The Effects of Numeral Classifiers and Taxonomic Categories in Chinese Speakers' Recall of Nouns

Shuping Huang (sphuang@staff.nsysu.edu.tw)
Department of Foreign Languages and Literature, National Sun Yat-Sen University
70 Lien-Hai Road, Kaohsiung 804, Taiwan ROC

Jenn-Yeu Chen (psyjyc@mail.ncku.edu.tw)
Department of Psychology, National Cheng Kung University
1 University Road, Tainan 701, Taiwan ROC

Abstract

It has been suggested that classifiers in Mandarin Chinese serve a semantic function of categorizing the nouns in terms of their perceptual and functional features. We investigated the classifiers' organizational utility in a recall task by contrasting it with that of taxonomic categories. Mandarin participants studied and recalled immediately two lists of nouns, one associated with four taxonomic categories and the other with four classifiers. The nouns were presented randomly in bare forms or in four columns headed by category names or classifiers. Comparable subjective clustering effects were found in the recall of taxonomically categorized nouns whether they were presented randomly or in columns. The recall of classifier categorized nouns showed no clustering when presented randomly, but some (though smaller) clustering when presented in columns. The findings suggest that classifiers do not serve the same function as taxonomic categories and that their semantic function may be limited.

Keywords: classifier; categorization; semantic memory; subjective clustering

Introduction

Categorization is basic to language use and cognition. We encounter an infinite number of entities everyday, and the terms we adopt to talk about them are merely reference to their "kinds," such as *tree, cup, love*, etc. As the referring to concepts is closely related to the linguistic labels of them, the relation between language and thought has long intrigued researchers in many fields. Along this line, a very relevant issue pertains to the language-particularities of nominal classification. Every language devises ways to mark classes of nouns, and the carve-up of nominal concepts varies drastically across languages. On the assumption that the grammar of a language may reflect its speakers' mind, the possibility that different nominal classifications are signs of varying worldviews has been vigorously pursued by linguists and psychologists (cf. Lucy, 1992; Nisbett, 2003). The use of classifiers, for example, is one way of signaling the class of a referred noun. Given that each entity has multiple semantic facets, classifiers provide unique ways to view objects in terms of a limited number of semantic parameters: material, animacy, shape, consistency, size, function, and orientation (Croft, 1994; Aikhenvald, 2004; Allan, 1977). Mandarin Chinese features a numeral classifier system. In numeral or deictic constructions, a noun is preceded by a classifier that specifies some salient perceived features of the referred entity, such as *yi ke shu* (one KE tree 'a tree') and z*he mei yingbi* (this MEI coin 'this coin'). Classifiers are believed to contribute to the semantics of a noun phrase (Tai, 1994). For example, *ke* is associated with plants, and *mei*, with small round and solid objects. Members in some classifier categories may be highly heterogeneous, but can mostly be accounted for by motivated extensions (Lakoff, 1987).

Whether the Chinese classifier system has to do with the organization of object concepts in the speakers' mind has attracted some attention recently. Kuo and Sera (2009), for example, found that Chinese speakers classified objects preferably by shape, which is the predominant semantics of Chinese classifiers. Zhang and Schmitt (1998) also showed that Chinese speakers, compared with English users, tend to group objects according to their associated classifiers. On the other hand, Saalbach and Imai (2007) argued that object concepts generally followed a universal principle of organization, and the role of classifiers was minor if any. It appears that the nature of the classifier effect may be complicated and further scrutiny is necessary.

Semantic Memory and Concept Organization

The question of what people know of "object concepts" pertains to our belief of knowledge organization. The structure by which a concept is stored for general purposes is known as "semantic memory," and it often reveals itself in cognitive tasks relevant to learning and understanding (Bransford, 1979). Recalling of a list, for example, pushes us to call for our knowledge structure. When participants are given a list of entities to memorize, the order and number of entities being recalled are found to reflect subjective clustering of concepts into smaller meaningful sequences (Bousfield, 1953; Tulving, 1962).

Of various relations among objects, taxonomy is attested as a prominent principle of concept organization. There could also be other schemes of object organization, for example, thematic relations (Lin & Murphy, 2001; Nisbett, 2003; Saalbach & Imai, 2007). In Chinese, the presence of a classifier in front of a noun invites us to ask the question of whether classifiers could be another scheme of object organization. In fact, the question has been raised and tested before. Using 16 nouns associated with four different

classifiers, Zhang and Schmitt (1998) presented the nouns in bare form and in a random order for Chinese and English participants to memorize and recall. They observed higher clustering of the nouns in the recall of the Chinese participants than in the recall of the English participants. Gao and Malt (2009) examined the role of classifier in recall with greater details and depth. Classifiers were first distinguished according to their relationship with the nouns. Three groups were identified based on Chinese speakers' ratings: well-defined, prototypical, and arbitrary. They selected nouns that paired with the three groups of classifiers and placed them in sentence frames. In one version of the sentences, classifiers were present whereas in another version (although equally grammatical), classifiers were absent. The sentences were shown in blocks defined by classifier groups and in a random order within a block to the Chinese and English participants for later recall of the nouns. The interesting findings of their study pertain to the recall and clustering of the nouns associated with well-defined classifiers. For these nouns, clustering was greater when classifiers were present than when classifiers were absent, but this classifier effect was apparent only in the recall of the Chinese participants. For the prototypical and arbitrary classifiers, noun recall also displayed a clustering trend, but the extent of clustering was comparable across languages and whether classifiers were present. In terms of the well-defined classifiers, Gao and Malt's (2009) results were consistent with those of Zhang and Schmitt (1998). However, Gao and Malt did not show whether a significant clustering beyond chance was present when classifiers were not explicitly used in sentences and when the classifiers were the prototypical and the arbitrary types. Furthermore, neither study contrasted classifiers with other well-known memory organizational schemes such as the taxonomic categories. Saalbach and Imai (2007) contrasted classifier with thematic and taxonomic relations in a categorization task, a similarity judgment task, and an inductive reasoning task. The authors observed a classifier effect in the similarity judgment task and the inductive reasoning task, and the effect was greater for the Chinese speakers than for the German speakers. However, the classifier effect was much smaller than the thematic and the taxonomic effects, both of which were comparable across languages. Zhang and Schmitt (1998) as well as Gao and Malt (2009) have demonstrated the differential effectiveness of classifiers (at least the well-defined type) as a memory organizational scheme for classifier and non-classifier languages. But, their work did not inform us of the relative effectiveness of classifiers as compared with the more universal organizational scheme such as the taxonomic categories. Saalbach and Imai's work (2007) shed much light on the contrasting roles of classifiers and taxonomic categories as an organizational scheme in speakers' semantic memory. But, the contrasting roles were not demonstrated in memory recall. The purpose of the present study was two folds. First, we wanted to determine whether Chinese speakers would spontaneously, as opposed to being explicitly primed to, detect the classifier relation among a set of nouns and to use the relation to help them organize the nouns in their memory. Second, we sought to determine the relative effectiveness of classifiers and taxonomic categories in aiding memory recall for Chinese speakers. Nouns that were associated with four well-defined classifiers and nouns that fell in four taxonomic categories were selected for study and recall. Experiment 1 presented the nouns in random order without the classifiers or the taxonomic categories. Experiment 2 presented the nouns in groups labeled by the classifiers or the taxonomic categories.

Experiment 1

Two lists of nouns were presented to speakers for recall, one with nouns organized by classifier relations (C-Nouns), and the other with nouns organized by taxonomic relations (T-Nouns). By observing how participants recalled the items of each list, we aimed to see whether these two schemes of categorization are equally activated in the participants' semantic memory for effective concept retrieval.

Method

Stimuli C-Nouns consisted of 20 nouns associated with 4 different classifier categories and T-Nouns comprised 20 nouns from 4 different taxonomic categories (Table 1). All nouns were two-character words, and within each list, the nouns shared no apparent orthographic similarity. To make sure the two lists of nouns are equally memorable, frequencies of occurrence were matched according to the data of Google Query. Importantly, the C-nouns did not share taxonomic relations, and the T-nouns did not share the same classifier. Because a Chinese noun was likely to go with more than one classifier, we relied on Chinese Sketch Engine to make sure the C-nouns selected had a high tendency to pair with a specific classifier.[1]

Table 1: C-Nouns and T-Nouns used in Experiment 1.

C-Nouns			
tiao2	***ba3***	***kuai4***	***mei2***
du2she2 'serpent'	*shou3qiang1* 'pistol'	*bu4liao4* 'cloth'	*jie4zhi3* 'ring'
yao1dai4 'belt'	*shu1zi0* 'comb'	*bing3gan1* 'cookie'	*you2piao4* 'stamp'
he2liu2 'river'	*fu3tou2* 'axe'	*mu4ban3* 'board'	*hui1zhang1* 'badge'
mian2bei4 'quilt'	*ji2ta1* 'guitar'	*fei2zao4* 'soap'	*zha4dan4* 'bomb'
tie3gui4 'rail'	*yu3san3* 'umbrella'	*hong2zhuan1* 'brick'	*ying4bi4* 'coin'

[1] Chinese Sketch Engine is devised by the Chinese WordNet Research Group, Academia Sinica, Taiwan. The sketch engine is provided by Lexical Computing Ltd., and our search was based on the database Gigaword, which is provided by Linguistic Data Consortium. Please visit http://wordsketch.ling.sinica.edu.tw to learn more information.

T-Nouns			
BODY	***CITY***	***MUSICAL INSTRUMENT***	***VEGETABLE***
jian1bang3 ‘shoulder’	*ba1li2* ‘Paris’	*da4gu3* ‘drum’	*cing1jiao1* ‘green pepper’
buo2zi0 ‘neck’	*man4gu3* ‘Bangkok’	*chang2di2* ‘flute’	*buo1cai4* ‘spinach’
xi1gai4 ‘knee’	*hua2sha1* ‘Warsaw’	*la3ba1* ‘trumpet’	*huang2gua1* ‘cucumber’
jiao3huai2 ‘ankle’	*kai1luo2* ‘Cairo’	*gang1cin2* ‘piano’	*suan4tou2* ‘garlic’
shou3zhou3 ‘elbow’	*ya3dian3* ‘Athena’	*xiang3ban3* ‘clappers’	*luo2buo0* ‘carrot’

Participants Twenty undergraduate students from National Cheng Kung University, Taiwan, participated in this experiment. They were native speakers of Mandarin Chinese, aging from 18 to 24 years old (mean age: 20.9 yrs). There were 13 females and 7 males.

Procedure The participants were told to memorize two lists of nouns. For each list, twenty bare nouns were presented on a computer screen in a random order at a rate of two seconds per word. At the end of presentation, the participants wrote down on a sheet of paper as many words as could be recalled in a period of three minutes. There were grids on the sheet that guided the participants to write down the nouns one in a separate line from the top to the bottom. The same list was tested five times. Half of the participants were given the C-Noun list before the T-Noun list, and the other half received the T-Noun list first.

Measures Following Frender and Doubilet (1974), we adopted the ratio of repetition (RR) as an indicator of subjective clustering. RR was defined as the “fraction representing the number of repetitions of items divided by the total item listed” (Bousfield, 1953), given by the equation in (1). Frender and Doubilet (1974) further suggested that the clustering tendency can be efficiently ascertained by comparing RR with its expected value in the protocol. The expected ratio of repetition (E_{RR}), or the rate of clustering happening simply by chance, can be determined by a simple formula, as defined in (2).

(1) RR: ratio of repetition

$$RR = \frac{r}{n-1}$$

(2) E_{RR} (Expected ratio of repetition)

$$E_{RR} = \frac{\sum_{i=1}^{c} e_i(e_i - 1)}{N(N-1)}$$

where,
r = the number of category repetitions in a subject’s recall
n = the number of items recalled
c = the number of categories in the protocol
e = the number of exemplars in a category
N= the total number of items in a protocol

For any two directly adjacent nouns on the recalled list, their sharing of the same categorical relation was counted one time of category repetition (r).

Results and Discussion

Participants tended to recall more T-Nouns (17.43 out of 20 items) than C-Nouns (15.04 out of 20 items): $F(1, 19) = 41.08$, $p < .0001$. They also tended to recall more nouns in later trials than in earlier ones: $F(4, 76) = 148$, $p < .0001$. This practice effect was similar for C-Nouns and T-Nouns, i.e., the interaction effect was not significant, $p > .08$.

Figure 1 shows the mean RR of participants’ recall of C-nouns and T-nouns in five successive trials. T-nouns consistently yielded a higher RR than the expected ratio of repetition. For recall of C-nouns, RR became slightly higher only in later trials. For our design of 4 categories X 5 members, the optimal RR is [4*(5-1)]/(20-1)=0.84. That means the RR of T-Nouns almost reached its highest point in later trials. In other words, participants almost always recalled T-Nouns in an organized pattern after the third trial. For both C-Nouns and T-Nouns, the growth of RR slowed down after the third trial.

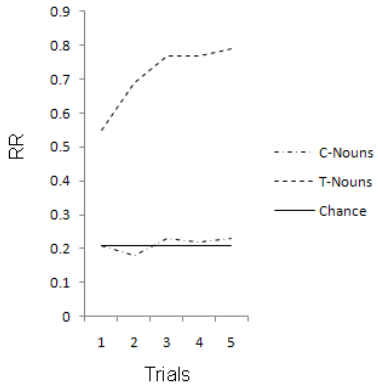

Figure 1: RR plot of C-Nouns and T-Nouns over 5 successive trials in Experiment 1.

The RR of T-Nouns (0.71) was higher than that of C-Nouns (0.21): $F(1, 19) = 450.6$, $p < .0001$. There was a significant practice effect. RR increased over trials: $F(4, 76) = 14.2$, $p < .0001$. A significant interaction was also observed, indicating greater practice effect for T-Nouns than for C-Nouns: $F(4, 76) = 6.7$, $p < 001$. Additional t-tests indicated that RRs were significantly greater than the chance for all trials of T-Nouns, p’s < .001, whereas the RRs were not significantly greater than the chance for C-Nouns, p’s > .2.

The results of this experiment showed that when nouns were presented in bare forms, Chinese speakers could spontaneously pick up the taxonomic relations among the nouns and used the cues to organize the nouns for effective memory recall. In contrast, the same participants failed to detect the classifier relations among the nouns, even when

they had studied the nouns multiple times, and their recall of the nouns showed no sign of clustering whatsoever. The results suggest that classifiers may not be inherently conceptual and do not seem to play an active role in Chinese speakers' semantic memory.

Experiment 2

Experiment 2 employed the same stimuli as Experiment 1, but differed in the way items were presented to the participants. Items of the same category were presented in the same column, and a linguistic label of their shared category was placed right above the column. For C-Nouns, the labels were the shared classifiers among category members, and for T-Nouns, the labels were taxonomic nodes that properly expressed the commonalities among category members. By labeling the nature of each category, we aimed to examine whether taxonomic and classifier categories would be equally activated for object organization when made explicit to the participants.

Method

Stimuli The same two lists of nouns from Experiment 1 were prepared in two booklet versions. In Version I, twenty nouns were printed on the same page, arranged into four columns corresponding to four categories. Shown above each column was a linguistic label of the category that properly represented the nature of the category. Three pages of C-Nouns were followed by another three pages of T-Nouns. Version II differed from Version I only in the order of the C-Noun pages and the T-Noun pages.

Participants Fifty-six undergraduate students taking the "Psychology of Language" course at National Cheng Kung University participated for an extra credit.

Procedure Participants were tested in the classroom in group. They were given 40 seconds to memorize twenty nouns on a page. After turning the page as instructed by the experimenter, they wrote down on a separate sheet as many words as they could recall in three minutes, in the same way as required in Experiment 1. The participants repeated the same list three times before switching to the other list. Roughly half of the participants received Version I of the booklet (n=27), and the other half received Version II (n=29).

Results and Discussion

T-Nouns (17.76 out of 20 items) were recalled better than C-Nouns (15.54 out of 20 items), $F(1, 55) = 98$, $p < .0001$. Nouns were recalled better in later trials, indicating a practice effect: $F(2, 110) = 445.9$, $p < .0001$. Practice effect was greater for recall of C-Nouns than for T-Nouns, $F(2, 110) = 30.8$, $p < .0001$.

Figure 2 shows the RR in the recall of C-Nouns and T-Nouns over three successive trials. A significant clustering effect is apparent in the recall of T-nouns. Importantly, given explicit linguistic cues, recall of the C-Nouns also shows significant clustering. For both C-Nouns and T-Nouns, RR was higher than the chance, but the RR of T-Nouns was higher than that of C-Nouns in all three trials.

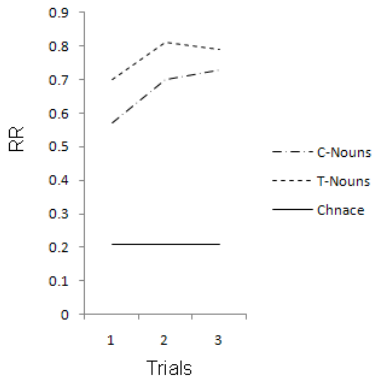

Figure 2: RR plot of C-Nouns and T-Nouns over 3 successive trials in Experiment 2.

The RR of T-Nouns (0.76) was higher than that of C-Nouns (0.66): $F(1, 55) = 30.7$, $p < .0001$. There was a significant practice effect indicating that RR increased over trials: $F(2, 110) = 33.9$, $p < .0001$. Practice effect was similar for T-Nouns and C-Nouns: $F(2, 110) = 2.4$, $p > .09$. T-tests showed that the clustering effect was greater than the chance for both T-Nouns and C-Nouns, p's $< .0001$.

Comparisons of the results of the two experiments (using the first three trials of Experiment 1) indicate that explicit linguistic labels of categories helped the amount of recall only slightly for both T-Nouns (from 16.02 to 17.76) and C-Nouns (from 13.23 to 15.54)). The labels also exerted minimal influence on the subjective clustering of the T-Nouns (from .67 to .76). In contrast, explicit classifier labels increased subjective clustering of the C-Nouns substantially (from .21 to .66).

General Discussion

The present study examined whether the organization of Chinese speakers' object concepts reflected the classifier system of the language as opposed to the taxonomic system. We found that when items were presented as bare nouns, participants detected taxonomic relations among the items, and accordingly clustered the items for effective recall. But for items sharing the same classifier, participants did not find the relation among them. Our findings were at odds with those of Zhang and Schmitt (1999). Following Pellegrino and Hubert (1982), they assessed the tendency of subjective clustering by "adjusted ratio of clustering" (ARC) which is a standardized score ranging from -1 to 1, with 0 indicating chance of accidental repetition. In Zhang and Schmitt's study, the ARC value of the Chinese speakers reached 0.68, which indicates a positive classifier effect. Computing the ARC with our data from Experiment 1, we obtained a value of 0.085, very close to zero, indicating little

classifier effect. By contrast, for recalling of taxonomically-related items, the ARC reached 0.637.

Our results from both experiments indicate an effect of classifier presence, which is in line with those of Gao and Malt (2009) with respect to their well-defined classifiers. The data from the first trials of our experiments reveal an increase of clustering from chance (.21) in Experiment 1 to .57 in Experiment 2. The increase was from .56 to .81 in Gao and Malt's data when classifiers were absent as opposed to present. Because Gao and Malt did not include a chance estimate, it is unclear whether the classifier-absent condition produced significant clustering. Our data showed no clustering beyond chance when classifiers were absent. More importantly, by including the taxonomic relations, we were able to show that whereas the taxonomic relations are an effective organizational scheme in semantic memory and is activated automatically, the classifier relations are not as effective and certainly do not emerge spontaneously without explicit mention.

Our results invite two possible explanations. Firstly, it is very likely for the classifier effect to be of a different nature from taxonomic relations. Although members in a classifier category usually share semantic similarities, the shared features are not always as salient as taxonomic relations. In fact, classifier categories reflect a variety of classifying rationales: shape, material, rigidity, etc. which makes them more arbitrary than taxonomic relations that operate only on in-built characteristics of objects. In one classifier category, it is common to involve more than one defining feature, for example, TOU for huge and animate beings. In addition to that, presentation of classifiers together with the stimuli may have invoked participants' linguistic knowledge rather than conceptual information. In previous studies, Wang, Guo, Bui, and Shu (2006) found a classifier effect only in naming of "Classifier + Noun" phrases, but not in naming of bare nouns. Similar result was also found in Alario, Yu, Geng, and Bi (2009) which suggested that items with similar shapes only produced an interference effect in naming of objects in nominal phrases containing a classifier. Following the notion of "thinking for speaking" (Slobin 1996), the effect of classifier could be a dynamic one. It may be activated when linguistic knowledge is demanded, such as when reading nouns along with a specific classifier.

The significant difference between C-Nouns and T-Nouns also invites another explanation, namely, "categorization" as a graded notion. Discussion of "categorization" often centered around categories that are constant and a priori, such as BIRD, TREE and ANIMAL. In the past century, psychologists began to take an interest in categories that are created spontaneously for immediate problem-solving, such as "things you take from your house when it is on fire" or "possible costumes to wear to a Halloween party" (Barsalou, 1983; Little & Lewandowsky, 2006). Given the term "*ad hoc* categories" (following Lawrence Barsalou), these spontaneously-activated categories had graded internal structure centralizing typical member(s), just like constant categories. Nevertheless, they were less effective for retrieving and clustering in recall experiments (Barsalou, 1983). Our findings seemed to be in resonance with Barsalou's, which called for rethinking of the nature of classifier categories as *ad hoc* rather than constant.

This claim certainly demands a thorough consideration regarding how *ad hoc* categories differ from constant ones. Philosophers such as Vervaeke and Green (1997) have argued the necessity to distinguish between "physical relation" and "superphysical relation". The former is typically known as the "kind-of" relation. It is also called "general-purpose taxonomy" that is autonomous, disregarding linguistic or social considerations. On the contrary, superphysical relations represent the so-called "special-purpose taxonomy" that is constructed for a specific demand, and is in this sense interactive with the immediate context. According to Barsalou (1983), common categories make superior mnemonic devices to *ad hoc* ones probably because their "concept-to-instance" associations are better established in memory, and we can activate them even without context. Unlike physical relations, superphysical relations have a short life span, and reflect a wide variety of relations from metaphor, metonymy, to propositional or perceptual similarities (Lakoff, 1987). Chinese classifier system as *ad hoc* categories was manifested in our two experiments: Classifier relations were employed by our participants only when the contexts (labels of categories) were given, but for taxonomic relations, participants could detect them even without an explicit cue.

The distinction between these two submodels of categorization is not always clear. Wierzbicka (1984), for example, regarded FRUIT as a superphysical relation whereas it was usually deemed a common category in psychological experiments (cf. Freedman and Loftus, 1971). Also, combining of common and ad hoc relations of a category is possible. For example, classifier KE introduces plants, and another classifier ZUO houses huge entities or buildings. They are organized with conceptual similarities and sometimes confront with taxonomic relations. For some classifying categories, the shared commonalities are easy to detect. That might account for the classifier effects found in some studies, such as Zhang and Schmitt (1999). In addition, frequency of category entrenchment plays a role. Barsalou (1983) suggests that an *ad hoc* category can shift to a well-established one when it is processed frequently. Constant use of classifiers to count or index objects is likely to entrench the classifier relations, and Chinese speakers are privileged to have an additional scheme of classification which is activated when the duty calls.

Conclusion

In cognitive linguistics, categorization is considered of crucial importance to language use (Croft & Cruse, 2004; Taylor, 2002; Ungerer & Schmidt, 1997). Nevertheless, Labov (1973, p. 342) was probably right in pointing out that it is "such fundamental and obvious part of linguistic activity that the properties of categories are normally assumed rather than studied." In the past few decades,

psychological experiments have enabled us to look at the nature of linguistic categorization in a less retrospective way. Although previous psychological studies did not always resonate in their answers, we believe we are collecting pieces of a puzzle to get a grand picture.

Relative to Zhang and Schmitt (1999) and Gao and Malt (2009), by two experiments, we found that classifier effect was observed in participants' recall of nouns in specific contexts. Nevertheless, we also made further comparison of classifier effect with taxonomy categories, demonstrating that the relatively weaker conceptual-organizing function of classifier knowledge found in Saalbach and Imai (2007) was also detected in noun-recall task. The results calls for a view to regard classifiers as *ad hoc* categories activated mostly when speakers need to think for speaking. Questions regarding the types of classifier categories and the degree of their entrenchment are yet to be determined in our future investigations.

Acknowledgments

This study is founded by National Cheng Kung University as part of the Landmark Project entitled "An Inter-disciplinary Approach to the Study of Natural Language Processing and Acquisition," directed by Prof. Jenn-Yeu Chen. The authors are grateful to Academia Sinica CWS team for letting us use the Chinese Sketch Engine, and to Adam Cheng-yi Li for collecting data in Experiment 2. All errors remain, of course, our own.

References

Aikhenvald, Y. A. (2003). *Classifiers: A typology of noun categorization devices*. New York: Oxford University Press.

Alario, X., Yu, X., Geng, J., & Bi, Y. (2009). *How visual shape may constrain lexical access: Evidence from Chinese classifier production*. Paper presented at the Architectures and Mechanisms for Language Processing. Barcelona, September 7-9, 2009.

Allan, K. (1977). Classifiers. *Language, 53*(2), 285-311.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*(3), 211-227.

Aikhenvald, A. Y. (2004). Nominal classification: Towards a comprehensive typology. *Sprachtypologie und Universalienforschung, 57*(2/3), 105-116.

Bransford, J. D. (1979). *Human cognition: Learning, understanding, and remembering*. Belmont, CA: Wadsworth.

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

Croft, W. (1994). Semantic universals in classifier systems. *Word, 45*(2), 145-171.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Freedman, J. L., & Loftus, E. F. (1971). The retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*, 107-115.

Frender, R., & Doubilet, P. (1974). More on measures of category clustering in free recall - Although probably not the last word. *Psychological Bulletin, 81*(1), 64-66.

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes, 24*(7), 1124-1179.

Kuo, J. Y.-c., & Sera, M. D. (2009). Classifier effects on human categorization: The role of shape classifiers in Mandarin Chinese. *Journal of East Asian Linguistics, 18*, 1-19.

Labov, W. (1973). The boundaries of words and their meaning. In C.-J. N. Bailey & R. W. Shuy (Eds.), *New ways of analyzing variation of English* (pp. 340-373). Washington D. C.: Georgetown University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: Chicago University Press.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*(1), 3-28.

Little, D. R., Lewandowsky, S., & Heit, E. (2006). Ad hoc category restructuring. *Memory and Cognition, 34*(7), 1398-1413.

Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge, UK: Cambridge University Press.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York: The Free Press.

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General, 136*(3), 485-501.

Tai, J. H.-Y. (1994). Chinese classifier systems and human categorization. In M. Chen & O. Tseng (Eds.), *In honor of Professor William S-Y. Wang: Interdisciplinary studies on language and language change* (pp. 479-494): Pyramid.

Taylor, J. (2002). *Cognitive grammar*. Oxford: Oxford University Press.

Tulving, E. (1962). Subjective organization in free recall of "unrelated words". *Psychological Review, 69*, 344-354.

Ungerer, F., & Schmid, H.-J. (1996). *An introduction to cognitive linguistics*. Essex: Longman.

Vervaeke, J., & Green, C. D. (1997). Women, fire, and dangerous theories: A critique of Lakoff's theory of categorization. *Metaphor and Symbol, 12*, 59-80.

Wang, L., Guo, J., Bui, Y., & Shu, H. (2006). Classifier congruency effect in the production of noun phrases. *Studies of Psychology and Behavior, 4*, 34-38.

Wierzbicka, A. (1984). Apples are not a "kind of fruit": The semantics of human categorization. *American Ethnologist, 11*, 313-328.

Zhang, S., & Schmitt, B. (1998). Language-dependent classification: The role of classifiers in cognition, memory and ad evaluation. *Journal of Experimental Psychology: Applied, 4*(4), 375-385.