeral Classifiers Influence Similarity Judgments?

Jessica Rett and Marilyn Shatz

INTRODUCTION

People classify things. That is, they treat things as members of classes or categories. Explaining the process of categorization involves explaining why we use the categories we do, and how we assign members to these categories. Early psychological work on categorization adopted the classical view of categories in which members of a category share certain criterial features, that is, the necessary and sufficient features an entity must have as a member of that category. Sets of criterial, or defining, features, then, are attributed to a person's mental representations of categories, and entities possessing the requisite features can be assigned to categories appropriately (Katz & Fodor 1963). Numerous arguments have raised serious questions about the value of characterizing categories in terms of criterial features (e.g., Putnam 1962, Quine 1951).

Alternative theories regarding the nature of categories propose similarity as the primary factor in determining category membership. These "resemblance theories" include network models, prototype models, frame models and Bayesian models (Brooks 1978, Medin & Schaffer 1978, Rosch & Mervis 1975). Although different approaches credit similarity with different degrees of responsibility for category formation, they all affirm that a person's decision about an item's category membership is less dependent on a set of criterial features than on how much an item resembles other items in the putative category.

However, there are many ways in which entities can resemble one another. For example, they can be judged as similar when they have similar observable properties such as shape, color and texture; they can also share internal properties such as having hearts or wires. For the purposes of this paper, we will adapt the terminology outlined in Gelman, Hollander, Star and Heyman (2000) with regard to categories and kinds: "Whereas a category is any grouping together of two or more discriminably different things, a kind (...) is a category that is treated by those who use it as being based in nature, discovered rather than invented, and capturing many deep regularities (pg. 204)." An example of a category, then, is "striped things," while an example of a kind is "tigers."

In assigning labels to items, languages do not treat all properties the same; items that share the same label are taken by speakers to mean that they have some essential properties in common (see Gelman *in press*). Some languages have additional linguistic means for noting similarities among entities. In particular, some languages group items together by marking them with the same numeral classifier. These markings typically allow for the grouping of items not by essential properties but by features likeshape (e.g., long and narrow) and size (e.g., size of body) (Downing 1996).

In this paper, we compare speakers of Japanese (a classifier language) with speakers of English (a language without clasifiers) to investigate the possibility that a person's knowledge of a language with numeral classifiers contributes to his/her judgments of the similarity of items. If speakers of Japanese judge items sharing a numeral classifier as more similar than do speakers of English, it would be evidence that classifier systems influence similarity judgments, and by inference, conceptual category assignment. We examine this question while addressing theoretical and methodological limitations in a previous classifier study. Our results provide the basis of a discussion of the factors implicated in similarity judgments, conceptual categories, and the nature of numeral classifier assignment.

The linguistic aspects of numeral classifiers

Numeral classifiers are morphemes attached to a numeral expression when an item is being counted. They occur in many non-Indo-European languages, including Japanese, Chinese, and Yucatec Mayan. Athough the types of categories organized by classifiers vary crosslinguistically (Aikhenvald 2000, Grinevald 2000), most classifier systems group together items similar in animacy, shape, or function (Uchida & Imai 1999). For instance, in Japanese, dogs and mice are counted using the classifier *hiki* (for small animals), both pencils and trees are counted using the classifier *hon* (for long, slender objects), and both cars and refrigerators are counted using the classifier *dai* (for machinery). Always attached to numbers, the classifiers do not convey numerical information but rather information about the counted items themselves. Thus, when referring to a specific number of objects in Japanese, a speaker must combine a number with one of a few hundred classifiers, the assignment of which reflects some attribute of the items being counted. However, there are no explicit, prescriptive rules governing either a speaker's choice of classifier or the linguist's description of classifier categories. Rather, linguists base their characterizations of a classifier category and its members on empirical work in which native speakers describe what sorts of entities take a given classifier (see Downing 1996).

There seems to be a strong covariation between classifier categories and ontological class membership. Uchida and Imai (1999) note, "Roughly speaking, the conceptual/semantic distinction between animals and inanimate entities is very strictly observed in the use of classifiers" (*p*. 51). This also is largely true for other semantic distinctions, such as those between one- two- and three-dimensional inanimate objects and between human and non-human animate objects. Nonetheless, although most classifier categories correlate with ontological class categories, classifier assignment is not fully predicted by semantic rules (Uchida & Imai 1999). There are some instances in which classifier assignment seems to contradict ontological class membership. Rabbits, for example, are classified with the classifier *wa* (a classifier that is used primarily to count birds) instead of with the small animal classifier *hiki* or the large animal classifier *too*. Instances like this, in which assignment to classifier category is not consistent with ontological class, should be especially revealing about the influence of classifier assignment on judgments of similarity.

There are several reasons why Japanese is an especially appropriate classifier language for us to compare to English. First, classifiers are common in Japanese; they are obligatory in all cases of counting objects, whereas they are optional in some languages[i], and they can also be used as definite articles. Example 1 illustrates the use of classifiers to count in Japanese, where the classifier <u>hon</u> (for long, slender items) is identified with CFR.

1. empitsu wa ni-hon aru

   pencil subject-particle 2-CFR exist

   There are two pencils.

Example 2 illustrates the use of a classifier as a definite article.

2. ippon empitsu

   1-CFR pencil

   a pencil

Second, although there is some dialectical variation, classifier assignment in Japanese is relatively standard for many entities, and variations on these assignments would largely be considered ungrammatical (Downing 1996, Denny 1986). This contrasts with some other classifier languages, such as Yucatec Mayan, which has only minimal restrictions on the assignment of classifiers, thus allowing speakers more freedom to vary the classifier they use with particular items.

Do classifiers have an effect on similarity judgments?

Schmitt and Zhang (1998) report that classifiers significantly affect similarity ratings and, by inference, conceptual organization and categorization. In their Study 1, groups of Mandarin speakers and English speakers provided pairwise comparisons for objects, some of which share the same classifier (e.g. *peanut* and *pill*, which share the Mandarin classifier *li* for tiny, grainlike objects, *p*.111). Subjects rated the pairs of objects on seven-point scales (1 = not at all similar, 7 = very similar). The ratings were compared crosslinguistically to determine whether and to what extent classifiers affected the subjects' judgments of similarity. The authors predicted that speakers of a classifier language (i.e., Mandarin Chinese) would rate objects that share a classifier as relatively more similar than objects that do not share a classifier (*p*. 109) and that this would *not* be the case for speakers of non-classifier languages (e.g. English).

The study yielded two interesting results. First, there was a main classifier effect: For *both* language groups, pairs sharing classifiers were rated as more similar than pairs of objects not sharing a classifier. Second, there was a significant interaction effect between the language and classifier variables: Mandarin-speaking subjects rated the classifier-sharing pairs as more different from the non-classifier-sharing ones than did the English speaking subjects. The interaction effect is in line with the authors' assertion that, if classifiers did indeed affect similarity judgments, the stimuli within different classifier variables would be treated significantly differently by the two language groups.

However, because they did not consider the covariation of ontological class and numeral classifier categories, their method of testing did not allow an analysis of the relative influence of classifiers and ontological class membership on similarity judgments. Yet, the semantic information conveyed in classifiers typically duplicates much of the information embodied in ontological classes (Uchida & Imai 1999). This suggests that classifiers themselves may contribute relatively little to similarity judgments beyond what ontological information can. Thus, a more adequate test for the influence of classifiers on similarity judgments requires an experimental design that recognizes the covariation between classifier categories and ontological class membership and separates those variables.[ii]

In our study, we cross two variables: 1) a classifier variable (same or different) analogous to Schmitt and Zhang 1998, and 2) a kind variable (same or different), signifying whether the two objects in the pair are of the same ontological class; that is, they share a second-order kind term in both of the testing languages (see Table 1). For example, in both Japanese and English, ducks and chickens are birds; apples and melons are fruits; rakes and shovels are tools.[iii] With the addition of the kind variable, we can separate the influence of classifiers from the influence of ontological class, and thereby focus on the few cases in which classifier category membership deviates from ontological class membership (as in the case of rabbits mentioned earlier). In so doing, we can clarify the role classifiers may play in similarity judgments, and by inference, conceptual categories. Study 1 reports on the judgments of 40 subjects (20 English speakers and

20 Japanese speakers), all of whom reside in the US. Study 2 reports on 15 additional subjects, Japanese speakers residing in Japan.

STUDY 1: JAPANESE SPEAKERS IN AMERICA

Method

Design

Study 1 had a 2 x 2 x 4 design. There were 2 between-subject variables, language of the subjects (English, Japanese) and type of question asked (similarity, kind). The degree to which subjects judge items as similar seems to be influenced by the version of the question asked (Diesendruck & Shatz 2001, Rips 1989). Therefore, we asked half of the subjects about similarity with the question, *How much is one item similar to the other?* and the other half the question, *How much is one item the same kind of thing as the other?* Subjects from each language group were randomly assigned to either the similarity or the kind question condition.

A within-subject condition variable included the 4 possible combinations of stimulus pairs in which items were the same or different on kind and classifier assignment, namely, SS, SD, DS, DD, where the first letter refers to the two items' kind memberships as same or different, and the second letter refers to the items' classifier assignments as same or different. Stimulus pairs in the SS condition are of the same kind and share the same classifier (e.g. dogs and mice are mammals and take the classifier *hiki,* for small animals). In the SD condition, stimulus pairs are of the same kind but take different classifiers (e.g. mice and rabbits are mammals but mice are counted with *hiki* while rabbits are counted with *wa*). Pairs in the DS condition are not of the same kind but share classifiers (e.g. snakes are <u>reptiles</u> and dogs are <u>mammals</u>, but all share the classifier *hiki*). Pairs in the DD condition are neither of the same kind nor do they share a classifier (e.g. rabbits are mammals and snakes are reptiles; rabbits are counted with *wa* and snakes are counted with *hiki*).

Subjects

Twenty native English speakers and 20 native Japanese speakers, all female[iv] and all residing in a Midwestern college town, provided judgments for 32 pairs of objects. English speakers were all between the ages of 18 and 25, were undergraduates at the university and were recruited from a large dormitory. All subjects had some experience with another language, but none reported it as their current primary language or that of their parents.

Japanese speakers were recruited from a weekly music class taught in Japanese for Japanese toddlers and their mothers. Fifteen of the women were between the ages of 26-35; five were over 36. All subjects had some experience with another language, but none reported it as their current primary language or that of their parents. All had resided in the United States, for 2 months to 10 years (*M* = 2;8 years). Fourteen of them had studied English in the United States, from 2 months to 2 years (*M* = 10.4 months), but only two subjects reported speaking English on a daily basis.

Stimuli

We conducted preliminary testing for standard classifier assignment with 7 different native Japanese speakers. After reading a brief explanation about what a classifier is, followed by an example of a classifier being used in a counting clause, they were requested to write the classifier they most frequently use when counting each item on a list of 120 words. From these 120 items, 92 words were selected according to the criterion that all 7 speakers had agreed on a classifier assignment. Pairs of 32 items were then selected to create 8 stimulus pairs for each of the 4 conditions. The pairs were constructed so that each word was used in two pairs in two conditions, once as the first word in the pair, once as the second. Because words could not be completely crossed, the conditions in which a given word appeared varied (e.g., *raccoon* is in SS and DS, while *refrigerator* is in SS and DD). All pairs of words were presented in the singular present tense without classifiers or determiners. See Appendix for a full list of the stimulus pairs by condition.

Procedure

Each language group was tested separately by a native speaker of the language. Each participant received a test booklet with the instructions, appropriate question, and stimuli written in her native language, with a horizontal scale numbered from one to seven printed after each word pair. The end points of the scale were marked in each language *not at all* and *very*, with the mid-point marked *somewhat.* The experimenters instructed the subjects to circle l and only l number for each pair. They urged subjects to take as much time as needed to do the task and that their answers would not be evaluated as right or wrong. Each subject rated all 32 pairs of words on the 7-point scales in response to either the similarity or the kind question.

All subjects filled out pre-test demographic questionnaires. Japanese subjects also filled out a post-test classifier assignment questionnaire. This questionnaire included a brief explanation of what a classifier is and provided an example of a classifier used in a counting clause. Subjects were instructed to write the classifier they typically used for each of the words in the test pairs.

Results

A preliminary 2 (language) x 2 (question: similarity or kind) repeated measures ANOVA revealed no main effect of question, ($F(1,38) = 1.82$, $p = .183$), but a significant language effect ($F(1,38) = 30.51$, $p < .000$), and an interaction between language and question ($F(3,36) = 7.76$, p

< .01). Because there was a main effect of language and an interaction effect, we chose to examine the influence of question type on the ratings of speakers of each language separately with two 2 (type of question: similarity or kind) x 4 (condition: SS, SD, DS, DD) ANOVAs. For English speakers, there was no significant effect of question or of interaction between question and condition, but only a significant condition effect, $F(3,16) = 91.79$, $p < .000$. For Japanese + speakers, there was no significant interaction effect, but there were main effects of condition ($F(3,16) = 22.64$, $p < .000$) and question ($F(1, 18) = 21.37$, $p = .021$), with the "same kind of thing" question rated higher on average ($M = 2.98$) than the similarity question ($M = 1.94$). Nonetheless, because question type was essentially a control variable whereas conditon was of primary interest, and because there were no significant interaction effects between question type and condition for either language group, we combined across question types in the further analyses investigating condition effects on the two groups of speakers.

A 2 (language) x 4 (condition: SS, SD, DS, DD) repeated measures ANOVA revealed a significant language effect, $F(1,38) = 25.57$, $p < .000$, and a significant condition effect $F(3,36) = 73.34$, $p < .000$. Overall, the SD (same kind, different classifier) condition received the highest similarity ratings ($M = 4.27$), SS (same kind, same classifier) the next ($M = 3.40$), then DS (different kind, same classifier) ($M = 2.74$) and finally DD (different kind, different classifier) ($M = 2.21$). Table 2 illustrates that each language group ranked the four conditions in the same order; they rated pairs with same kind information higher than they did pairs with different kind information. However, English speakers gave pairs in each of the four conditions higher similarity ratings (overall $M = 3.84$) than did the Japanese speakers (overall $M = 2.46$). Table 2 also reveals that English speakers varied their ratings more than did the Japanese speakers. The rating differences between pairs of conditions for English speakers ranged from 2.77 to .37 with a mean of 1.60, compared to a range of 1.35 to .03 and a mean of .68 for Japanese speakers. The main effect of language, then, seems largely to be a consequence of an overall difference in the way the two language groups used the scale's range. Nonetheless, the main effects need to be interpreted further in light of a language by condition interaction, $F(3,36) = 15.41$, $p < .000$.

We used Kremer-Tukey post-hoc tests to explore the condition and interaction effects. Table 3 presents the *p*-values associated with a subset of these tests, namely, the differences in the means of the six-pair-wise comparisons for the four conditions for each of two language groups. Table 3 reveals that for both groups, participants rated pairs significantly higher when both kind and classifier information were the same than when they were both different (Column 1), and they rated pairs significantly higher when only kind information was the same compared to when only classifier information was the same (Column 2). English speakers rated pairs significantly higher when both types of information were the same, compared to only same classifier information (Column 3), but Japanese speakers did not. Both groups of speakers rated pairs significantly higher when only kind information was the same (Column 4). English speakers rated pairs in which only kind was the same significantly more highly compared to same for both kind and classifier, whereas Japanese speakers showed only a trend in this direction (See Column 5 and Table 2). Japanese speakers but not English speakers rated pairs taking same classifiers but not sharing kind categories as significantly higher than pairs differing on both (Column 6). However, English and Japanese speakers did not differ significantly in their ratings for DS pairs ($p > .6$), nor did they differ significantly in their ratings for DD pairs ($p > .1$), whereas they differed significantly in their similarity ratings for SS and SD pairs ($p < .001$ for both conditions.)

In sum, the post-hoc comparisons support the conclusion that kind accounts for similarity ratings far more than classifiers do, regardless of the language raters speak: On 7 of 8 within-language comparisons, conditions with pairs having kind in common were rated significantly higher than were those that did not. Importantly, this was so even when classifier information was available (see Column 2). However, there were three within-language findings suggesting that Japanese speakers may have been influenced somewhat by shared classifiers: When there were kind differences in both conditions, commonality of classifier produced significantly higher ratings in the DS than the DD condition by Japanese speakers but not English speakers (Column 6); moreover, the pairs of the SD condition engendered significantly higher ratings by the English speakers than did those in the SS condition, but the differences in the ratings by Japanese speakers for those 2 conditions did not reach significance (Column 5). Nor did the SS pairs produce significantly higher ratings by the Japanese speakers than DS pairs (Column 3).

To assess the influences of same kinds and classifiers over different kinds and classifiers on each of the two groups, we averaged the conditions in which pairs shared kind (SS and SD) into a *same kind* condition (SK) and the conditions in which they differed (DS and DD) into a *different kind* condition (DK); analogously, we averaged SS and DS into a *same classifier* condition (SC) and SD and DD into a *different classifier* condition (DC). With these new conditions, we conducted four paired t-tests, two for each language group, assessing in one whether the ratings differed significantly for same versus different kind pairs and in the other whether the ratings so differed for same versus different classifier pairs.

The findings support the results of the post-hoc comparisons in suggesting a major influence of kind. For the English speakers, same kind (M = 4.86) differed significantly from different kind (M = 2.83), t (1,19) = 17.71, $p < .000$, two-tailed), and same classifier (M= 7.32) differed from different classifier (*M* = 8.07), t(1,19) = -4.63, *p* < .000, two-tailed. For the Japanese speakers, same kind (*M* = 2.81) also differed significantly from different kind (*M* = 2.12), t(1,19) = 2.64, *p* = .016, two-tailed), but same classifier (*M* = 2.47) did not differ significantly from different classifier (*M* = 2.45), t(1,19) = .214, *p* = .83, two-tailed. Thus, both groups judged shared kind as highly important in assessing similarity. Overall, English speakers found the pairs with different classifiers significantly **more** similar than those with shared classifiers whereas there was virtually no difference between the two groups of pairs for Japanese speakers.

Discussion

Study 1 provided some provocative results about kind and classifier influences on similarity judgments in speakers of classifier and non-classifier languages. First, the results indicate that shared kind -- when two entities are members of the same ontological class -- plays a major role in similarity judgments for speakers of both sorts of languages. Second, there are some suggestions of an additional minor influence of classifiers on Japanese speakers: when kind was not shared, shared classifier pairs (DS) resulted in significantly higher similarity ratings than non-shared classifier pairs (DD); same kind did not result in significantly higher ratings over different kind when classifiers were shared (SS vs. DS); and overall, Japanese different classifier pair ratings (DC) were not significantly different from same classisfier pair ratings (SC). In contrast, English speakers rated SS pairs significantly higher than DS pairs and the DC pairs significantly higher than the SC pairs. However, the findings of non-significance for Japanese ratings should be interpreted with caution. Although, the data could be interpreted as showing that influence of kind similarity of particular pairs was sometimes mitigated in Japanese speakers by classifier information, it could also be that the lack of significance in the differences in Japanese ratings is a consequence of their compression of the rating scale.

Recall that there was a major language effect which we argued was due to differences in scale use. In particular, English speakers produced significantly higher ratings on same kind pairs than did Japanese speakers, although the latter still showed a preference for same kind information over same classifier information in rating pairs. It is possible that our Japanese speakers were unusual in the degree to which they compressed the rating scale. As wives of graduate students living in a foreign land, they may have been more subdued in their performance than Japanese speakers in Japan. To investigate this possibility, we collected data on Japanese speakers living in Japan for a second study comparing them to our original participants.

Recall too that Japanese speakers in the US who had been asked the same-kind-of thing question had given higher ratings overall than had those asked the similarity question. Although the question asked had not interacted with stimulus conditions in Study 1, we tried nonetheless in Study 2 to maximize the likelihood that Japanese speakers in Japan would attend not just to kind commonalities but to classifier commonalities as well by asking all the speakers in Japan only the similarity question. Thus, Study 2 should allow us to interpret more confidently the findings suggesting possible classifier influences on the similarity judgments of Japanese speakers.

STUDY 2: JAPANESE SPEAKERS IN JAPAN

Method

Subjects

Fifteen Japanese speakers, 4 female and 11 male, were drawn from a university in a small city near Tokyo, Japan. All were between the ages of 18-25, and all identified Japanese as their primary language and their parents' primary language. Every subject had studied English in Japan from 7 to 11 years, (*M* = 7.66 years). Only one subject reported knowing a language other than Japanese and English, and that language was Chinese. These speakers were compared to the 10 Japanese speakers living in the U.S. tested in Study 1 with the similarity question.

Stimuli and procedure

The stimuli and procedure in this study are the same as those in Study 1, with one exception: all subjects were tested with the question, *How similar is one item to the other?*

Results

We conducted one 2 (location of speakers: U.S. or Japan) x 4 (condition) repeated measures ANOVA for uneven samples, comparing the Japanese speakers in the U.S. to the Japanese speakers in Japan. There was a significant effect of condition, $F(3, 21) = 26.95$, $p \leq .000$, but no effect of location, $p = .124$. Overall, Japanese speakers in the U.S. had a mean similarity rating of 1.92, and speakers in Japan had a mean rating of 2.47. However, the interaction effect between location and condition met the conventional level of significance, $F(3, 21) = 3.10$, $p = .049$. Table 4 lists the means by conditions for the two sets of Japanese speakers by location. As can be seen there, regardless of location, Japanese speakers who heard the similarity question ranked the conditions in the same order as all the English and Japanese speakers in the U.S. had done: SD, SS, DS, and DD. Table 4 also reveals that Japanese speakers in Japan varied their ratings more than the subset of Japanese speakers in the U.S. The mean rating difference for the former was 1.37, as compared to a mean of .66 for the latter. Moreover, the mean variation in similarity judgments made by Japanese speakers in Japan fell closer to that of English speakers (*M* = 1.6) than that of Japanese speakers in the U.S. (*M* = .68 for all U.S. Japanese speakers.) Thus, although the location effect was not significant, the means and variation in ratings produced by the Japanese speakers in Japan suggests that they compressed the scale somewhat less than their U.S. counterparts.

To explore the condition and interaction effects, we conducted a series of Kremer-Tukey post-hocs. There were no significant differences between the groups of speakers on any of the 4 conditions; but an examination of Table 5 reveals that Japanese speakers in Japan had more significant comparisons involving same kind pairs than did Japanese speakers in the U.S. answering the similarity question. The former rated same kind pairs significantly higher than different kind pairs regardless of classifier information, whereas the only comparison to reach significance for the former involved SD-DD pairs. The lack of significance in the comparisons of the US Japanese ratings is likely due both to the small number of speakers hearing the similarity question (N = 10) and the more compressed rating scale they used.

We again assessed the overall influences of kind and classifier by first combining same kind pairs (SS and SD) into one condition (SK) and different kind pairs (DS and DD) into another condition (DK), making the analogous combinations for classifier (SS and DS into SC, SD and DD into DC), and by then conducting two 2-way (location of Japanese speaker: Japan-U.S.) x 2 (condition: same-different) ANOVAs, one for kind and one for similarity, on the Japanese subjects who responded to the similarity question. The kind analysis revealed a significant effect of condition, $F(1,23) = 48.73$, $p \leq .000$, and a significant condition-location interaction, $F(1,23) = 9.52$, $p = .005$. All speakers rated SK pairs highly (*M* = 3.02) compared to DK pairs (*M* = 1.58), ($p = .035$). Japanese speakers in Japan rated SK pairs more highly (*M* = 3.48) than did their U.S. counterparts (*M* = 2.32) but did not differ from them on DK ratings (*M*

= 1.57). The influence of location was not significant, $p > .1$. There were no significant effects associated with the classifier analysis.

Discussion

Our comparison of Japanese speakers in the U.S. and Japan on the similarity question has revealed that the U.S. sample of Japanese speakers was quite subdued in the use of the rating scale compared to the Japanese speakers in Japan, who rated same kind pairs higher than did their U.S. counterparts and whose range of ratings more closely resembled English speakers than Japanese speakers in the U.S. Importantly, the Japanese speakers in Japan produced the same ranking of stimulus pairs as had been found with the U.S. groups, rating SD highest, then SS and DS, down to DD.

A comparison of the English speaker and U.S. Japanese speaker condition-comparisons (lines 2 and 1 respectively, Table 2) to the Japanese speaker in Japan condition-comparisons (line 2, Table 5) reveals that 2 of the 3 possible classifier influences on Japanese speakers may be unstable. Japanese speakers in Japan, like English speakers but unlike U.S. Japanese speakers, rated SS pairs significantly higher than DS pairs (Column 3 of Table 5) and did not rate DS pairs significantly higher than DD pairs. However, like U.S. Japanese speakers, their ratings on SD pairs were higher but not significantly so than on SS pairs (Column 5 of Table 5). Importantly, no groups of Japanese subjects ranked same classifier pairs (SC) significantly higher than different classifier pairs (DC).

In sum, the results of Study 2 strengthen our finding of a major effect of kind commonality on similarity judgments by speakers of both classifier and non-classifier languages. They also clarify the possibility of influences of classifier commonality on speakers of a classifier language. When Japanese speakers use the rating scale more fully by expanding their range of ratings, their ratings show relatively little influence of classifier, and then largely as a minor mitigator of kind information.

GENERAL DISCUSSION

The findings from our studies on speakers of English (a non-classifier language) and Japanese (a classifier language) add important information for interpreting prior research findings on the influence of classifiers on similarity judgments. Schmitt and Zhang (1998) did not include the kind variable that we included when testing for the effect of classifiers and so could not disentangle the influence of kind from classifier**.** They found a main effect of classifier for speakers of both classifier and non-classifier languages: Both Mandarin and English speakers rated pairs sharing classifiers as different from pairs not sharing them. We can now see that this was very likely because many of their same-classifier pairs apparently shared kind as well. Indeed, when kind is separated from classifier, one finds no main effect of classifier for speakers of English or of Japanese, but a main effect of kind. Moreover, the rather unexpected higher ranking of the mean ratings for SD over SS pairs in our study suggests that many of our speakers, regardless of language spoken, often found the particular kind commonalities in SD pairs more compelling than pairs that shared both kind and classifier.

Schmitt and Zhang (1998) had also found an interaction between language and classifier: Mandarin speakers rated the shared classifier pairs even more differently than did the English speakers. Again, their finding is at odds with our findings on the SC versus DC analyses.

In our studies, English speakers rated these conditions significantly differently (rating DC more similar than SC pairs), whereas there was no significant difference between SC and DC pairs for Japanese speakers either in Japan or in the U.S. Since the SC and DC pairs both include same and different kind pairs, it is hard to argue that classifier commonality per se is a major factor in similarity judgments.

There are two possibilities that might help resolve some of the differences between the Schmitt and Zhang findings and ours. One is that Mandarin speakers may be influenced differently from Japanese speakers by classifiers. There may be some evidence for this in Study 4 of Schmitt and Zhang (1998): Although both Chinese and Japanese speakers took account of shared classifiers in their similarity judgments, the breadth of the language-specific classifiers affected the two groups of speakers differently. Again in this study, however, kind and classifier are conflated.

Chinese speakers may also differ from Japanese speakers with regard to the dialectical variability they show in making classifier assignments. Despite our controlling for stimulus classifier assignment with preliminary testing, we discovered, based on the post-test questionnaire, that there was still some variability in Japanese speakers' spontaneous classifier assignments to some of the test stimuli. All or most of the 20 U.S. Japanese subjects chose the expected classifier for more than 80% of the test words. There were five words for which more than half of the subjects chose unexpected classifiers. All or most of 15 Study 2 subjects in Japan also chose the expected classifier for more than 80% of the words. There were six words for which more than half of the subjects chose unexpected classifiers. Given previous findings on the subjectivity of classifier assignment even in the most standard of systems (see Downing, 1996; Denny, 1986), it is unlikely that a subject sample representing diverse dialectical regions would agree completely on classifier assignment for almost three dozen words. This is very likely true for Mandarin speakers as well as Japanese speakers. Further explorations of how best to control for classifier assignment should be made, but we note that neither Schmitt and Zhang

(1998) nor other research on the influence of numeral classifiers (e.g., Gao 1999, Lucy & Gaskins 2001) includes any attempt to control for variation in classifier assignment.

The other (more interesting, we think) possibility for resolving the differences is that particular pairs vary with regard to the relative importance of classifier; for some pairs, kind information may be paramount, and for other pairs, less so, thereby allowing for some influence of other pair characteristics such as classifier. Although an examination of our data did not reveal particular pairs that seemed to be "kind" outliers, the fact that the SD pairs resulted in higher mean similarity ratings than SS pairs suggests that even pairs sharing kind can differ among themselves with regard to how similar any given pair seems based on kind information. We recognize the imprecise nature of our kind variable. As mentioned in footnote 2, there is no clarity on how to specify level of shared ontological class. We used a rather ad hoc method of selecting pairs with second-order kind terms common in both languages of testing. Future research should address how best to specify the kind variable and whether different construals of kind influence similarity judgments.

These considerations not withstanding, our findings illustrate clearly that kind is the major influence on similarity judgments compared to classifier for both speakers of a classifier language (Japanese) and a non-classifier language (English). Future research should address whether this finding can be replicated using our design with speakers of other classifier languages. Assuming our finding can be extended to other languages, it should give pause to researchers who use classifier languages to argue for a major Whorfian influence of language on cognition. Although we have not directly assessed differences in conceptual organization between speakers of classifier and non-classifier languages, we have identified common kind as a major factor in the similarity judgments of speakers of both sorts of languages and have shown that when classifier commonality is separate from kind commonality, it has little effect on such judgments. To the extent that the classifiers in a language's classifer system correlate with ontological kinds, that system provides no new semantic information to conceptual categorization. Therefore, unlike Schmitt and Zhang (1998), who argue for linguistic relativity based on their claims for a classifier effect, we believe classifiers offer a relatively weak basis on which to assume that speakers of different languages have substantively different conceptual categories from speakers of non-classifier languages.

References.

Aikhenvald, A. (2000). Classifiers: A Typology of Noun Categorization Devices. New York. Oxford University Press.

Brooks, L.R. (1978). Non-analytic Concept Formation and Memory for Instances. In E. Rosch & B.B. Lloyd (eds.) Cognition and Categorization (pp. 169-211). Hillsdale, NJ: Erlbaum.

Denny, J. P. (1986). The Semantic Role of Noun Classifiers. In C. Colette (ed.) Noun Classes and Categorization. Amsterdam & Philadelphia: John Benjamins.

Diesendruck, G. & Shatz, M. (2001). Two-Year-Olds' Recognition of Hierarchies: Evidence from their interpretation of the semantic relation between object labels. Cognitive Development, 16(1), 577-594.

Downing, P. (1996). Numeral Classifier Systems: The case of Japanese. Philadelphia. John Benjamins Publishing Company.

Gao, Y. (1999). Mental Representation of Chinese Classifiers. Dissertation Abstracts International: Section B: The Sciences & Engineering 59:12-B.

Gelman, S. A. (in press). The essential child: Origins of essentialism in everyday thought. New York. Oxford University Press.

Gelman, S., Hollander, Star & Heyman (2000). The Role of Language in the Construction of Kind. Psychology of Learning and Motivation 39, 201-263.

Gleitman, L. (1990). The Structural Sources of Verb Meanings. Language Acquisition 1(1), 3-55.

Grinevald, C. (2000). A Morphosyntactic Typology of Classifiers. In G. Senft (ed.) Systems of Nominal Classification. New York: Cambridge University Press.

Katz, J. J. & Fodor, J. (1963). The Structure of a Semantic Theory. Language 39, 190-210.

Lucy, J. & Gaskins, S. (2001). Grammatical Categories and the Development of Classification Preferences: A comparative approach. In *M*. Bowerman and S. Levinson (eds.) Language Acquisition and Conceptual Development. Cambridge: Cambridge University Press.

Medin, D. and Schaffer, *M*. (1978). Context Theory of Classification Learning. Psychological Review 85(3), 207-238.

Putnam, H. (1962). It Ain't Necessarily So. Journal of Philosophy 59, 658-670.

Quine, W. V. (1951). Two Dogmas of Empiricism. Philosophical Review 60, 20-43.

Rips, L. (1989). Similarity, Typicality, and Categorization. S. Vosniadou and A. Ortony (eds.) Similarity and Analogical Reasoning. Cambridge University Press.

Rosch, E. & Mervis, C. B. (1975). Family Resemblances: Studies in the internal structure of categories. Cognitive Psychology 7(4), 573-605.

Schmitt, B. H. & Zhang, S. (1998). Language Structure and Categorization: A study of classifiers in consumer cognition, judgment and choice. Journal of Consumer Research 25(2), 108-122.

Uchida, N. & Imai, *M*. (1999). Heuristics in Learning Classifiers: The acquisition of the classifier system and its implications for the nature of lexical acquisition. Japanese Psychological Research 41(1), 50-69.

Table 1

Stimulus Conditions, Study 1

		CLASSIFIER	
		same	different
KIND TER	same	SS	SD
	different	DS	DD

Table 2

Mean Ratings by Condition and Language Group, Study 1

	SS	SD	DS	DD
English	4.30	5.42	3.02	2.65
U.S. Japanese	2.49	3.12	2.46	1.77

Table 3.

Condition Comparison Probabilities by Language Group, Study 1

	1	*2*	*3*	*4*	*5*	*6*
	SS*DD	SD*DS	SS*DS	SD*DD	SS*SD	DS*DD
English	.000	.000	.000	.000	.000	.638
U.S. Japanese	.018	.042	1.00	.000	.059	.026

Table 4.

Mean Ratings by Condition, Similarity Question, Study 2

	SS	SD	DS	DD
U.S. Japanese (10)	2.03	2.59	1.81	1.34
Japan Japanese (15)	3.24	3.72	1.61	1.53

Table 5.

Condition Comparison Probabilities by Language Group, Study 2

	1	*2*	*3*	*4*	*5*	*6*
	SS*DD	SD*DS	SS*DS	SD*DD	SS*SD	DS*DD
U.S. Japanese (10)	.494	.608	1.00	.007	.657	.545
Japan Japanese (15)	.002	.001	.003	.000	.948	1.00

Appendix: Stimuli by Condition

Same kind term - Same classifier

1	raccoon - whale	**mammal - mammal**	*hiki - hiki*
2	dog - mouse	**mammal - mammal**	*hiki - hiki*
3	squirrel - cat	**mammal - mammal**	*hiki - hiki*
4	melon - apple	**fruit - fruit**	*ko - ko*
5	onion - potato	**root veg - root veg**	*ko - ko*
6	car - bicycle	**vehicle - vehicle**	*dai - dai*
7	rake - knife	**tool - tool**	*hon - hon*
8	microwave - refrigerator	**appliance - appliance**	*dai - dai*

Same kind term - Different classifier

9	duck - chicken	**bird - bird**	*hiki - wa*
10	hawk - seagull	**bird - bird**	*wa - hiki*
11	mouse - rabbit	**mammal - mammal**	*hiki - wa*
12	radish - turnip	**root veg - root veg**	*hon - ko*
13	bush - fern	**plant - plant**	*hon - mai*
14	shovel - rake	**tool - tool**	*ko - hon*
15	hammer - screwdriver	**tool - tool**	*hon - ko*
16	dryer - stove	**appliance - appliance**	*ko - dai*

Different kind term - Same classifier

17	seagull - raccoon	**bird - mammal**	*hiki - hiki*
18	whale - duck	**mammal - bird**	*hiki - hiki*
19	snake - dog	**reptile - mammal**	*hiki - hiki*
20	apple - onion	**fruit - root vegetable**	*ko - ko*
21	potato - melon	**root vegetable - fruit**	*ko - ko*
22	stove - car	**appliance - vehicle**	*dai - dai*
23	bus - microwave	**vehicle - appliance**	*dai - dai*
24	screwdriver - dryer	**tool - appliance**	*ko - ko*

Different kind term - Different classifier

25	chicken - squirrel	**bird - mammal**	*wa - hiki*
26	cat - hawk	**mammal - bird**	*hiki - wa*
27	rabbit - snake	**mammal - reptile**	*wa - hiki*
28	turnip - bush	**root vegetable - plant**	*ko - hon*
29	fern - radish	**plant - root vegetable**	*mai - hon*
30	refrigerator - hammer	**appliance - tool**	*dai - hon*
31	knife - bus	**tool - vehicle**	*hon - dai*
32	bicycle - shovel	**vehicle - tool**	*dai - ko*

[i] For example, Yucatec Mayan, a classifier language, requires classifiers only when the number of objects being counted is ten or less. The first ten numbers in Yucatec are traditional Yucatec numbers, whereas numbers after ten are borrowed from Spanish and do not require classifiers.

[ii] In their Study 4, Schmitt & Zhang (1998) compared Chinese and Japanese speakers' similarity judgments on a different set of stimulus pairs. Both groups rated pairs with shared classifiers significantly differently from pairs not sharing a classifier, but again kind and classifier were conflated.

[iii] How to determine the appropriate level of shared kind terms is a difficult issue, as Gleitman (1990) has so articulately noted. We recognize that at higher levels of an ontological hierarchy many "peculiar" members of a classifier category may share an even higher-order term with other members of their classifier category, e.g., both rabbits and birds are living things. Yet, at the lower level, they are clearly of different kinds. Moreover, not all living things share a single classifier.

[iv] Male speakers of Japanese in the vicinity were much more fluent in English on average than were females, most of whom were spouses of graduate students. Thus, we chose to limit the subjects for this study to females.