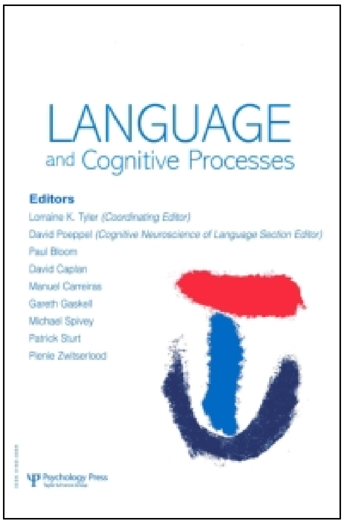

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

Mental representation and cognitive consequences of Chinese individual classifiers

Ming Y. Gao [a]; Barbara C. Malt [b]
[a] Harrisburg Area Community College, Harrisburg, PA, USA [b] Lehigh University, Bethlehem, PA, USA

First Published:September2009

To cite this Article Gao, Ming Y. and Malt, Barbara C.(2009)'Mental representation and cognitive consequences of Chinese individual classifiers',Language and Cognitive Processes,24:7,1124 — 1179

To link to this Article: DOI: 10.1080/01690960802018323

URL: http://dx.doi.org/10.1080/01690960802018323

LANGUAGE AND COGNITIVE PROCESSES
2009, 24 (7/8), 1124–1179

Psychology Press
Taylor & Francis Group

Mental representation and cognitive consequences of Chinese individual classifiers

Ming Y. Gao
Harrisburg Area Community College, Harrisburg, PA, USA

Barbara C. Malt
Lehigh University, Bethlehem, PA, USA

Classifier languages are spoken by a large portion of the world's population, but psychologists have only recently begun to investigate the psychological reality of classifier categories and their potential for influencing non-linguistic thought. The current work evaluates both the mental representation of classifiers and potential cognitive consequences for speakers of Mandarin Chinese. We first provide a taxonomy of 126 common classifiers and a large sample of the objects classified by each as a tool for this and future research. We then present four studies investigating potential variation in the mental representation of the classifier categories. The data provide evidence that at least three forms of mental representation need to be distinguished. Finally, we present a fifth study investigating the impact of this variation on the cognitive consequences of classifier knowledge. This study suggests that the differences identified in Experiments 1–4 have important implications for the likelihood of finding cognitive consequences.

Correspondence should be addressed to Barbara Malt, Department of Psychology, 17 Memorial Drive East, Lehigh University, Bethlehem, PA 18015, USA. E-mail: barbara.malt@Lehigh.edu

We thank Susan Barrett, Constance Cook, Sue Gao, Padraig O'Seaghdha, and Martin Richter for useful input on the project, Mutsumi Imai and Henrik Saalbach for assistance in obtaining corpus-based classifier frequency information as well as providing helpful comments on an earlier draft of this paper, and Zhang Yuping and Shu Hua for providing the corpus frequency information. Portions of the research were supported by a Grant-in-Aid of Research to Ming Gao from Sigma Xi, the Scientific Research Society, and by NIMH grant MH51271 to Barbara Malt and Steven Sloman.

http://www.psypress.com/lcp DOI: 10.1080/01690960802018323

Psychological studies of categorisation have overwhelmingly focused on the groupings of entities labelled by common nouns such as *dog*, *ball*, *bottle*, etc. But many of the world's languages, while having nouns that function the same way that nouns in English and other Indo-European languages do, also have a second type of morpheme that picks out groups of objects: noun classifiers. In Mandarin Chinese, for instance, if a sentence indicates a specific quantity of something or if a demonstrative is used, the noun naming the objects must be preceded by a classifier. Whereas in English, one would say *a rope*, a Chinese speaker would say *yi tiao shengzi*, in which *yi* is the numeral 1, *shengzi* means 'rope', and *tiao* is a classifier indicating a long thing (literally, 'one long-thing rope'). Similarly, *that rope* would be expressed as *na tiao shengzi*, 'that long-thing rope'. Most classifiers are used with more than one noun, and classifiers can therefore be thought of as defining categories: the set of nouns (or the entities they refer to) that occur with a particular classifier.

Classifier languages – which include most East and Southeast Asian languages (Adams & Conklin, 1973; Allan, 1977; T'sou, 1976), some Australian aboriginal languages (Dixon, 1982), and some native American languages (Berlin, 1968, Dixon, 1982; Lucy, 1992; Haas, 1967) – are spoken by a large portion of the world's population, and noun classification has long been familiar to linguists. Psychology's general introduction to this phenomenon is more recent and is largely attributable to Lakoff's (1986, 1987) intriguing discussion of the Japanese classifier *hon* as an example of categorisation. *Hon* is used with nouns labelling a diverse set of objects that include not only long, slender objects such as sticks, canes, pencils, and candles, but also trees, hair, martial arts contests, judo matches, rolls of tape, telephone calls, baseball hits, and medical injections (Downing, 1984, p. 13; Lakoff, 1986, 1987, p. 104). Such diversity in category membership has sometimes been interpreted to suggest that classifiers are arbitrary grammatical devices indicating only 'unit' (of something) (e.g., Greenberg, 1972; Noss, 1964, also cited in Hiranburana, 1979; see also Burling, 1965). However, Lakoff argued that the diverse membership is explainable in terms of motivated extensions from central examples. For instance, he suggested that hits in baseball are classified by *hon* because baseball bats (being long and thin) are central examples of *hon*, and hits emanate from them and follow a trajectory that is long and thin. Lakoff thus considered classifier categories to reflect the 'imaginative aspects of mind' (1987, p. 113) and concluded that they are a form of conceptual category.

If the groupings of objects picked out by classifiers are indeed meaningful to speakers, then classifier categories may have cognitive consequences beyond whatever role they play in sentence processing. Denny (1976) has suggested that classifier groupings tend to highlight how humans interact with the classified entities physically, functionally, or socially, whereas groupings picked out by nouns tend to highlight properties intrinsic to the objects. If speakers of classifier languages have available a stable, well-learned organisation of things in the world that differs from that given by nouns such as *dog*, *ball*, *bottle*, they may differ from speakers of other languages in their behaviour on cognitive tasks ranging from memory retrieval to reasoning and property inference.

But much remains to be understood about both parts of the proposition. Should classifier categories truly be considered conceptually coherent and meaningful to their users? And to the extent that they are, do they have any cognitive consequences beyond their role in sentence processing? The existing evidence and arguments, which we discuss next, tell a story on both these points that is not entirely straightforward. Our research aims to shed further light on both these issues with regard to Mandarin Chinese.

HOW MEANINGFUL ARE CLASSIFIER CATEGORIES?

Many linguists in addition to Lakoff have considered classifier categories to be meaningful to speakers, for several compelling reasons (Adams, 1986; Allan, 1977; Burling, 1965; Erbaugh, 1984; Hiranburana, 1979; Pulman, 1978; Sanches, 1977; T'sou, 1976). First, although some classifier categories include a highly heterogeneous set of objects such as in the case of Japanese *hon*, others have a more restricted set with a more transparent basis such as shape or kind. Second, even among those with heterogeneous membership, there is often a subset of members that have a clear similarity on some dimension to each other. Third, across languages (including ones with no close genetic relation), there appear to be some common bases for classifier categories such as shape, animacy, function, and social status. This fact suggests that at least some classifier categories are grounded in properties that are universally salient to human perception/cognition (Lyons, 1977; Mithun, 1986; Pulman, 1978). And finally, native speakers of a classifier language have clear and shared intuitions about what category a new object belongs to (Allan, 1977; Burling, 1965; Hiranburana, 1979; T'sou, 1976). If classifier use depended entirely on memorisation of individual instances, such productivity would not exist.

Thus a convincing case exists that classifiers as a whole are not merely empty syntactic devices. But at the same time, concluding that all classifier groupings, for all languages, are meaningful to speakers of the language may overlook important complexities of the situation. As the case of Japanese *hon* illustrates, for at least some classifiers the nouns occurring with a given classifier can be quite diverse. Although Lakoff (1987) made the case for motivated links that brought the diverse members into the category, those links may be opaque to current speakers of the language, for whom much of the membership may be experienced as purely arbitrary. Furthermore, even when one can specify features that seem to determine membership, it can be unclear why some objects are included in a category but others aren't. For example, *tou* classifies many large-headed animals but does not apply to horses. Thus a non-trivial degree of arbitrariness seems to exist along with any transparent motivations. Finally, there may be variation from language to language in how meaningful or coherent the language's classifier categories are to its speakers, similar to the case for gender systems in Romance and other languages. Some researchers take gender assignment to be arbitrary with respect to the meaning of words (e.g., Caramazza & Miozzo, 1997) and others to be meaningfully grounded (in part) in characteristics of the entities named (Zubin & Kopcke, 1981). But the truth is not necessarily one or the other; languages may vary in the degree of semantic meaningfulness of their gender system and its consequent impact in other cognitive tasks (Sera, Elieff, Forbes, Burch, Rodriguez, & Dubois, 2002; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). In a related vein, Imai and Saalbach (in press) suggest that classifiers in Japanese are used much less frequently and are less salient to Japanese speakers than is the case for Chinese. This cross-language diversity may extend to how speakers experience the categories, with classifiers functioning as more arbitrary syntactic markers in some languages and as more meaningful in others. (Schmitt and Zhang, 1998, suggest that Japanese classifier categories tend to be broader in scope than Chinese, presumably carrying less specific semantic information as a result.) Thus it may not be possible to make blanket statements about the meaningfulness of classifiers systems across languages based on consideration of one language or an amalgam of evidence from various languages. It may not be possible even to make a blanket statement about the meaningfulness of classifier categories within a single language, if there is notable variability across its classifiers.

ARE THERE COGNITIVE CONSEQUENCES OF SPEAKING A CLASSIFIER LANGUAGE?

Experimental research examining cognitive consequences of speaking a classifier language is sparse, but several series of studies have begun to explore the implications. Schmitt and Zhang (1998; see also Zhang & Schmitt, 1998) presented words referring to concrete objects to speakers of Chinese and English (and, in some tasks, Japanese) in tasks including similarity ratings, feature listing, clustering in recall, and inference and choice. They found differences in performance suggesting an impact of the Chinese classifier category relations among the stimuli. For instance, Chinese speakers rated objects belong to the same Chinese classifier category higher in similarity than speakers of English did, and Chinese produced classifier-related features sooner in their feature lists than speakers of English did. Japanese performance was not influenced by classifier distinctions present in Chinese but not Japanese. They concluded in favour of a pervasive effect of classifier category knowledge on Chinese thought. More recently, Saalbach and Imai (2007) tested Chinese and German speakers on tasks including forced choice categorisation, similarity judgements, property induction, and speeded word-picture matching. They selected stimuli so that the degree of the impact of classifier category membership could be compared to that of taxonomic and thematic relations. Like Schmitt and Zhang, they found that Chinese speakers' similarity judgements of objects were influenced by shared classifier category membership. They also found some impact of classifier category membership for induction judgements about 'blank' properties for which general world knowledge is irrelevant. However, they otherwise found no evidence that Chinese participants' performance was influenced by classifier category relations among the stimuli (beyond the degree of similarity among objects in the same classifier category that German speakers were also sensitive to), and they found that taxonomic relations dominated the patterns of responding for Chinese as it did for Germans. On this basis, they argued that the effect of classifier knowledge on task performance is minor at best for Chinese speakers and that taxonomic relations are the primary organising relation for them. Thus, the existing evidence is mixed regarding the nature and extent of cognitive consequences of speaking a classifier language.

THE CURRENT INVESTIGATION

This mixed evidence about the cognitive consequences of speaking a classifier language, along with the complexities regarding the status of classifiers as meaningful categories, suggest that more consideration needs to be given to the nature of classifier categories and their mental representation on a language-by-language and category-by-category basis. If the categories vary in meaningfulness or coherence between or within languages, conclusions regarding their psychological status will need to be nuanced. Furthermore, such variation may be crucial in understanding whether or when classifier category knowledge may have an impact on other aspects of cognition.

The current investigation evaluates possible variation in the mental representation of classifiers for speakers of Mandarin Chinese and the implication of this variation for cognitive consequences. Mandarin Chinese is spoken by approximately 900 million people and constitutes the classifier language probably most readily studied by cognitive psychologists, in terms of both availability of native speaker participants and access to expertise in the language. To set the stage for our research and assist future research, we first present further background on the nature of this system and a compilation of 126 common classifiers along with a large sample of the objects that are classified by each, organised according to the type of semantic feature(s) that appear to account for at least some of the categories' members. We then present four studies investigating potential variation in their mental representation. The data provide evidence that at least three forms of mental representation need to be distinguished. Finally, we present a fifth study investigating the impact of this variation on retrieval of information from memory. This study suggests that the differences identified in Experiments 1–4 have important implications for the likelihood of finding cognitive consequences.

THE MANDARIN CHINESE CLASSIFIER SYSTEM

Classifiers such as Japanese *hon*, Chinese *tiao*, and others used by Schmitt and Zhang (1998; Zhang & Schmitt, 1998) and Saalbach and Imai (2007), are said to carry meaning (to a greater or lesser extent) about features of the entities being classified. However, classifiers of this sort represent only part of a larger system of noun classification in most languages, and Chinese is no exception. Linguists have made distinctions among classifiers according to their apparent semantic function. The most common distinction is between classifiers that 'quantify' or in some way provide a measure of the amount of what is being referred to, and those that 'qualify' or identify a characteristic (other than amount) of what is being referred to (Adams & Conklin, 1973; Burling, 1965; Denny, 1976, 1986; Dixon, 1982; Downing, 1984; Lucy, 1992; Lyons, 1977). Chao (1968) provides the most detailed breakdown of Chinese classifiers, making nine distinctions. Collapsing across several of Chao's finest distinctions, Chinese noun classifiers can be described as belonging to five main categories: (1) group, (2) container, (3) standard measure, (4) temporary, and (5) individual classifiers.

Group classifiers signify a group of any size, ranging from two (e.g., *dui* means 'a pair', so *yi dui qinglu* means 'a pair of lovers'), to hundreds or thousands (e.g., *qun* means 'group', 'crowd', or 'flock', so *yi qun mianyang* means 'a flock of sheep'). Container classifiers are those that denote containers of all kinds and are especially useful for indicating the amount of an entity that is labelled by a mass noun, such as water, beer, or rice. For example, *bei* means 'glass', and *yi bei pijiu* means 'a glass of beer'; *chepi* means 'railroad (cargo) car', and *yi chepi xi gua* is 'one carload of watermelons'. Standard measures refer to units of measurement such as inch, metre, and kilogram. For example, *gong jin* means 'kilogram', and *san gongjin yu* is 'three kilograms of fish'. Temporary classifiers are an open class of words that can be temporarily used as units for counting or measuring things. For example, *lian* (meaning 'face') can be adopted for use as a classifier to talk about what is on someone's face, such as sweat, water, blood, dust, or mud; thus, *yi lian tu* means 'a faceful of dust'; *liang zhuozi wenjian* (in which *liang* = two, *zhuozi* = table or desk, and *wenjian* = document) means 'two deskfuls of documents'. Virtually any word that denotes an object having a surface on which something can land or be placed can be given a temporary use as a classifier.

Members of the last group, individual classifiers, are used to classify individual objects. Some are used with only a small number of nouns (e.g., *zhan* is used only for lamps and electric lights) but others are used with a large number of nouns (e.g., as already illustrated, *tiao* is used in talking about a large number of different entities, both animate and inanimate, including but not restricted to things that are physically long and thin). In general, individual classifiers categorise countable objects. The groupings of objects picked out by individual classifiers are as a whole the least like groupings familiar to speakers of non-classifier languages, and for this reason they have been of the greatest interest in terms of their mental representation and potential impact on non-linguistic thought. However, keeping in mind their place within the larger classification system may be important to understanding aspects of their cognitive status such as their salience to native speakers and their likelihood of impacting non-linguistic cognition. We return to this point in the General Discussion.

Within the category of individual classifiers, classifiers can be further divided according to the type of feature that appears most relevant to determining membership in the category (to whatever extent there are such discernable features). Drawing on analyses of classifier systems for other languages (Adams, 1986; Adams & Conklin, 1973; Allan, 1977; Burling, 1965; Dixon, 1982; Hiranburana, 1979; Sanches, 1977; Supalla, 1986; T'sou, 1976), we suggest that Chinese individual classifiers, like those of other languages, can be divided into two broad groups: those in which membership is linked to shape attributes such as length or roundness, and those in which membership is linked to multiple shared features and generally restricted to either animate or inanimate entities. Within each of these broad groups, a number of sub-groups can be discerned. They are identified in Appendix A, which presents common classifiers and a substantial sampling of the objects they classify (compiled as described later).

The majority of characters in Chinese that function as classifiers also are used in other contexts as other parts of speech. Of the 126 classifiers presented in Appendix A, dictionary definitions (*Xiandai Hanyu Cidian* [*Modern Chinese Dictionary*], 1984) show that only 19 of them, or 15%, function solely as classifiers. Ninety-three (73.8%) also function as nouns, 13 (10.3%) also function as verbs, and one (0.8%) also functions as an adjective.

For most (about 92%) of the 107 classifiers with non-classifier uses, there is some relation, close or more distant, perceivable between its other use and its use as a classifier. For instance, *chuang* means 'bed' when used as a noun, and when used as a classifier, it classifies quilts, blankets, cotton-padded mattresses, and beddings. Likewise, *ban* when used as a noun means 'work shift', and when used as a classifier it classifies mass transit vehicles that run on a fixed schedule such as buses, trains, ships, and aeroplanes. Thus for many, but not all, classifiers, part of the knowledge about what nouns occur with a given classifier may derive from knowledge of the associated noun, verb, or adjective meaning. In Appendix A we list the non-classifier use along with each classifier word where such a use exists.

A LIST OF 126 FAMILIAR CLASSIFIERS AND THE OBJECTS THEY CLASSIFY

An essential tool for behavioural research on Mandarin Chinese individual classifiers is a list of commonly recognised classifiers and the objects that they classify. To provide the basis for our studies and serve as a resource for future research, we compiled such a list, provided in Appendix A. Although several Chinese classifier dictionaries have been published (Chen, Che, Chen, & Zhang, 1988; Jiao, 2001), our list has several advantages for researchers. First, we isolate individual classifiers, whereas in the dictionaries, group, container, standard measure, temporary, and individual are intermingled without identification. Second, we isolate familiar ones, whereas the dictionaries do not discriminate them from those that are rarely used and unfamiliar to many current speakers of Chinese. Third, we provide an organisation according to the kinds of attributes that may account for at least some of their membership. We also provide information about the meaning associated with the classifier word when used in non-classifier contexts, which may shed light on the origin of the classifier and its meaning to speakers, as discussed above. Fourth, our list is presented in English, providing English-speaking researchers with a convenient information source. Finally, we have added frequency information for each classifier from a Chinese text corpus (obtained after completion of the research presented here as a resource for future studies).

To compile the list, classifiers were collected during a 10-month period in which author Gao added to the collection every classifier observed that would fit the criterion for individual classifiers (i.e., categorising countable nouns). Sources included Chinese books, newspapers (mainly *People's Daily, Overseas Edition*), dictionaries (e.g., Chen et al., 1988; *Xiandai Hanyu Cidian* [*Modern Chinese Dictionary*], 1984; *Han Ying Cidian* [*A Chinese-English Dictionary*], 1980), casual conversations with other native Chinese speakers (mainly Lehigh University graduate students and their spouses), and his own knowledge of Chinese. The 126 classifiers listed thus represent all or virtually all those likely to be familiar to college-educated speakers of Mandarin Chinese.

To confirm that the classifiers collected are currently used in Mandarin Chinese, six native speakers of Mandarin Chinese from Beijing (three graduate students at Lehigh University and three college-educated spouses of graduate students) were paid to judge their familiarity. A questionnaire listed each classifier accompanied by several examples of nouns that it may classify. Participants judged whether each classifier would be familiar to native speakers of modern Mandarin Chinese. Three classifiers in the original set were judged unfamiliar by one or more participants and are eliminated from the list presented here. The remaining 126 individual classifiers were judged familiar by all six.

The general classifier *ge*, used for any noun that does not fall into a more specialised classifier category, can also substitute for the more specialised classifiers (Erbaugh, 1986; Lyons, 1977), and it often does so in casual conversation and in children's speech. Nevertheless, as the data just reported and those below confirm, the other 125 classifiers and their applicability to substantial numbers of nouns are well known to educated speakers of Mandarin Chinese.

Concurrent with collection of the classifiers, and using the same sources, examples of nouns occurring with each classifier were recorded. As with the classifiers, all examples found during the collection period were included. Because the nouns that can be used with any particular classifier usually do not constitute a closed set, this compilation cannot represent every possible classifier-noun combination. (See the Chinese classifier dictionaries for additional examples.) To confirm that the nouns observed with each classifier are generally accepted as constituting a grammatical and sensible use, ten college-educated native speakers of Mandarin Chinese from Beijing (six living in Beijing and four living in eastern Pennsylvania, USA, at the time of testing) were paid to judge the appropriateness of the noun-classifier combinations. A questionnaire listed each of the 126 classifiers followed by the nouns compiled for it plus two filler items not likely to be judged as occurring with that classifier (i.e., nouns that would normally occur with a different classifier). Participants circled any of the nouns listed that they felt would not occur with the given classifier. Filler items were reliably rejected by all subjects, and a small number of other nouns were rejected by five or more subjects. The nouns included in the list in Appendix A were judged acceptable by at least six of the ten subjects, and the great majority were accepted by all ten.

EXPERIMENTS 1–4: THE MENTAL REPRESENTATION OF CHINESE NOUN CLASSIFIER CATEGORIES

In four studies we explored the nature of knowledge about individual Mandarin Chinese noun classifiers, drawing on views of category structure from the general categorisation literature (e.g., Lakoff, 1987; Murphy, 2002; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Smith & Medin, 1981). Based on native speaker intuitions along with consideration of the membership of classifier categories illustrated in Appendix A, we propose that Chinese classifier categories fall into three distinct types differing in (1) the knowledge that the classifier word itself evokes for native speakers when considered in isolation, and (2) the relation of the nouns used with a given classifier to one another and to the knowledge evoked by the classifier word. Experiments 1–4 were aimed at testing the validity of the three hypothesised category types.

We propose that some classifier categories constitute 'Well-Defined' categories. These are ones in which all objects that are referred to using a given classifier share one or more features; the feature(s) fully define the scope of the category. For example, *ke* is used to classify all and only green, growing things with stems and leaves. Likewise, *ben* is used only with objects that are bound into a book-like form, including books, notebooks, textbooks, dictionaries, handbooks, user's manuals, atlases, and calendars in the form of a book. The features that define the category can often, although not always, be deduced from the non-classifier use of the word. Regardless, native speakers have a clear intuition of associated knowledge when hearing the classifier word, knowledge that fully defines the scope of the category.

We propose that a second subset of the classifier categories is 'Prototype' categories. These are ones in which there is a typical sort of thing associated with the category, which can often be deduced from the non-classifier use of the word. Furthermore, native speakers consistently associate certain features with the classifier word based on these exemplars and/or the non-classifier use of the word. Besides the typical things, though, the categories include other objects that are more distant from the central features or prototype evoked by the classifier word. For some of these objects the resemblance to the prototype is relatively strong, and for other objects, the relation to the category is more distant, producing a gradient of typicality. For instance, *ke* (written with different characters from *ke* used for plants) classifies pearls, peas, soy beans, teeth, buttons, bullets, stars, man-made satellites, and bombs. The prototype appears to be a small, three-dimensional, roundish object. Pearls, peas, and soy beans are all quite close to this prototype. Others such as buttons, bombs, mines, stars, and satellites are somewhat more removed, but their connection can be fairly readily perceived. A mine is not very small, but it is normally three-dimensionally round. A satellite or star is definitely not small, but when viewed from far away (as people do), they do appear to be very small, like a pearl. The classifier *tiao*, often said to classify long, thin things (e.g., Saalbach & Imai, 2007; Schmitt & Zhang, 1998; Zhang & Schmitt, 1998), does include many objects that closely fit this prototype, such as ropes, braids, snakes, and belts, but it also includes many that fit it less well. For instance, fish, towels, and trousers have some resemblance to this shape. Furthermore, *tiao* classifies some entities – such as ships, news, brave men, experience, blankets and sheets, and mini-skirts – that have no transparent connection to the prototype or other category members. Thus the native speaker knowledge associated with the classifier word provides only a notion of a category prototype, and the complete membership must be mastered on the basis of partial overlap with it, sometimes supplemented by memorisation.

Last, we propose that some classifier categories are 'Arbitrary' categories for which there is no set of defining features nor any prototype evoked by the classifier word giving a sense of coherence to the set of objects classified by the classifier. For instance, *zun* is commonly used to classify two types of objects only – large guns and statues of Buddha. Another example is *zhuang*, used with (business-type) deals, cases (as in murder cases), and something on one's mind/a worry. Historically there may have been motivated links from a prototype or from one another, but the non-classifier use of the words do not provide any clear indications of the basis of the category. In some cases, there may be perceivable connections between some members of the category (for instance, *zong* classifies business deals and large sums of money; *bu* classifies films, literary works and long novels, as well as telephones), but native speakers do not have a clear intuition of features associated with the classifier word. To the current speaker of Chinese, the connection of the classifier to membership in these categories may be perceived as arbitrary and learnable only through memorisation.

We note that classifier categories vary on additional dimensions – such as category size and the type of semantic feature(s) that link members of a category – besides the ones captured in our three proposed types. This variation is surely reflected in the mental representations of native speakers. However, it is not clear what implications, if any, variations on those dimensions have for understanding whether classifier categories constitute meaningful conceptual categories or create potential for cognitive consequences outside of the language arena. Perhaps of greater interest for our key issues is to consider whether our Well-Defined categories are better described as taxonomic categories, as suggested by the observation that some of the categories in this group can informally be described as plants, vehicles, books, and guns. However, exactly what constitutes a taxonomic category to cognitive psychologists is not clear-cut. They are sometimes defined as stable, established groupings in memory (Barsalou, 1983; Ross & Murphy, 1999) in contrast to ad hoc or goal-derived ones. The organisation of groupings into multi-level hierarchies also is often invoked (e.g., Markman, 1989; Murphy, 2002). Perhaps most important is the notion that they group together things that share some sort(s) of intrinsic features rather than external relations such as 'used together' or 'next to' (Markman, 1989). But there are many ways that things can be grouped by such features, because they can share properties that are perceptual, functional, or more abstract, or a combination (Markman, 1989). Many classifier categories across our three types meet one or more of the various criteria offered for taxonomic status. In practice, psychologists have tended to take groupings of objects labelled by English nouns to be taxonomic categories and to assume that those labelled by phrases (such as *things to take on a picnic* or *green things*) or by other linguistic devices, such as classifiers, are of some other sort. However, the notion of using nouns of any language to identify what conceptual groupings are taxonomic implies universality of the conceptual groupings, which in many domains does not actually exist. For instance, the particular groupings of artifacts picked out by each noun in the domain of household containers varies widely across languages (Malt, Sloman, Gennari, Shi, & Wang, 1999), as do those picked out by names for drinking vessels (Kronenfeld, Armstrong, & Wilmoth, 1985). Finally, as implied by this last observation, although some classifier categories resemble groupings labelled by English nouns, they are not necessarily fully comparable; for instance, *ben* can be summarised as applying to books, but English does not have a single word that groups magazines with novels and dictionaries and other bound materials, and *ting* may be summarised as applying to guns, but not all things called *gun* in English are classified by *ting*.

We therefore suggest that the diversity of the category members, and how tightly linked the members are to the knowledge evoked by the classifier word, better capture the distinction between the Well-Defined and other two category types we describe than invoking the notion of taxonomic categories. Furthermore, the underlying variables of degrees of relatedness of category members to one another and to featural knowledge evoked by the classifier word are more directly relevant to understanding to what extent classifier categories are meaningful to native speakers. Likewise, these variables are more transparently relevant to implications for cognitive consequences since they may directly influence the possibility of a stimulus in the world evoking a classifier category and its processing or encoding being influenced by classifier category membership. We believe our description to have greater explanatory power as a result. We do note, though, that it is no doubt more than coincidence that our Well-Defined categories correspond more strongly overall to categories labelled by English nouns than do those in either of the other two category types. Nouns, while capturing groupings that often vary across languages, nevertheless usually capture groupings of things that overlap on multiple dimensions (Rosch & Mervis, 1975), thereby limiting the perceived diversity of the category members and providing a basis for a strong link between knowledge of members and the knowledge evoked by the noun itself. Not all our Well-Defined classifier categories necessarily correspond to groupings named by the nouns of any particular language, but they are more likely to resemble such groupings than those of our other types.

If the three distinctions in category type that we focus on have psychological reality, then the differences should be reflected in tasks designed to tap the knowledge native speakers have about the categories. In Experiment 1, we examined the mean and variability in typicality ratings for the three proposed category types. In Experiment 2, we examined the number of features produced to classifier words and extent of agreement in feature listings for the three hypothesised types. In Experiment 3, we tested the ability to identify the classifier from the features generated in Experiment 2, and in Experiment 4 we gathered judgements of to what extent the nouns in a category embody the central features produced in Experiment 2. We discuss predictions for the different classifier types for each task below.

EXPERIMENT 1: TYPICALITY

If mental representations of classifier categories differ in their nature as we hypothesise, these differences should be apparent in typicality ratings. Members of Well-Defined categories should all fully meet the criteria for membership and they should therefore all be viewed as relatively high in typicality. Consequently there should be relatively little variability in the typicality ratings for their members. Members of these categories may nevertheless show some degree of variability because frequency or familiarity of objects can influence typicality ratings (Armstrong, Gleitman, & Gleitman, 1983; Ashcraft, 1978; Malt & Smith, 1982) in addition to object features.

Members of Prototype categories should average lower in typicality than members of Well-Defined categories, since they will have substantial range in their closeness to the category prototype. By the same token, the ratings should show a higher degree of variability.

Expectations for typicality ratings for Arbitrary categories are less clear cut. By hypothesis there are no semantic features circumscribing category membership, so it is impossible to predict how high or low on average the ratings will be. However, we can predict that any typicality variability that does exist will be driven by the other sorts of variables known to affect typicality ratings, namely, familiarity or frequency. Because it would have been difficult to pose straightforward questions to participants about the familiarity or frequency of some of the classifier category members (it is not clear how one would judge the familiarity or frequency of a brave man or a slogan or news), we chose to evaluate this possibility by asking for judgements of the frequency of particular classifier-noun combinations. A noun that is rarely encountered, either because its referent rarely occurs in Chinese speakers' environment, or because the referent is rarely discussed, should yield a lower classifier-noun frequency rating than one more commonly encountered. Any variability in typicality that does exist may be related to this variable.

Experiment 1a: Typicality ratings

Method

Participants. Twenty native speakers of Mandarin Chinese from Beijing and nearby cities, all graduate students or spouses of graduate students at Lehigh University at the time of testing, were recruited via an advertisement on a computer bulletin board and were paid to participate.

Materials. Twenty-four classifiers were selected from the list in Appendix A, eight to represent each proposed type. Because some classifiers could be considered borderline between two different types (see General Discussion), the eight selected to represent each type were chosen as ones that seemed to be clear cases. Each classifier was paired with up to 18 nouns. For most classifiers, the total number of nouns that had been collected was 18 or fewer, and all the nouns collected were included (along with a few others observed in the interim: 'desk lamp' for *zhan*, 'phonebook' for *ben*, and 'pilot' for *ming*). For three classifiers, one item each from the original compilation was accidentally omitted ('soup' for *di*, 'pencil' for *gen*, and 'ladder' from *jia*); all others were included. For two classifiers that are used with a large number of nouns that cannot be exhaustively listed (*ke*, used for most things called *plant* in English, and *duo*, used for most things called *flower* in English), the original compilation for Appendix A had not listed individual exemplars. Nouns (12 and 8 respectively) for these were generated by author Gao, attempting to reflect the diversity of exemplars possible (see Table 1 for complete stimulus list). Booklets were prepared listing each of the eight classifiers (in Chinese), with each one followed by the list of nouns. Four classifier lists appeared each page, and the order of pages varied across participants.

Procedure. Participants were run individually or in small groups. They were given written instructions in English asking them to imagine that they

TABLE 1
Stimuli for all experiments

Classifier type	*Classifier*	*Nouns*
Well-Defined	Ke (棵)	树 (tree), 草 (grass), 玉米 (corn), 白菜 (cabbage), 松树 (pine tree), 柳树 (willow tree), 果树 (fruit tree), 梨树 (pear tree), 苹果树 (apple tree), 古树 (ancient tree), 高树 (tall tree), 矮树 (short tree)
	Liang (辆)	公共汽车 (public bus), 小汽车 (car), 卡车 (truck), 自行车 (bicycle), 吉普车 (jeep), 拖拉机 (tractor), 火车 (train) 坦克车 (tank)
	Zhan (盏)	油灯 (oil lamp), 电灯 (electric lamp), 台灯 (desk lamp), 日光灯 (fluorescent lamp)
	Di (滴)	水 (water), 油 (oil), 眼泪 (tear drop), 血 (blood), 汗 (sweat), 口水 (saliva), 酱油 (soy sauce), 醋 (vinegar)
	Ben (本)	书 (book), 杂志 (magazine), 画报 (pictorial), 小说 (novel), 字典 (dictionary), 电话簿 (phonebook)
	Duo (朵)	花 (flower), 白云 (white cloud), 菊花 (chrysanthemum), 玫瑰花 (rose), 玉兰花 (magnolia), 纸花 (paper flower), 大红花 (big red flower), 小白花 (small white flower)
	Ming (名)	教师 (teacher), 教授 (professor), 护士 (nurse), 医生 (doctor), 科学家 (scientist), 律师 (lawyer), 记者 (journalist), 工人 (worker), 学生 (student), 作家 (writer), 战士 (soldier), 演员 (actor/actress), 政客 (politician), 警察 (policeman), 海员 (sailor), 飞行员 (pilot)
	Ting (挺)	步枪 (rifle), 机关枪 (machine gun), 冲锋枪 (assault rifle)

(*Continued*)

TABLE 1 (*Continued*)

Classifier type	*Classifier*	*Nouns*
Prototype	Gen (根)	棍子 (stick), 筷子 (chopstick), 稻草 (straw), 一根蜡烛 (candle), 手指 (finger), 头发 (hair), 针 (needle), 线 (thread), 绳子 (rope), 神经 (nerve)
	Jia (家)	人家 (household), 商店 (store), 饭馆 (restaurant), 超级市场 (supermarket), 银行 (bank), 电影院 (cinema), 医院 (hospital), 工厂 (factory), 公司 (company), 报社 (newspaper publisher), 电台 (radio station), 旅馆 (hotel), 旅行社 (travel agency), 出版社 (publishing house)
	Ke (颗)	珍珠 (pearl), 黄豆 (soy bean), 钮扣 (button), 牙 (tooth), 地雷 (land mine), 子弹 (bullet), 炸弹 (bomb), 星星 (star), 人造卫星 (man-made satellite)
	Tou (头)	猪 (pig), 鹿 (deer), 牛 (cattle), 毛驴 (donkey), 狮子 (lion), 大象 (elephant), 蒜 (garlic)
	Ba (把)	雨伞 (umbrella), 手枪 (pistol), 茶壶 (tea pot), 刀 (knife), 螺丝刀 (screwdriver), 剪子 (scissors), 钳子 (pliers), 锤子 (hammer), 勺 (spoon), 笤帚 (broom), 小提琴 (violin), 椅子 (chair), 钥匙 (key), 尺子 (ruler)
	Tiao (条)	绳子 (rope), 线 (line), 辫子(braid), 蛇 (snake), 鱼 (fish), 小溪 (stream), 河 (river), 水渠 (canal), 毛巾 (towel), 路 (road), 裤子 (trousers), 裙子 (skirt), 毯子 (blanket), 标语 (slogan), 新闻 (news), 经验 (experience), 人命 (human life) 好汉 (true man)
	Jia (架)	飞机 (airplane), 航天飞机 (space shuttle), 直升飞机 (helicopter), 眼镜 (eye glasses), 机器 (machine), 钢琴 (piano), 手风琴 (accordion), 电子琴 (electronic keyboard), 照相机 (camera)

(*Continued*)

TABLE 1 (*Continued*)

Classifier type	*Classifier*	*Nouns*
	Tai (台)	机器 (machine), 电视机 (TV set), 录音机 (recorder), 收音机 (radio), 机车 (locomotive), 拖拉机 (tractor), 电子计算机 (computer), 节目 (stage performance)
Arbitrary	Bi (笔)	交易 (business deal), 钱 (money), 现金 (cash), 款 (fund), 开支 (expense)
	Bu (部)	电影 (film), 文学著作 (literary work), 长篇小说 (novel), 电话 (telephone set)
	Zun (尊)	大炮 (big gun), 佛像 (figure of Buddha)
	Dao (道)	墙 (wall), li笆 (twig fence), 门 (door), 防线 (line of defense), 菜 (dish), 手续 (procedure), 阳光 (sunlight)
	Dun (顿)	饭 (meal)
	Ju (具)	尸体 (corpse), 棺材 (coffin)
	Zhuang (桩)	大事 (big matter), 小事 (small matter), 案子 (case), 买卖 (business deal), 心事 (load on one's mind)
	Zong (宗)	买卖 (business deal), 巨款 (large sum of money)

needed to teach American students studying Chinese about classifiers. They were further told that because it is often difficult to summarise the use of a classifier in words, it is helpful to show students good examples of its use. They were asked to first consider, for each classifier shown, what it means or what idea it conveys when used as a classifier, and then to indicate, for each noun in the list, how well it reflects the idea expressed by the classifier (regardless of how often the noun occurs with the classifier). They were asked to give their answers using a scale of 0 to 7, where 0 meant that the noun listed would not be used with that classifier and the numbers 1 to 7 indicated how well the noun reflected the idea of the classifier, from 'very poorly' to 'perfectly'. The instructions emphasised that participants could select all high or all low numbers, or a combination of numbers, for the nouns of a given list.

Results and discussion

Mean typicality ratings and standard deviations for the three category types are given in Table 2. As predicted, the nouns of Defining Features categories showed high mean ratings, close to the top of the scale, and the nouns of Prototype categories (as well as of Arbitrary categories) received lower mean ratings. The effect of category type was significant, $F(2, 38) = 29.46$, $MSe = 0.16$, $p < .0001$, and pairwise comparisons confirmed a significant different between Well-Defined and Prototype categories, $F(1, 38) = 44.67$, $p < .0001$, as well as between Well-Defined and Arbitrary, $F(1, 38) = 43.71$, $p < .0001$. Thus, it seems that the nouns associated with Well-Defined classifier categories more clearly embody the semantic features conveyed by the classifier word than do those of the other two types.

Also as expected, there were significant differences in the extent of variability in ratings for nouns of the different category types (as reflected in standard deviations), with Well-Defined categories showing the least variability. The effect of category type on standard deviation was significant, $F(2, 38) = 35.03$, $MSe = 0.09$, $p < .0001$, and pairwise comparisons confirmed a significant difference between Well-Defined categories and each of the other two, $F(1, 38) = 35.26$, $p < .0001$ for Prototype and $F(1, 38) = 65.26$, $p < .0001$ for Arbitrary. This outcome supports the idea that the set of categories we have designated as Well-Defined contain members that meet some clear criterion for category membership.

One might ask whether the effects we have found are consistent across the different classifiers representing each of the three hypothesised category structures. With only eight items (classifiers) per category type, traditional item analyses would lack power and be uninformative. Because we use the same items in Experiments 2–4 to further test the hypothesis of distinguishable category types, it is more informative to examine the pattern of results by items across the four experiments to see if any items systematically deviate from the others across measures. We therefore defer consideration of this question until the results of Experiments 1–4 have all been presented.

TABLE 2
Mean typicality ratings (and standard deviations) as a function of classifier type, Experiment 1a

	Well-Defined	*Prototype*	*Arbitrary*
Typicality			
Mean	6.45	5.62	5.63
SD	0.39	0.63	0.73

Although we did not have specific predictions for Arbitrary categories on these two measures, it may seem surprising that the mean typicality rating and standard deviation for these categories were similar to those of the Prototype categories. If participants' ratings of typicality reflected solely the extent to which an object tends to have the features conveyed by a classifier, and if Arbitrary classifiers convey little semantic information and/or their members have little relation to that information, one might expect low typicality ratings with little variance. However, as noted above, past research has found that variables such as familiarity and frequency can influence typicality ratings. In the absence of meaningful featural information, ratings may be driven by such variables. Experiment 1b examined the extent to which the ratings of Experiment 1a might be related to frequency for all three category types.

Experiment 1b: Frequency ratings

Ratings of the frequency of each classifier-noun combination were collected in order to determine whether these frequency judgements were correlated with typicality. If the typicality variation for Arbitrary categories is due to this variable, the correlation should be strong for this category type. Past work (Malt & Smith, 1982) has shown that significant variation in perceived typicality exists for categories known to be meaningful – those labelled by nouns – independent of the contribution from frequency, and so if Prototype classifiers carry more inherent meaning for participants, the typicality range observed in Experiment 1a should be less strongly related to frequency.

Method

Participants. Twenty native speakers of Mandarin Chinese from Beijing and nearby cities, all graduate students or spouses of graduate students at Lehigh University at the time of testing, were recruited via an advertisement on a computer bulletin board and were paid to participate. Most had not participated in Experiment 1a.

Materials. Booklets were the same as in Experiment 1a.

Procedure. Instead of typicality, participants judged how frequently each of the classifier-noun combinations is used in modern Chinese. Responses were made on a 0–7 scale where 0 meant that the combination was not used and the numbers 1 to 7 indicated values from very infrequently to very frequently.

TABLE 3
Mean frequency ratings (and standard deviations) as a function of classifier type, Experiment 1b

	Well-Defined	*Prototype*	*Arbitrary*
Frequency			
Mean	6.10	6.22	5.75
SD	0.68	0.65	0.85

Results and discussion

Mean frequency ratings and standard deviations are given in Table 3. Prototype classifier-noun combinations were rated most frequent and Arbitrary classifier-noun combinations least, and the difference among the types was significant, $F(2, 38) = 8.69$, $MSe = 0.14$, $p < .001$. The key purpose of the ratings, however, was to evaluate their relation to typicality. The correlation of typicality and frequency ratings for Well-Defined categories was .7; for Prototype categories it was .33; and for Arbitrary categories it was .62, $p < .01$ for all. The substantial correlation for Arbitrary categories compared with Prototype categories suggests that feelings of frequency or familiarity did play a more major role in creating variability in typicality ratings for Arbitrary categories, with the variability for Prototype categories presumably more strongly driven by semantic features. The high correlation of frequency and typicality judgements for Well-Defined categories suggests that much of the variation in typicality judgements for these categories was also due to feelings of frequency or familiarity. This differential pattern of correlation across the three category types, in combination with the pattern of typicality rating means and standard deviations, supports the notion that the classifiers we considered to pick out Well-Defined categories do classify objects that meet clear criteria for membership (with very limited typicality variation resulting primarily from frequency/familiarity), those we considered to pick out Prototype categories classify objects that vary in their featural relation (and hence typicality) to a central idea conveyed by the classifier word, and those we considered to pick out Arbitrary categories classify objects with less semantic relation to the classifier word but with variable typicality resulting primarily from frequency/familiarity variation.

EXPERIMENT 2: FEATURE LISTING

Experiment 2 was designed to investigate more directly the nature of featural information associated with the three proposed category types. Features have proven to have substantial predictive value in accounting for performance on a variety of category-related tasks including naming choices (Sloman, Malt, & Fridman, 2001), word priming (McRae, deSa, & Seidenberg, 1997), similarity judgements (Tversky, 1977) and the gradual convergence of children's word use with adults' (Ameel, Malt, & Storms, 2008). Thus although features may not capture all knowledge associated with a category name (e.g., Ahn & Kim, 2000; Murphy & Medin, 1985), they stand to provide useful insight into the mental representations of classifier categories.

Participants were asked to list central features associated with each of the classifier categories studied in Experiment 1. If mental representations of Chinese classifier categories differ in a way that corresponds to the three hypothesised types, their responses should show different patterns for the diversity of features listed across participants, and also for the number of participants listing the single most frequently mentioned feature. For diversity of features listed, we predicted that Well-Defined categories would result in the smallest total number of different (unique) features produced across participants, because there should be a clear set of criterial features that come to mind when thinking of the classifier and its members. For Prototype categories more different features should come to mind across participants due to the range of membership. For Arbitrary categories, since there is no central tendency to guide retrieval of features from memory, an even larger number might emerge across participants. Conversely, we predicted that agreement among participants on the single most frequently listed feature will be lowest for the Arbitrary categories since there is no central idea to guide retrieval, and it will be highest for the Well-Defined ones.

Method

Participants. Twenty native speakers of Mandarin Chinese from Beijing and nearby cities, all graduate students or spouses of graduate students at Lehigh University at the time of testing, were recruited via an advertisement on a computer bulletin board and were paid to participate. Most had not participated in Experiments 1a or 1b.

Materials. The Chinese characters indicating each of the 24 classifiers used in Experiment 1 was printed at the top of a piece of paper. Below it the participant was asked to indicate if he or she was familiar with its use as a numeral classifier. Below that question was the one of key interest asking the participant to record in Chinese the features of things classified by this classifier. The order of classifier pages in the booklets was varied randomly.

Procedure. Participants were run individually or in small groups. As in Experiment 1, the task was framed in the context of teaching English speakers about classifiers. Participants were asked to imagine that they were helping to develop a new Chinese language textbook and were told that they would be asked to think of objects that go with certain classifiers and then write down their features, so that learners would know what sort of objects could go with the classifiers. Examples of the sorts of features that might be listed were given for three classifiers not among the 24 being studied.

Results and discussion

Because our prediction for feature diversity concerned the number of different (unique) features produced across participants for a category type (not the number of features individual participants produced), this analysis used items rather than participants as the unit of analysis. The number of unique features was tabulated for each classifier within a category type and the mean across classifier stimuli for that type was obtained. Mean numbers of unique features listed and standard deviations for the three category types are given in Table 4. The number of participants who listed the most frequently produced feature for each type (averaged across the eight classifiers in the category type) is also given. There was a significant effect of category type on the mean number of unique features listed, $F(2, 21) = 4.07$, $MSe = 26.38$, $p < .05$, with Arbitrary categories resulting in the largest number and Well-Defined categories resulting in the fewest, as predicted. The number of unique features listed for Well-Defined categories differed significantly from the number for Arbitrary categories, $F(1, 21) = 7.43$, $p < .05$, and the number for Prototype categories also differed significantly

TABLE 4
Mean number of unique features generated and mean number of participants listing the most frequent feature (and standard deviations) as a function of category type, Experiment 2

	Well-Defined	*Prototype*	*Arbitrary*
Unique features generated			
Mean	14.00	15.63	21.00
SD	3.66	5.07	6.32
Participants who listed the most frequent feature			
Mean	13.50	13.87	7.75
SD	4.17	4.16	2.76

Note: Measures are calculated across items in a category type (see text).

from the number for Arbitrary ones, $F(1, 21) = 4.38$, $p < .05$, although the number for Well-Defined did not differ significantly from that for Prototype.

Also as predicted, the number of participants who listed the single most frequently given feature for classifiers within each category type showed a reversed pattern. Agreement was lowest for Arbitrary categories, and it was higher for Well-Defined and Prototype categories, producing a significant effect of category type, $F(2, 21) = 6.68$, $MSe = 14.11$, $p < .01$. As we expected, it seems that when people try to think of features of Arbitrary category classifiers, they retrieve diverse pieces of information and do not converge on a shared set as most important. Slightly inconsistent with our expectation, agreement was about the same for Well-Defined and Prototype categories. Although we had expected good agreement for Well-Defined categories, it seems that the central information associated with Prototype categories is also constrained enough that equally good agreement on the most frequent feature results. (This may be because the information retrieved is primarily that associated with typical exemplars; the next experiment demonstrates that this information is, nevertheless, less tightly tied to the classifier word than is the case for Well-Defined categories.) The difference between Well-Defined and Arbitrary was significant, $F(1, 21) = 9.37$, $p < .01$, as was the difference between Prototype and Arbitrary, $F(1, 21) = 10.63$, $p < .01$, although the difference between Well-Defined and Prototype was not. This outcome along with the results for number of features listed strongly supports our expectation that the features that come to mind for Well-defined categories are relatively limited and consistent across participants, while those that come to mind for Arbitrary categories are diverse and less constrained.

EXPERIMENT 3: IDENTIFYING CLASSIFIERS FROM FEATURES

Experiment 3 was designed to confirm and validate the findings from Experiment 2 by determining whether participants could identify the intended classifier category from the features most frequently generated by participants in Experiment 2. The features were presented to participants one at a time, beginning with the most frequently generated, so that we could evaluate not only whether participants could determine the correct classifier but how many features were required before they could identify the classifier. If our analysis of the classifier types is correct, participants should be able to identify Well-Defined classifiers most successfully, and from the fewest features, because the features involved specify the category fully. Participants should be able to identify the Prototype classifiers fairly well, but not as well as the Well-Defined ones, because the features listed may not be quite as closely tied to the classifier word itself. They should have the hardest time identifying the Arbitrary classifiers since the features are most diverse and least closely linked to the classifier.

Method

Participants. Twenty college-educated native speakers of Mandarin Chinese from Beijing and nearby cities, all living in eastern Pennsylvania, USA, at the time of testing, participated without compensation. They were referred for participation through friends or acquaintances and were contacted by phone or e-mail. None had participated in Experiments 1 or 2.

Materials. For each of the 24 classifiers used in the previous experiments, the four most frequently mentioned features from Experiment 2 were chosen. (In cases of ties, the feature that had appeared first on the tally list was used.) Each set of four features for a given classifier was printed on a sheet of paper, listed in the order of frequency in Experiment 2. To the right of each feature appeared a blank space for participants to record their guess about the classifier. Booklets were assembled containing all 24 classifier lists, with the order of pages varied across participants. The complete set of features used for each classifier, in order of frequency, is given in Appendix B.

Procedure. Participants were run individually or in small groups. They were told that the features on each page described certain objects or ideas, and that their task was to determine which classifier should be used for the objects or ideas described. The features on the first page were covered with a sheet of paper. When the experimenter signalled the participant to start, the participant slid the sheet down so that the first feature was revealed. The participant then either recorded a possible classifier in the blank next to it and slid the sheet to expose the next feature, or decided that she had no guess and moved to the next feature without recording a response. This procedure was repeated for each feature for all of the 24 classifiers.

Results and discussion

The maximum score for each participant for each of the three classifier types was eight, since there were eight classifiers presented per type. The mean scores and standard deviation for each type after reading only the first feature, and after reading all four features, are given in Table 5. As predicted, the mean number of correct guesses after reading only one feature showed a significant effect of classifier type, $F(2, 38) = 78.44$, $MSe = 0.79$, $p < .0001$, with the greatest number of correct guesses for Well-Defined categories and the fewest for Arbitrary. Each of the three means differed significantly from

TABLE 5
Mean number of items correctly identified after reading one feature and after reading all four features (and standard deviations) as a function of category type, Experiment 3

	Well-Defined	*Prototype*	*Arbitrary*
Number of classifier categories correctly identified after reading one feature			
Mean	4.85	2.55	1.40
SD	0.88	1.15	0.68
Number of classifier categories correctly identified after reading all the four features			
Mean	5.75	4.50	2.65
SD	1.02	1.47	1.20

the other two (Well-defined vs. Prototype, $F(1, 38) = 67.23$, Well-defined vs. Arbitrary, $F(1, 38) = 151.27$, and Prototype vs. Arbitrary, $F(1, 38) = 16.81$, all p values .001 or less).

Also, as predicted, the mean number of correct guesses after reading all four features showed a significant main effect of classifier type, $F(2, 38) = 34.64$, $MSe = 1.40$, $p < .0001$, with the largest number of correct guesses for Well-Defined categories and the fewest for Arbitrary. Each of the three means differed significantly from the other two (Well-defined vs. Prototype, $F(1, 38) = 11.13$, Well-defined vs. Arbitrary, $F(1, 38) = 68.43$, and Prototype vs. Arbitrary, $F(1, 38) = 24.37$; all p values .01 or less).

The differences in performance across the three category types in this task were quite striking. Participants apparently gained substantial information relevant to determining the correct classifier from just a single feature for Well-Defined categories but relatively little for the other two types of categories. This outcome is consistent with the idea that Well-Defined categories are well-specified by a small number of identifiable features. After all four features, the number of classifiers participants could correctly identify for Prototype categories came closer to that of Well-Defined but still lagged a bit, and the Arbitrary categories remained relatively low even under this circumstance. This outcome supports the proposal that central features for Prototype categories are fairly closely tied to the classifier but determine it less well than the features associated with Well-Defined categories do, and that it is difficult or impossible to specify features of Arbitrary categories that pick out the classifier in a reliable way.

EXPERIMENT 4: JUDGING HOW MUCH AN OBJECT EMBODIES CENTRAL FEATURES

The previous three experiments involved direct consideration of the classifier words themselves, and the last two involved explicit consideration of the relation of semantic features to classifiers. If the hypothesised three types of classifier categories exist, then one would also expect that the central features generated to classifier names in Experiment 2 would be viewed as true of the individual objects or nouns that refer to them constituting the category members to differing degrees across the types. In this experiment, participants were shown the nouns used in the previous three experiments and asked to judge to what extent each one embodied the features of the associated classifier as generated in Experiment 2. No mention was made of the classifiers themselves or of the phenomenon of noun classification. We expected that nouns belonging to Well-Defined classifier categories would be judged to have the four features generated for the relevant category to a very high degree. Nouns belonging to Prototype classifier categories should embody the features generated to their category to a more variable extent, with the result that the overall mean judgement would be somewhat lower than for Well-Defined Categories (and the standard deviation would be higher). Nouns belonging to Arbitrary classifier categories would embody the features less still, since these categories are not grounded in a set of features shared by all or a core set of members.

Method

Participants. Twenty native speakers of Mandarin Chinese from Beijing and nearby cities, living in eastern Pennsylvania, USA, at the time of testing, participated without compensation. They were referred for participation through friends or acquaintances and were contacted by phone or e-mail. Most had not participated in any of the earlier experiments.

Materials. The four central features generated for each classifier in Experiment 2 (and provided in Appendix B) were written at the top of a piece of paper, followed by each of the nouns used with the classifier in the previous experiments (but the classifier itself was not presented). A blank space for giving a rating was provided next to each noun. Each sheet also contained a 7-point scale, with 1 labelled ‘embodies the features to a very low degree’ and 7 labelled ‘embodies the features to a very high degree’. Booklets were assembled containing all 24 classifier lists in a random order.

Procedure. Participants were run individually or in small groups. They were asked to judge, using the scale given, how much each of the nouns listed on the page embodied the four features listed on the top portion of the page. They were asked to use 1 for nouns that embodied hardly any of the four features, 4 for one that embodied all four of the features but to a very low degree, and 7 for a noun that embodied all four features to a very high degree.

TABLE 6
Mean ratings (and standard deviations) of how much nouns embody central features as a function of category type, Experiment 4

	Well-Defined	*Prototype*	*Arbitrary*
Mean	5.92	4.47	3.99
SD	0.44	0.42	0.57

Results and discussion

The mean embodiment rating and standard deviation across the eight classifiers for each type are given in Table 6. As predicted, there was a significant effect of classifier type, $F(2, 38) = 186.07$, $MSe = 0.1$, $p < .0001$, with the highest mean for Well-Defined categories and the lowest for Arbitrary. Each of the three means differed significantly from the other two (Well-defined vs. Prototype, $F(1, 38) = 192.97$, Well-defined vs. Arbitrary, $F(1, 38) = 343.65$, Prototype vs. Arbitrary, $F(1, 38) = 21.59$, all *p*s less than .0001). In addition, standard deviations varied as predicted, with members of Well-Defined categories showing less variability in means than members of Prototype categories. There was a main effect of category type on standard deviation, $F(2, 38) = 44.76$, $MSe = 0.06$, $p < .0001$. Well-Defined and Prototype categories differed significantly from one another, as we predicted, $F(1, 38) = 62.64$, $p < .0001$ (as did Well-defined and Arbitrary, $F(1, 38) = 71.35$, $p < .0001$). Prototype and Arbitrary categories did not differ from each other in degree of variability $F(1, 38) = 0.28$, $p > .60$; apparently some Arbitrary category members were felt to fit the listed features better than others despite the generally weak association of features to the members as shown in the previous experiments. Thus, this experiment supports the idea that members of the three different classifier category types share central features of the categories to differing degrees, with those in Well-Defined categories most consistently and strongly possessing those features.

ITEM VARIABILITY IN EXPERIMENTS 1–4

We now consider the question of whether the individual stimulus items (classifiers) within each hypothesised type behaved as we expected them to. Table 7 presents the mean value across participants for each item for the seven measures taken across the four experiments. To determine if any item did not behave in accord with its hypothesised type, we calculated the item mean and standard deviation for each measure and identified each score that was more than one standard deviation above or below the mean for that

TABLE 7
Item means across participants for measures from Experiments 1–4

	Expt. 1 Mean typicality rating/s.d.	*Expt. 2* Mean # unique feats. listed/# Ps listing most freq.	*Expt. 3* # correctly classified after 1st feature/ after 4th feature	*Expt. 4* Mean embodiment rating
Well-defined				
Ke	6.43 (1.06)	15/10	18/20	6.05
Liang	6.37 (1.29)	11/13	18/19	6.76
Zhan	5.61 (1.63)	18/10	20/17	5.5
Di	6.78 (0.55)	10/20	5/20	4.7
Ben	6.85 (0.44)	11/13	18/19	6.77
Duo	6.64 (0.70)	19/11	14/13	5.98
Ming	6.45 (1.08)	11/20	4/1	5.79
Ting	5.77 (2.17)	17/11	1/4	5.68
Prototype				
Gen	6.10 (1.32)	14/16	7/11	4.88
Jia	5.74 (1.23)	15/9	4/9	4.66
Ke	5.87 (1.45)	9/16	7/5	3.98
Tou	6.09 (1.51)	14/18	6/12	5.74
Ba	5.69 (1.69)	11/12	14/17	5.14
Tiao	4.95 (2.11)	17/20	10/7	3.29
Jia	5.27 (1.97)	20/10	3/19	4.4
Tai	5.89 (1.60)	25/10	0/9	4.65
Arbitrary				
Bi	6.33 (1.17)	12/12	4/11	5.79
Bu	5.63 (1.79)	30/6	7/11	3.76
Zun	6.13 (1.77)	24/8	0/0	5.65
Dao	5.09 (2.08)	26/9	0/4	2.39
Dun	6.85 (0.49)	12/10	0/8	5.6
Jü	6.18 (1.68)	22/8	19/15	4.28
Zhuang	5.49 (1.82)	20/3	1/3	3.76
Zong	4.48 (2.72)	22/6	1/2	3.25

measure. We then looked to see if the outliers were systematically associated with particular items. For the most part, outlier scores were distributed in a fairly scattered fashion, with each item showing outlying scores on no more than three of the seven measures. Only two items had outlying scores on a majority of the measures (four). Thus overall, consistency among items within the hypothesised types is reasonably good.

The first of the items showing relatively poor correspondence on the measures to the rest of its type was *ting,* placed in the Well-Defined group. In Experiment 1 *ting* had a lower mean typicality rating with higher variability than the other Well-Defined classifiers, and participants in Experiment 3 were not very successful at identifying the classifier from either the first feature given or all four features. We had considered *ting* to pick out a well-defined category because this classifier is used for various types of guns and only guns. However, the connection between the non-classifier meaning of the word (which is used as both an adjective and verb to mean 'straight', 'erect', or 'straighten') is not transparent, unlike in the other Well-defined cases, and the exemplars may have been viewed as somewhat low in typicality because they did not reflect properties associated with the classifier word through its other uses. (The high standard deviation may have been simply because there were few exemplars rated.) As for the poor ability to identify the classifier based on features, this outcome may have arisen because although *ting* classifies only guns, not all types of guns take *ting* as their classifier, and participants sometimes thought of the other relevant classifier when viewing the features. Overall, it seems that *ting* has a resemblance to classifiers in the Arbitrary group by virtue of lack of meaningful connection of the non-classifier use of the word to the entities classified, but otherwise resembles other classifiers of the Well-defined type.

The other classifier showing poor correspondence on four measures to its group was *tiao* in the Prototype type. *Tiao* showed lower typicality ratings with higher variance than the others, and a low mean rating of how much exemplars embodied the four listed features, despite high agreement among participants on the top listed feature and good ability to identify the classifier from the features. We interpret this pattern as consistent with our earlier comments about *tiao*; specifically, it is a category that has a clear prototype meaningfully associated with the non-classifier use of the word, but its exemplars are particularly diverse with some belonging to the category only by metaphorical extensions (which in some cases may not be transparent to current speakers). Thus *tiao* is correctly placed as exemplifying the Prototype type but represents a case of particularly high diversity in membership.

Two other items contained a smaller number of outlying scores that are worth remarking on. The first is *ming*, which we considered to be Well-Defined but for which participants were very poor at identifying the classifier from features. Although *ming* classifies only one clearly specifiable type of thing (people of different professions), there is another classifier, *wei*, that is used for very similar referents but under slightly different discourse conditions (when the emphasis is less on the profession itself). The features listed caused many participants to think of *wei* instead of *ming* (perhaps reflecting a difference in word frequency or some other variable affecting accessibility in memory). The low score thus does not necessarily argue against the status of *ming* as Well-Defined.

The second case of concern is *jü*, which we considered Arbitrary but for which participants were quite good at identifying the classifier from features. *Jü* shows a dissociation between the non-classifier use (where it means 'utensil' or 'tool') and the classifier use (in which it classifies corpses and coffins), but because the set of things it classifies is so small, the features listed pointed very clearly to that set of things. It is possible that this classifier is better considered Well-Defined if the classifier word in isolation activates knowledge of these features (in our feature listing task, however, participants were asked to think of exemplars of the category as well).

Because *ting* and *Jü* arguably are members of types other than the one had we placed each in, we recalculated the data from Experiments 1–4 omitting these two items. For Experiments 1, 2, and 4, the omissions shifted outcomes only a fraction of a decimal point (due to the small number of exemplars they contributed to the tasks) and in no way changed the patterns observed. In the case of Experiment 3, in which participants identified classifiers from features, the omissions strengthened the reported pattern considerably since the change entailed removing the item that had been least well identified for Well-Defined classifiers and the one best identified for Arbitrary classifiers. Thus conclusions from the experiments remain unchanged.

EXPERIMENT 5: CLASSIFIERS AND ORGANISATION IN MEMORY

Having demonstrated that classifiers differ in the knowledge that the classifier word evokes in isolation and in the relation of nouns used with a classifier to one another and to the knowledge evoked by the word, we now turn to our second key issue: whether this variation has implications for the potential for finding cognitive consequences of speaking a classifier language. If classifier categories differ in their mental representations, then the potential for cognitive consequences may vary depending on the particular categories at stake. Specifically, if some classifier categories constitute more coherent, meaningful groupings, labelled by words that invoke semantic knowledge more closely associated with the category members, those may be the ones most likely to produce effects on responding in tasks such as organising and retrieving information from memory, drawing inferences, judging similarity, and so on. In this experiment we chose to study the potential effects of classifier categories on the storage and retrieval of information from memory since (a) the presence of an alternative organisational system that cross-cuts the standard categories delineated by nouns such as *dog*, *ball*, and *bottle* will arguably most readily be seen in a task that centrally involves storing and retrieving information, and (b) our type distinctions based on degree of relatedness among category members and of the members to knowledge associated with the classifier word are particularly likely to be relevant to the storage and retrieval of information from memory. Because Zhang and Schmitt (1998; Schmitt & Zhang, 1998) have obtained positive results in a similar task, the existence of their work also allows a comparison of results when categories are distinguished by type.

Organisation of information in memory and its impact on retrieval can be demonstrated in a recall task. When items from different categories labelled by nouns (such as cities or flowers) are presented to people randomly intermixed, there is usually some clustering in the recall order (defined as the occurrence of sequences of two or more words from the same category; Bousefield & Cohen, 1955). That is, items from the same noun category tend to be recalled adjacent to one another, even though they were not adjacent in the input. Such clustering presumably reflects a tendency to notice the category relations among items in the list, to store them in a way that reflects these relations, and to retrieve them using the category names (e.g., *cities* or *flowers*) as retrieval cues. In our experiment, participants were exposed to sets of nouns from different classifier categories and were asked to recall the nouns. We evaluated whether they produced clusters of nouns corresponding to the classifier categories at a greater than chance level, and whether the degree of any such clustering differed depending on which of the three classifier category types was represented in the noun set.

Although Schmitt and Zhang (1998; Zhang & Schmitt, 1998) presented lists of bare nouns to participants in their clustering experiment, we chose to embed the nouns in sentence contexts. We wanted to be able to compare any impact of classifier category membership with and without the classifier actually present (which was possible through slight alterations in the sentence frame so that use of a classifier was or was not grammatically necessary). Because classifiers normally occur only in linguistic contexts, if a noun's classifier category membership affects how the noun is stored in memory, the effect may be strongest when the classifier has been linguistically present and activates knowledge of the classifier category membership. At the same time, however, presenting simple classifier-noun phrases (to contrast with bare noun lists) would draw a great deal of attention to classifier category membership, and any clustering effect would not necessarily reflect that present when processing nouns under more natural conditions. Our paradigm allows us to test for possible effects of classifier category membership under conditions similar to those in which nouns (with or without their classifiers) are normally encountered.

Because some classifier groupings do resemble groupings labelled by common nouns (for instance, those containing plants, animals, or vehicles), it is important to be able to distinguish any effect of classifier category membership from that of groupings that may be salient to speakers of non-classifier languages. For that reason, we also presented our sentences (translated into English) to a group of English-speaking participants for comparison.

If speaking a classifier language gives participants a way of organising and retrieving information that is distinct from and in addition to that of categories salient for other reasons, then we would expect Chinese participants overall to recall more items and to show more clustering of items related to classifier category membership than English-speaking participants. If the use of classifier category organisation is triggered primarily when explicitly processing classifiers, we would expect to see this effect more pronounced in the sentences with classifiers present. However, if these effects are modulated by the nature of the classifier category, so that only the more coherent and meaningful categories produce them, we would also expect to see an interaction with classifier category type such that there is little or no effect of classifier category membership with Arbitrary categories and the strongest effect with Well-Defined categories.

Method

Participants. Twenty-eight native speakers of Mandarin Chinese from Beijing and nearby cities, all living in eastern Pennsylvania, USA, at the time of testing, participated without compensation. They were referred for participation through friends or acquaintances and were contacted by phone or e-mail. Most had not participated in any of the previous experiments. Twenty-eight native speakers of English who were undergraduates at Harrisburg Area Community College served as the control group. They received extra credit toward their grade in a child development course for participating.

Materials. Each packet of materials contained four blocks of trials, with 16 sentences containing the target nouns in each block. In the block with Mixed classifiers, each of the 16 nouns came from a different classifier category. In the block with Well-Defined categories, nouns came from four different Well-Defined categories (four nouns from each); similarly, in the block with Prototype categories four nouns came from each of four Prototype categories, and in the block with Arbitrary categories four nouns came from each of four Arbitrary categories. Block type was a within-participant factor, with each participant receiving one Mixed, one Well-Defined, one Prototype, and one Arbitrary block. Presence or absence of classifiers in the sentences was a between-subjects factor, so all four blocks for a participant either contained explicit classifirs or did not.

The 16 sentences in each block were active voice sentences created from frames such as (in the English version) *Mary bought a [noun]* and *John made a [noun]*, with the nouns filled in according to block type. The verbs in the sentence frames were selected such that they would be able to form a meaningful sentence with many different nouns, so that the same 16 sentence frames could be used in each block. The same sentence frames were used in the Classifiers Present and the Classifiers Absent version of the stimuli, with one small modification. In the Classifiers Present version, the reference was to a single object (in English, for instance, *Mary bought a tree*), which required a classifier to be used before the noun in the Chinese. In the Classifier Absent version, number was not specified (in English, for instance, *Mary bought trees*), and so a classifier was not required in Chinese. Table 8 illustrates with 3 of the 16 English sentences from each block, with an indication of which Chinese classifier would be used in each in the Chinese version.

The classifier categories used were randomly chosen, within each type, from the pool of classifier categories used in the previous experiments. For the Mixed block, six Well-Defined, five Prototype, and five Arbitrary classifiers were randomly chosen and then one noun from each category that had been selected was randomly chosen. For the other blocks, four categories were randomly chosen and then four nouns from each were randomly chosen. Two random orders of sentences were used within each condition (defined by the combination of language, presence or absence of classifier, and block). Different participants received the four blocks in different orders (with sentence frames in the same order in each block for a participant).

The packets were organised into four sections corresponding to the four blocks. The first page of each was a sheet of paper with a rectangular window sized to exposure one of the 16 sentences at a time. The second page contained the 16 sentences. Next was a response sheet on which participants provided their recall. This was followed by a brief comprehension test containing three questions about the sentences of that block, designed to ensure that participants would pay attention to the whole sentence. The final page of each section (except for the last) was a set of simple math problems that served as a distracter task between blocks.

Procedure. Chinese participants were run individually or in small groups. English-speaking participants were run in a single group. In each group, participants were randomly assigned to either Classifiers Present or Classifiers Absent packets, with 14 participants in each.

The procedure was a modified version of that used by Bousfield, Sedgewick, and Cohen (1954) and Bousfield and Cohen (1955) (see also Daneman & Carpenter, 1980). Participants read instructions that told them they would be reading a series of sentences and that they should understand what the sentences said because there would be a comprehension test at the end. The instructions then told them that they needed to pay special attention to the last word of each sentence because they would be asked to recall it. The instructions explained how to use the cover sheet to expose one sentence at a time and told participants that they should study each one for a

TABLE 8
Sample sentences from each block, English version, Experiment 5

Mixed Classifiers Block
Classifiers Present version
Mary bought a tree. (classifier Ke, Well-defined)
John talked about a household. (classifier Jia, Prototype)
Mary sold a meal. (classifier Dun, Arbitrary)
Classifiers Absent version
Mary bought trees.
John talked about households.
Mary sold meals.
Well-defined Classifiers Block
Classifiers Present version
Mary bought a bicycle. (classifier Liang, Well-defined)
John talked about a student. (classifier Ming, Well-defined)
Mary sold a jeep. (classifier Liang, Well-defined)
Classifiers Absent version
Mary bought bicycles.
John talked about students.
Mary sold jeeps.
Prototype Classifiers Block
Classifiers Present version
Mary bought a chopstick. (classifier Gen, Prototype)
John talked about a supermarket. (classifier Jia, Prototype)
Mary sold a rope. (classifier Gen, Prototype)
Classifiers Absent version
Mary bought chopsticks.
John talked about supermarkets.
Mary sold ropes.
Arbitrary Classifiers Block
Classifiers Present version
Mary bought a fence. (classifier Dao, Arbitrary)
John talked about a deal. (classifier Bi, Arbitrary)
Mary sold a telephone. (classifier Bu, Arbitrary)
Classifiers Absent version
Mary bought fences.
John talked about deals.
Mary sold telephones.

Note: Material in parentheses identifies the classifier used in the Chinese version of each sentence and the classifier type it belongs to. Full stimulus sets contained 16 sentences per block.

few seconds until a bell rang, which would be their signal to slide the cover sheet down and begin studying the next sentence.

When all participants understood the instructions, the first block began, with the experimenter ringing the bell every six seconds until the 16 sentences of a block had been studied. Participants were then asked to turn to the next page and record the last words of the 16 sentences on the response sheet. After a minute and a half, the experimenter signalled them to move to the next page, which was the comprehension test. They were given one minute to complete this test and then signalled to move to the next page, which contained the math problems. Two minutes were allotted for the problems. The remaining three blocks were completed following the same procedure, except that no math problems followed the final recall test.

Results and discussion

If classifier categories help organise and retrieve information in memory, Chinese participants should recall more items overall than English speakers, and they should show more clustering of items related to classifier category membership than English speakers. If the use of classifier category organisation is triggered primarily when explicitly processing classifiers, these effects should be more pronounced in the sentences with classifiers present. However, if the effects are modulated by the nature of the classifier category, such that only the more coherent and meaningful categories produce them, we would expect to see clustering and a higher recall level most strongly for Well-Defined categories.

Table 9 presents the mean and standard deviations for number of items recalled for Chinese and English speakers for each block, with classifiers present and absent. An overall ANOVA on the number of items recalled showed a main effect of block, $F(3, 156) = 11.98$, $MSe = 2.83$, $p < .0001$, but no main effect of the presence or absence of the classifier in the sentence $F(1, 52) = 0.02$, $MSe = 0.16$, and no main effect of speaker, $F(1, 52) = 0.92$, $MSe = 6.45$. The lack of main effect of speaker indicates that Chinese speakers did not have higher levels of recall overall than English speakers did. Thus their knowledge of classifier categories did not across the board help them to better recall the items. The lack of a main effect of presence or absence of the classifier in the sentence across both groups of speakers is less meaningful, because we did not expect English speakers to show a difference. More important is whether this effect interacts with speaker group (Chinese vs. English). The data show a trend toward a differential impact on the two groups of speakers, with Chinese recalling slightly more items when classifiers were present and English speakers recalling slightly fewer, but the interaction of presence or absence of classifier with speaker was not significant, $F(1, 52) = 2.00$, $MSe = 14$, $p > .10$. Thus explicitly processing classifiers, overall, seems to at best improve recall for Chinese speakers only slightly. We defer considering the main effect of block and its relation to the speaker variable until after presenting the results for clustering.

Clustering was defined as the occurrence of sequences of two or more words in the same category in a participant's recall for a block. To code the

TABLE 9
Mean number of items recalled (and standard deviations) as a function of speaker group, block, and classifiers present or absent, Experiment 5

		Block			
Speaker	*Classifiers Present or Absent*	*Mixed mean SD*	*Well-defined mean SD*	*Prototype mean SD*	*Arbitrary mean SD*
Chinese	Classifiers Present	8.86	9.86	7.50	7.79
		2.28	2.32	2.14	1.93
	Classifiers Absent	8.29	9.14	7.14	7.21
		1.86	1.79	1.96	1.53
English	Classifiers Present	7.64	8.64	7.36	7.00
		2.27	1.55	1.91	1.57
	Classifiers Absent	8.57	8.43	8.14	7.29
		1.74	2.03	2.03	2.30

data, each time a noun was found occurring immediately before or after another noun from the same classifier category, one point was given to the noun. Thus a two-noun cluster received two points total and a three-noun cluster received three. Then we calculated the ratio of clustering to the total number of items recalled by the person for that block, reflecting the proportion of recall for that block that involved clustering. Data from the Mixed block were not coded, since there is no clustering possible.

Table 10 presents the mean and standard deviations for clustering scores for Chinese and English speakers for the remaining three blocks, with and without classifiers present. An overall ANOVA on the clustering scores showed a strong main effect of block $F(2, 104) = 12.03$, $MSe = 0.83$, $p < .0001$, and a main effect of speaker $F(1, 52) = 9.30$, $MSe = 0.64$, $p < .005$, but no main effect of the presence or absence of the classifier in the sentence, $F(1, 52) = 0.10$, $MSe = 0.01$. The main effect of speaker indicates that Chinese speakers did have higher levels of clustering than English speakers did. Because the previous analysis showed no overall difference in number of items recalled, this difference cannot be attributed to different memory abilities per se. Thus, overall knowledge of classifier categories seemed to result in more organised recall, if not greater overall levels of success in recalling items. The lack of a main effect of presence or absence of the classifier in the sentences across both groups of speakers is again not meaningful, because we did not expect English speakers to show this difference. Again, the data show a trend toward a differential impact on the two groups of speakers, with Chinese showing slightly more clustering when classifiers were present and English speakers showing slightly less, but the

TABLE 10
Mean ratio of clustering to total number of items recalled (and standard deviations) as a function of speaker group, block, and classifiers present or absent, Experiment 5

		Block		
Speaker	*Classifiers Present or Absent*	*Well-Defined mean SD*	*Prototype mean SD*	*Arbitrary mean SD*
Chinese	Classifiers Present	.81	.28	.37
		.19	.28	.32
	Classifiers Absent	.56	.32	.39
		.31	.26	.26
English	Classifiers Present	.39	.29	.27
		.26	.28	.24
	Classifiers Absent	.39	.43	.24
		.24	.26	.23

interaction of presence or absence of classifier with speaker was not significant, $F(1, 52) = 1.45$, $MSe = 0.10$, $p > .20$. Thus, overall, explicitly processing classifiers seems to increase clustering for Chinese speakers only to a small degree at most.

We now consider the significant main effects of block that emerged for both recall and clustering. For both measures, the most notable variation across conditions lies in the relation of Well-Defined categories to the rest: Well-Defined categories produced the greatest recall for both Chinese and English speakers, and they also produced the greatest amount of clustering for both Chinese and English speakers. To the extent that effects are the same for speakers of a classifier language and a non-classifier language, one must attribute them to something other than the status of the nouns as classifier category members. Membership in some Well-Defined categories overlaps to a substantial extent with membership in some categories labelled by nouns in both English and Chinese, so the extent of shared benefit may arise from their relation to the noun categories for speakers of both languages (see also Saalbach & Imai, 2007). We also note that, surprisingly, recall levels were second highest for both English and Chinese speakers for the Mixed block, in which nouns from 16 different classifier categories were presented. Because there were no evident groupings of the nouns available (either classifier or otherwise) to aid in recall, we must attribute the relatively good recall for this block to a particularly high level of memorability of either the individual nouns used in this condition or the combination of sentence frames with nouns. (For instance, *John dreamed about a corpse*, a Mixed stimulus sentence, may have been more memorable than *John dreamed about a business*, the corresponding sentence in the Arbitrary category block.)

To the extent that the patterns of differences in recall and clustering are shared by both Chinese and English speakers, we cannot rule out the possibility that the observed variation across blocks should be attributed to such extraneous effects.

However, two key findings indicate that a portion of the variation in performance across blocks differs for Chinese and English speakers, and hence this portion should be attributed to the knowledge of classifier categories. First, the main effect of speaker in the clustering results indicates that the higher levels of clustering shown by Chinese speakers are meaningful. Since the nouns and sentence frames were the same (except for language) for both speaker groups, the greater degree of clustering cannot be attributed to anything about the individual nouns or frames. Thus, classifier knowledge is implicated in the clustering. Second, the main effect of speaker in the clustering data is qualified by a significant interaction of speaker with block, $F(2, 52) = 6.33$, $MSe = 0.44$, $p < .002$, indicating that the Chinese advantage varies across blocks. As Table 9 shows, this interaction reflects primarily the fact that the clustering advantage for Well-Defined categories for Chinese speakers far exceeds that for English speakers. Indeed, pairwise comparisons show that (collapsing across presence and absence of classifiers), for Chinese speakers Well-Defined categories showed a highly significant difference in the level of clustering compared to both Prototype, $F(1, 104) = 29.66$, $p < .0001$ and Arbitrary, $F(1, 104) = 18.00$, $p < .0001$, whereas for English speakers the effects were minimal: Well-Defined categories did not differ significantly in clustering from Prototype, $F(1, 104) = 0.17$, and differed only marginally from Arbitrary, $F(1, 104) = 3.6$, $p < .06$. Chinese speakers seem to benefit from the classifier category membership of Well-Defined category nouns above and beyond any advantage these groupings afford for English speakers.

That this difference is tied to classification knowledge is also bolstered by the observation that the Chinese advantage for Well-defined categories is greater with the classifier present than absent, $F(1, 156) = 6.05$, $MSe = 0.42$, $p < .02$, although performance for the two versions of the English sentences was identical. The classifier category relations are more salient when the classifiers are present, and apparently Chinese speakers take advantage of the relations more readily in this case.

In sum, Chinese and English speakers showed broadly similar degrees of recall of nouns from sentence lists and similar degrees of clustering despite the availability of classifier category membership knowledge that Chinese speakers might have used in three of the four blocks to help store and retrieve the information from memory. Overall, the data are consistent with Saalbach and Imai's (2007) suggestion that knowledge of a classifier system may have a rather minimal impact on cognitive processes relative to the processes that speakers of classifier and non-classifier languages share. However, the data do reveal a noteworthy effect of classifier category membership in clustering for Chinese speakers under certain circumstances: namely, when the nouns involved come from Well-Defined classifier categories, and especially if classifiers are present in the sentences containing them. Even though the set of nouns involved in such cases may also be meaningful as a grouping to English speakers, the Chinese speakers seemed to be particularly sensitive to the existence of the category relations among the nouns and to make greater use of the presence of this relation in their recall. Experiments 1–4 provided clear evidence that Chinese classifier categories differ in the meaningfulness and coherence of the grouping to speakers of Chinese. The data from the current experiment demonstrate that these differences are reflected in the measured degree of impact of classifier category membership on task performance, especially when the presence of a classifier makes salient the relationship.

GENERAL DISCUSSION

We now return to the two broad questions we posed in the Introduction.

How meaningful are classifier categories?

Variability among classifier categories

At the outset we asked how meaningful classifier categories are, and we suggested that the truth may not be all-or-none: Classifier categories might vary in the nature of their mental representation within languages, as well as possibly between languages. The current investigation addressed the within-language possibility for Mandarin Chinese. Our studies provided converging evidence from several tasks that to current speakers, some classifier categories are well-defined, some have a prototype structure, and some are essentially arbitrary in terms of what they classify. It is possible that by taking a historical perspective one could provide an explanation for all classifier category memberships, but from the perspective of current speakers it seems that the links between objects and their classifiers vary considerably in transparency, and sets of classifier categories can be identified that differ from one another in the overall extent and variability of the transparency within the categories.

To the extent that classifier categories are meaningful, they provide an important case of cross-cutting category organisation (e.g., Ross & Murphy, 1999) that will need to be accounted for in models of knowledge for speakers of classifier languages, alongside the distinctly different organisation of the same entities given by knowledge of nouns. The need to be able to account for the co-existence of these cross-cutting organisations supports the value of approaches such as connectionism in which different patterns of activation can occur across the same representational units. Indeed, given that speakers of classifier languages successfully use classifier words even when the associated categories seem arbitrary to them, all classifier categories must in some way be accounted for in the representations. Being able to differentiate between representations that allow articulation of coherent, meaningful knowledge and those that provide for systematic usage without feelings of coherent knowledge may be an interesting challenge for such models. The fact that some of the classifier categories (mostly within the Well-Defined group, but to some extent also within the Prototype group) have a strong resemblance to groupings labelled by some Chinese nouns will also provide challenges, since patterns of activation may in some cases need to closely but not perfectly resemble those for nouns and others will need to be unique, yet all will need to be evoked in the appropriate grammatical contexts alongside activation of relevant nouns.

Further characterising the nature of classifier category representations

Our division of classifier categories into the three types was based on inspection of the sets of nouns presented in Appendix A and on author Gao's intuitions about the features shared (or not) among a given set, and the semantic information, if any, conveyed by their classifier. The converging evidence from the experiments provides independent confirmation of the reality of the proposed differences, but we do not have an objective method for determining what type every classifier falls into a priori. In addition, we studied only a small sample of the 126 classifiers that we determined to be familiar to modern speakers of Chinese. Further distinctions may be identifiable within the three types that we described, and some categories may sit at the borders between types. For instance, as our initial description of Prototype categories indicated, in some cases one can see a family resemblance among all the members of the category (e.g., for *ke*, one can make the case for some overlap of perceptual features of less typical members such as stars, satellites, and bombs with typical members such as pearls, peas, soy beans, and buttons), but for others, it is not clear what features the least typical members might share with moderately typical or highly typical members (e.g., for *tiao*, there is no apparent relation of ships, news, brave men, and mini-skirts to somewhat long, thin fish, towels, and trousers or clearly long, thin ropes, braids, and snakes). For the latter case, it may be that a subset of the members are treated as arbitrary, although most are treated as meaningfully related to the category. Similarly, *kou* for the most part classifies things that have a mouth-like opening, such as person, pig, pot, bell, and well for water, but it also classifies long, wide-bladed, sword-like knives, which seem to be arbitrary members. A different sort of case is *ma*, which is used only with a noun referring to a matter (as in an important matter). It could be said to belong to the Well-Defined type since there is no variability in category membership and the criterion for using the classifier is clear. However, the classifier word when used in its non-classifier sense means a number symbol, and when used as a classifier it seems to evoke little knowledge related to what the category membership consists of. From that perspective, *ma* could be considered an Arbitrary category. Thus the full set of classifiers is likely to exhibit a gradation across the three types that we described, with some being perhaps hard to categorise definitely as one or another. The item variability shown in the results of Experiments 1–4 suggests, not surprisingly, that our stimulus selection did capture some such differences.

Implications for learning Chinese

Our focus has been on understanding the classifier category knowledge of mature native speakers of Chinese. However, the outcome also has implications for language learning and teaching. If there are different types of categories, then different types of information need to be acquired in order to master the use of different classifiers. It may be helpful to second-language learners of Chinese to be explicitly taught about the variability in the nature of classifier categories, and to understand that they cannot always generalise use of a classifier to the same degree for all of the classifiers they learn. Without such an understanding, students may tend to either chronically over-generalise classifiers, thinking that all behave like Well-Defined ones do, or chronically under-generalise them, thinking that all contain a high level of unpredictability like Prototype or Arbitrary ones. Furthermore, it may be helpful in teaching individual classifiers to identify not only some semantic features associated with a classifier and examples of nouns that take the classifier, but to discuss to what extent the use of the particular classifier is fully described by these features and examples.

Because Chinese classifiers are also learned slowly and with some difficulty by native speaking children (Erbaugh, 1986), who do not master many of them until taught in school, the same sorts of information may be useful in instructing child native learners. In fact, instruction for native speakers teaches classifier use as if classifiers were purely grammatical devices, and this instruction does not impart to the children the idea that the set of nouns classified by a particular classifier can be thought of as constituting a category having, in many cases, some degree of coherence. As a result, native speakers generally acquire implicit knowledge of when to apply each classifier word but have little explicit understanding of the nature of the system. Greater explicit understanding of the nature of the system might speed children's learning, although empirical evaluation of their ability to benefit from such information would be needed.

Are there cognitive consequences of speaking a classifier language?

At the outset we noted that the existing evidence from work by Schmitt and Zhang (1998) and Saalbach and Imai (2007) is mixed regarding the extent of impact of classifier categories on non-linguistic cognitive processes. Our data from the memory task is consistent with Saalbach and Imai's finding that knowledge of classifier categories can influence performance in some tasks but its influence is quite modest in scope and may be secondary to other sources of influence on performance in the task. The data from Experiments 1–4 suggest that part of the reason for this limited influence may be that Chinese classifier categories are varied in the extent to which they represent meaningful, coherent groupings to native speakers. It should not be surprising if groupings labelled by a grammatical device carrying little or no meaning have little effect on the organisation of meaningful information in memory, especially given the availability of alternative bases for thinking about and organising the information.

Imai and Saalbach (in press) note that Japanese classifiers are used much less frequently and are less salient to Japanese speakers than is the case for Chinese, and so may have little influence on non-linguistic cognitive processes. Although Chinese classifiers may be salient compared to Japanese, one must likewise consider their larger context for a Chinese speaker. Several factors may militate against a strong impact on non-linguistic processes. First, as we have discussed, Chinese uses classifiers with nouns only in sentences in which a specific number of something is identified or a demonstrative is used. Nouns are frequently mentioned in sentence constructions that do not require use of any classifier. (Erbaugh, 1986, found that only about 3% of utterances in a sample of casual conversation with most referents physically present contained individual classifiers other than *ge*.) Second, in informal discourse when an individual classifier is needed, it is common for speakers to use the general classifier *ge*, which fulfills the grammatical requirement for a classifier but is semantically empty, instead of choosing a classifier that carries some degree of meaning. Third, as our Experiments 1–4 demonstrate, there is considerable variability among the individual classifiers in the degree to which they carry information that has predictive value about the object that they classify. Fourth, the relatively late mastery of the classifier system (Erbaugh, 1986) may reduce its potential to shape cognitive processes. And finally, as we discussed earlier, individual classifiers are only a piece of a larger classifier system, the rest of which involves rather different types of semantic knowledge. Individual classifiers provide only one of multiple ways of grouping entities by means of noun classification, and the categories they pick out may lack salience as a result. When considered in this broader perspective, it is perhaps more surprising that effects on non-linguistic cognition emerge at all than that they often do not.

Relation of our results to those of Schmitt and Zhang

Given this line of reasoning, why have Schmitt and Zhang (1998) and Zhang and Schmitt (1998) found what seems to be a rather pervasive effect of classifier category knowledge on a variety of non-linguistic tasks? Our data suggest that stimulus selection may contribute to the existence of these effects. The classifier categories that they used were primarily restricted to those we would consider to be Prototype, with a few Well-defined categories and no Arbitrary categories. Especially in Zhang and Schmitt (1998), in cases where the classifier can take a diverse set of objects with some having no clear connection to other members of the category, the particular items selected to represent the category seem to be often (though not always) those with a more transparent relation (usually perceptual). Schmitt and Zhang (1998) also included three classifiers we consider to be container classifiers (indicating, for instance, 'bottle of X') rather than individual classifiers. Such item sets may be more likely to form coherent groupings in memory, trigger appreciation of classifier-related similarities, etc., than if they had made other possible selections. This is not to suggest that their selection was inappropriate for its purpose. However, it does suggest that the interpretation of their effects may need to be qualified in terms of its generalisability.

In light of Schmitt and Zhang's results using stimuli that span what we considered to be both Well-Defined and Prototype categories, however, one might wonder why we did not obtain more evidence of clustering for Chinese speakers for the Prototype stimuli. Following from the preceding point, part of the answer may lie in the particular nouns we presented to participants. We randomly selected from all the nouns we had established as being classified by a particular classifier, and as a result our stimuli may have included more of those having connections to the category that are purely arbitrary from the perspective of the current speaker. Our paradigm also differed from theirs in other respects, most notably that our nouns were embedded in sentences whereas theirs were presented in lists of nouns alone. Although our participants knew they needed to pay special attention to the nouns in order to recall them later, they were processing them in the context of sentences that created thoughts or images of events such as a purchase or a dream or a theft, etc. The salience of relations among the nouns per se was likely reduced, or at least was working against competing elements that may have influenced the organisation in memory. One could argue that our paradigm therefore was not designed to best reveal to what extent classifiers categories can serve as organisational structures in memory. We agree that it was not optimally designed to reveal the maximum extent to which they might, under conditions in which there is little other information to process or guide storage and retrieval. However, we suggest it better reflects the extent to which classifiers may influence memory processes under conditions in which they are normally encountered.

The classifier present vs. absent effect

We did find that the degree of classifier-based clustering (where it occurred, namely, for Well-Defined categories) varied depending on whether the relevant classifier was present in each sentence or not. This heightened effect occurred despite the fact that participants were reading whole sentences and not focusing on the classifiers per se. From that perspective it might seem impressive that the presence of the classifier has such a notable effect on triggering a classificatory schema. However, it must also be noted that other than in the control (Mixed) block, each block of sentences contained four nouns from each of four classifier categories, and each participant received three blocks of trials with this characteristic. The presence of repeated classifiers may have therefore become rather salient to participants and drawn their attention to the shared category memberships. The fact that the clustering effect was minimal, even for the Well-Defined categories, when classifiers were absent from the sentences, again suggests that under conditions in which people are normally encountering nouns (that is, embedded in sentences carrying meaning that goes substantially beyond that embodied in the noun), the impact of classifier category membership on the organisation of information in memory may be quite modest.

CONCLUSIONS

Classifier languages are spoken by a large portion of the world's population, and the recent interest in classifiers systems and their potential for cognitive consequences is overdue. To understand well the implications for non-linguistic consequences, however, it is important to have a full perspective on the nature of the individual portion of the classifier system and its place in the larger system. We have tried to provide such a perspective. It is also important to understand the nature of the knowledge that underlies use of different classifier categories for current speakers of the language. We have found that distinctions among three category types need to be drawn. These distinctions appear to have important implications for the likelihood of finding cognitive consequences.

REFERENCES

Adams, K. L. (1986). Numeral classifiers in Austroasiatic. In C. Craig (Ed.), *Noun classes and categorization* (pp. 241–261). Philadelphia: John Benjamins.

Adams, K. L., & Conklin, N. F. (1973). Toward a theory of natural classification. *Papers from the ninth regional meeting of the Chicago Linguistic Society,* 1–10.

Ahn, W., & Kim, N. S. (2000). The causal status effect in categorization: An overview. In D. Medin (Ed.), *The psychology of learning and motivation*, *40* (pp. 23–65). New York: Academic Press.

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

Ameel, E., Malt, B. C., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language*, *58*, 262–285.

Armstrong, S. L., Gleitman, L., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*, 263–308.

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory and Cognition*, *6*, 227–232.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*, *11*, 211–277.

Berlin, B. (1968). *Tzeltal numeral classifiers: A study of ethnographic semantics.* The Hague: Mouton.

Bousfield, W. A., & Cohen, B. H. (1955). The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. *Journal of General Psychology*, *52*, 83–95.

Bousfield, W. A., Sedgewick, C. H., & Cohen, B. H. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology*, *67*, 111–118.

Burling, R. (1965). How to choose a Burmese numeral classifier. In M. Spiro (Ed.), *Context and meaning in cultural anthropology* (pp. 243–265). Glencoe, IL: Free Press.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, *64*, 309–343.

Chao, Y-r. (1968). *A grammar of spoken Chinese.* Berkeley, CA: University of California Press.

Chen, B., Che, G., Chen, H., & Zhang, Z. (1988). *Hanyu liangci cidian [A dictionary of Chinese numeral classifiers]*. Fuzhou, Fujian Province: Fujian People's Publishing House.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Denny, J. P. (1976). What are noun classifiers good for? *Papers from the twelfth regional meeting of the Chicago Linguistics Society*, 122–132.

Denny, J. P. (1986). The semantic role of noun classifiers. In C. Craig (Ed.), *Noun classes and categorization* (pp. 297–308). Philadelphia: John Benjamins.

Dixon, R. M. W. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton.

Downing, P. A. (1984). *Japanese numeral classifiers: A syntactic, semantic, and functional profile.* Ph.D. dissertation. Ann Arbor, Michigan: UMI Dissertation Information Service.

Erbaugh, M. S. (1984). 'Scissors, paper, stone:' Perceptual foundations of noun classifier systems. *Papers and Reports on Child Language Development*, *23*, 41–49.

Erbaugh, M. S. (1986). Taking stock: the development of Chinese noun classifiers historically and in young children. In C. Craig (Ed.), *Noun classes and categorization*. Amsterdam/Philadelphia: John Benjamins.

Greenberg, J. H. (1972). Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Stanford University Working Papers on Language Universals*, *9*, 1–39.

Haas, M. R. (1967). Language and taxonomy in Northern California. *American Anthropologist*, *70*, 358–362.

Han Ying Cidian A Chinese-English Dictionary. (1980). Beijing: Commercial Press.

Hiranburana, S. (1979). A classification of Thai classifiers. *Southeast Asian Linguistic Studies*, *4*, 39–54.

Imai, M., & Saalbach, H. (in press). Categories in mind and categories in language: Do classifier categories influence conceptual structures? In B. C. Malt & P. Wolff (Eds.), *Words and the world: How words capture human experience*. Oxford: Oxford University Press.

Jiao, F. (2001). *A Chinese-English Dictionary of Measure Words. Beijing*, China: Sinolingua.

Kronenfeld, D. B., Armstrong, J. D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extensionist semantic view. In J. W. D. Dougherty (Ed.), *Directions in cognitive anthropology* (pp. 91–113). Urbana, IL: University of Illinois Press.

Lakoff, G. (1986). Classifiers as a reflection of mind. In C. Craig (Ed.), *Noun classes and categorization* (pp. 13–51). Philadelphia: John Benjamins.

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

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

Lyons, J. (1977). *Semantics*. Cambridge, UK: Cambridge University Press.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory and Cognition*, *10*, 69–75.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

McRae, K., de Sa, V., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.

Mithun, M. (1986). The convergence of noun classification systems. In C. Craig (Ed.), *Noun classes and categorization* (pp. 379–397). Philadelphia: John Benjamins.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Noss, R. B. (1964). *Thai reference grammar*. Washington, DC: Foreign Service Institute.

Pulman, S. G. (1978). Shape classifiers and natural categories. *University of Essex Language Center Occasional Papers*, *20*, 35–57.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Ross, B. H. & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, *38*, 495–553.

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

Sanches, M. (1977). Language acquisition and language change: Japanese numeral classifiers. In B. G. Blount & M. L. Sanches (Eds.), *Sociocultural dimensions of language change* (pp. 51–62). New York: Academic Press.

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*, 108–122.

Sera, M., Elieff, C., Forbes, J., Burch, M. C., Rodriguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*, 377–397.

Sloman, S. A., Malt, B. C., & Fridman, A. (2001). Categorization versus similarity: The case of container names. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 73–86). Oxford: Oxford University Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Sun, H. L., Sun, D. J., Huang, J. P., Li, D. J., & Xing, H. B. (1996). The description on the corpus system of modern Chinese studies. In Z. S. Luo & S. L. Yuan (Eds.), *The studies of Chinese and Chinese character in computer era*. Beijing: Tsinghua University Publisher.

Supalla, T. (1986). The classifier system in American Sign Language. In C. Craig (Ed.), *Noun classes and categorization* (pp. 181–214). Philadelphia: John Benjamins.

T'sou, B. K. (1976). The structure of nominal classifier systems. In P. N. Jenner, L. C. Thompson, & S. Starosta (Eds.), *Austroasiatic studies* (pp. 1215–1247). Honolulu: University of Hawaii Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Vigliocco, G., Vinson, D. P., Paganelli, D., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*, 501–520.

Xiandai Hanyu Cidian [*Modern Chinese Dictionary*]. (1984). Beijing: Commercial Press.

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

Zubin, D. A., & Kopcke, K.-M. (1981). Gender: A less than arbitrary grammatical category. In R. A. Hendrick, C. S. Masek, & M. F. Miller (Eds.), *Papers from the Seventeenth Regional Meeting of the Chicago Linguistic Society*, pp. 439–449. Chicago, IL: Chicago Linguistic Society.

APPENDIX A

Chinese individual numeral classifiers and some associated nouns. Bracketed information is the meaning of the classifier word when used as a noun, verb, or adjective. Numbers in parentheses are frequencies for the word when used in sentences as a classifier, derived from a corpus of approximately 10 million words tagged for part of speech (Sun, Sun, Huang, Li, & Xing, 1996). The corpus contained texts from newspapers, literature, and oral language material (drama). Frequencies are number of occurrences divided by 1,000.

Predominantly shape-based

Saliently one-dimensional

1. *duan* (0.1597) [a section of something that extends saliently in one dimension] rope, stick, road, railway, speech, article, life, experience
2. *gen* (0.0921) [root (of a plant), indicating a stick-shape object] stick, chopstick, straw, candle, finger, hair, needle, thread, rope, nerve, pencil
3. *gu* (0.0872) [strand] thread, rope, water, flood, airstream, cold current, warm current, fragrant smell, offensive odour
4. *jie* (0.0147) [section, length] something that consists of natural sections in length, or something that is often cut into sections, such as train car, cell battery, stick, rope, pipe, chalk, period of lesson (in school)
5. *jie* (0.0122) [to cut (into halves)] an arbitrarily cut section of something that extends in one dimension, used for wood, stick, wire, bamboo pole, road
6. *liu* (0.0016298) [tuft, lock, skein] thread, knitting wool, hair
7. *lü* (0.0163) [wisp, strand, lock] thread, hemp, smoke, sunlight, moonbeam
8. *pie* (0) [one particular stroke of a Chinese character] moustache (which resembles the stroke in shape), eyebrow
9. *shu* (0.007334) [to tie, to bundle up] something in a long shape of a bundle, bunch, sheaf, used for fresh flowers, straw, sunlight, flash light
10. *si* (0.0211871) [a thread-like thing] hair, vision, breeze, smile, warmth
11. *tiao* (0.8947) [a slender, long-shape thing, often flexible] rope, line, plait, snake, fish, stream/brook, river, canal, towel, road, trousers, skirt, blanket, slogan, news, experience, life, brave/true man
12. *zhi* (0.009) [tree branch, twig] tree branch, match, pencil, pen, cigarette, arrow, gun
13. *zhi* (0.1556) [a stick-like long thing] candle, pencil, pen, cigarette, flower, thermometer, gun, pistol, spear, arrow, hand, arm, feather, troop

Saliently two-dimensional

14. *mei* (0.0554) coin, badge, medal, stamp, missile
15. *mian* (0.0464) [surface] mirror, silk banner, flag, wall, big drum

16. *pan* (0.0212) [a plate] magnetic audio tape, video tape, mosquito-repellent incense (coiled in a shape of a plate), grinding stone, chess match
17. *pian* (0.2119) [a flat, thin piece, slice, or a stretch of land] bread, meat, tree leaf, snow flake, farming field, desert, forest, white/dark cloud
18. s*han* (0.0171126) [a leaf-shape thing] used door, window, sail, partition
19. z*hang* (0.2772) [to spread open/flat] paper-like things, or something that has a flat surface, including paper, photo, ticket, diploma, certificate, stamp, postcard, phonograph record, carpet, cattle hide, pancake, desk, table, bed, mouth, bow, fishing net

Saliently three-dimensional
20. *ban* (0.0016298) [a segment/section (of an orange, etc.)] orange, mandarin, tangerine, garlic
21. *di* (0.022) [to drip (in drops)] water, oil, tear, blood, sweat, saliva, soup, vinegar
22. *ke* (0.1019) [something small and roundish in shape] pearl, soy bean, button, tooth, mine, bullet, bomb, star, (man-made) satellite
23. *kuai* (0.4645) [a lump-shape thing] soap, candy, cake, meat, stone, wrist watch, cloth, handkerchief, lawn, farming field,white/dark cloud
24. *li* (0.0139) [a grain-like thing] rice, salt, sand, grain, seed, sweat, button, bullet
25. *quan* (0.0318) [a circle] water, grease stain, hills, mountains, wreath
26. *tuan* (0.022) [a collection of something in a ball shape] cotton, thread, knitting wool, paper, wire, hemp, dough, fire, smoke, dark cloud
27. *tuo* (0.0016298) [a big lump] iron bar, lead bar, mud
28. *wan* (0.0513) [a ball, pellet] Chinese medicine, marble
29. *xing* (0) [a star] light (in a distance), oil
30. *ya* (0.0008149) [tooth, or something with a shape of a tooth, indicating a shape of a crescent moon] moon, watermelon, pancake
31. *zhou* (0.0008149) [a spool (for thread)] thread, (a scroll of) Chinese painting
32. *zuo* (0.0049) [a tuft] hair, beard

Salient feature
33. *ba* (0.1516) [a handle] things that have a handle, such as umbrella, pistol, teapot, knife, screwdriver, scissors, pliers, hammer, spoon, broom, violin, chair, key, ruler
34. *ding* (0.0089638) [crown of the head, top] something that has a top, such as cap, hat, straw hat, tent, mosquito netting, umbrella
35. *dong* (0) [a hole] (stone) bridge, big (arch) gate
36. *gan* (0.0016298) [shaft or arm] things that have shaft or arm, such as rifle, steelyard, flag, pen, pencil

37. *guan* (0) [a pipe] something that has a pipe-like shape, such as hunting gun, bamboo flute, hair brush (for writing or painting)
38. *jia* (0.0179) [a frame, stand] things that have a frame, such as airplane, space shuttle, helicopter, ladder, eye glasses, machine, piano, accordion, electronic keyboard, camera
39. *kou* (0.0459) [mouth] something has a shape of a mouth, such as pot, bell, water well, person, pig, coffin, knife
40. *yan* (0.05857) [an eye] things that have a big opening, such as water well, water spring, roof window, cave house

Multiple shared features, animate vs. inanimate

Animate/Human

41. *dai* (0.1646072) [generation] emperor, people
42. *hu* (0.1157) [household] family, residents
43. *ming* (0.449) [name] people of different professions, such as teacher, professor, nurse, doctor, scientist, lawyer, journalist, worker, student, writer, soldier, actor/actress, politician, policeman, sailor
44. *ren* (0.0089638) [to hold the post of] president (of country or institution), mayor, chairman, company/factory head
45. *tai* (0.0016298) [fetus] boy, girl, twins, also used for animals, such as piglets, puppies, etc.
46. *wei* (1.0919) [an individual, a person] professor, teacher, mister, miss, parent, policeman, comrade [more polite than the general classifier *ge*]

Animate/Animal

47. *pi* (0.0212) horse, mule, cloth (a bolt of)
48. *tou* (0.0619) [a head] big animals, such as pig, deer, cattle, donkey, lion, elephant, garlic (a head of)
49. *wo* (0.0008149) [nest, litter, brood] birds, chickens, eggs, pigs, children
50. *zhi* (0.3308) [single, alone, one of a pair] bird, fly, mosquito, bee, chicken, goat, sheep, tiger, elephant; also used for hand, foot, leg, eye, ear, shoe, sock, boat, watch, suitcase, music/tune

Inanimate/Natural object

51. *duo* (0.0244466) flowers, white cloud
52. *ke* (0.0685) all plants with stems and leaves (the whole plant), such as tree, grass, corn, cabbage
53. *lun* (0.0269) [a wheel] the sun and the moon only (especially, red sun, and bright moon)
54. *pao* (0.0081489) urine, shit
55. *tan* (0.0032595) [to spread (on the ground) a small pool of liquid, mud] water, blood, mud, shit

56. *zhu* (0.014668) [stalk and the part of the root that is above the ground] plants only, small tree, big tree, seedling
57. *sheng* (0.1548) [sound] gun shot, thunder, shout, crying, coughing, knocking

Inanimate/Artifact (concrete)
58. *ban* (0.0147) [a work shift] transportation on fixed schedule, such as bus, train, ship, airliner
59. *ben* (0.119) [a book (a bound copy of printed materials)] book, magazine, pictorial, novel, dictionary
60. *bian* (0) [a braid] garlic (a braid of), hair
61. *bu* (0.1597) [part] film, literary work (especially one of good quality, and in a form of a book), long novel, telephone
62. *ce* (0.0473) [copy, volume] book
63. *chu* (0.0537825) [place, location] physical wound, typographical error, household
64. *chuang* (0.0032595) [bed] quilt, cotton-padded mattress, bedding
65. *dao* (0.075) [way, course, path] wall, fence, door, gate, defence line, dish, procedure, sun rays
66. *dong* (0.0139) building
67. *du* (0.0049) [to block up] wall, fence
69. *dun* (0.0782) [pause] meal
70. *fa* (0.0032595) [to fire] bullet, artillery shell
71. *fen* (0.1719) [share, portion/part of a whole] newspaper, magazine, exam paper, homework, meal, gift, job
72. *feng* (0.0929) [to seal] letter, telegram
73. *fu* (0.0505) [the width of cloth (a bolt of)] picture, painting, ad, poster, map
74. *gua* (0.0008149) [to hang something on a hook] a set of something tied/strung together, such as firecrackers (a string of), bead curtain, a horse and cart
75. *ji* (0.0032595) [a dose] Chinese herbal medicine, decoction of medicinal ingredients
76. *jia* (0.4433) [family, home] household, store, restaurant, hotel, supermarkets, bank, cinema, hospital, factory, company, news agency, travel agency, publishing house
77. *jia* (0) [harness, horse-drawn vechicle] horse-drawn cart, cattle-drawn cart, horse-drawn sleigh
78. *jian* (0.1002) [room] any rooms, including bedroom, living-room, kitchen, bathroom, study, office, classroom, workshop
79. *jian* (0.3724) [a piece] clothes, shirt, coat, overcoat, jacket, sweater, luggage, matter/thing, work/job, case

80. *jü* (0.3675) [sentence] speech, talk, poem
81. *jü* (0.0024447) [utensil, apparatus] corpse, coffin
82. *juan* (0.0823) [book, volume] book, writings/works (in a form a book)
83. *liang* (0.119) all ground vehicles including bus, car, truck, bicycle, jeep, tractor, train, tank
84. *long* (0.0008149) [ridge (in a farming field)] farming land, roof tile
85. *pian* (0.1019) [a complete article] article, report, editorial, commentary, review, novel, prose
86. *qi* (0.0407444) [scheduled time/date] magazine (one issue of), pictorial, training class, students/trainees (in one training class), project
87. *qi* (0) [a rectangular piece of land in a field, separated by ridges (usually for growing vegetables)] vegetable, plant
88. *shen* (0.0554) [body] suit, clothes, dress, strength, skills in martial arts, foreign flavor/Western style
89. *sou* (0.0253) all ships (especially big in size) including speedboat, ocean liner, warship, oil tanker
90. *suo* (0.0016298) [cartridge clip] bullet
91. *suo* (0.0774) [location] house, villa, residence, school, kindergarten, university, hospital, club, church
92. *tai* (0.1247) [platform, stage, stand, support] for machine, TV set, recorder, radio, computer, locomotive, tractor, performances
92. *tang* (0.0701) [(frequency of) scheduled transportation] regular bus, train, ship, ocean liner, airliner
93. *tie* (0) [to paste, to stick] medicated plaster
94. *ting* (0.0008149) rifle, machine gun, submachine gun
95. *wei* (0.0024447) [taste, flavor] ingredient (of a Chinese medicine prescription)
96. *ye* (0.0416) [page, leaf] paper, book, text, article, novel, document
97. *zhan* (0.0081) [a small cup] oil lamp, bulb lamp, fluorescent lamp
98. *zhang* (0.0407444) [chapter] book, novel, thesis, dissertation
99. *zhuang* (0.0423741) building
100. *zhuo* (0.0122233) [table] used for food, feast, people, guests
101. *zun* (0.0073) [respect] statue of a Buddha, artillery piece
102. *zuo* (0.2135) [seat, stand, pedestal, base] bell, stone tablet, pagoda, bridge, house, temple, building, factory, church, grave, reservoir, forest, mountain, village, city

Inanimate/Artifact (other)

103. *bi* (0.0742) [pen/pencil] (business) deal, sum of money, cash, fund, expense

104. *chang* (0.2249) [arena, field] battle, fight, war, illness, storm, rain, disaster, nightmare, film, concert, dancing ball, opera, play, ball (basketball, football,volleyball, tennis ball, etc.) match
105. *chu* (0.0179) [a big section/episode of a legend] a dramatic piece, including opera, play
106. *dian* (0.022) [spot, dot, indicating a point (as in a point of view), and a tiny amount] view, suggestion, criticism, request, ink spot/stain, blood spot/stain
107. *ji* (0.009) [a collection of literary works, volume, part, used for film, TV play
108. *jie* (0.2868) [due time] something that occurs in a fixed sequence, such as congress, president, students (enrolled in the same year), Olympics, the Asian Games
109. *ma* (0.013) [number symbols] matter
110. *men* (0.035) [branch, class, category] branch of learning, knowledge, art, subject, course, craftsmanship, artillery piece
111. *mu* (0.0236) [curtain] (an act of) play, reminiscence of an earlier event
112. *qi* (0.0334) (an occurrence of an) accident, theft, robbery, burglary, murder
113. *qiang* (0.007334) [(thoracic) cavity] love, regret, warmth, enthusiasm, anger, hatred
114. *qu* (0.0097786) [tune melody] song, music, melody, solo, duet, trio, quartet, etc.
115. *shou* (0.0464) song, poem, nursery rhyme
116. *tang* (0.0024447) [hall] lesson (as in school), furniture
117. *xi* (0.013) [feast] banquet, talk, conversation (with someone)
118. *xian* (0.0114) [thread] hope, light, life/energy
119. *xiang* (0.5623) [item] plan, suggestion, decision, order, decree, measure, task, work, activity, invention, discovery, result (of an experiment), cause, (business) deal record
120. *ze* (0.0081489) [norm, rule] a piece of writing, such as news, ad, commentary, fable
121. *zhan* (0.0008149) [to stop] way, distance
122. *zhao* (0.0032595) [a move (in chess)] move (in chess), good idea
123. *zhen* (0.0782) [(a short) duration of time] wind, rain, cold spell, laughter, applause, footsteps, knockings (on the door), gun shots
124. *zhuang* (0.0162977) [stake, pile] (big/small) matter, case, (business) deal, worry/concern
125. *zong* (0.0048893) [ancestor, faction/sect] business deal, (a large sum of) money

General classifier

126. *ge* (8.5547) Generally used for nouns that do not have a special classifier, but also often used as a substitution for some specific classifiers (especially in casual speech); the nouns may include person, boy, girl, man, woman, student, teacher, sun, moon, week, month, fruit, apple, pear, orange, watermelon, country, nation, state, province, city, county, district, school, place, forest, desert, grassland, park, game, festival, story, idea, question, problem, experiment, investigation, solution, method, opportunity, ceremony, dish, plate, sofa, table, chair, news, film, play, and dream

APPENDIX B

The four most frequently produced features for each classifier from Experiment 2, in order of frequency. These features were also used for classifier identification in Experiment 3 and embodiment ratings in Experiment 4

Well-defined	
Ke	is a plant, is slender/long, has a (long) trunk, has roots
Liang	has wheels, is a vehicle, travels/moves, is a transportation tool
Zhan	is a lamp, produces light/is a lighting device, contains/burns oil or liquid, is round/saucer-shaped
Di	is a liquid, is very small in quantity, is roundish/ball-shaped, is falling (in the air)
Ben	is made of paper (bound together), is a book, has many pages (can be turned), is rectangular in shape/structure
Duo	is a flower/flower-like object, is roundish/ball-shaped, has a irregular shape/not identical, is beautiful/colourful
Ming	is a human/human profession, calls for respect, can be used as a title, has blood
Ting	is a gun/weapon, has a long/stick-like shape, has legs/supporting structure, is large in size
Prototype	
Gen	is slender/long, is hard/inflexible, is straight, is soft/flexible
Jia	is a family/whole made up of many people, is place involving human activity/business, is an organisation involving a building, are people of blood relations (living together)
Ke	is a roundish, ball-like object, is small in size, is solid/3-D, has a smooth/shining surface
Tou	is a certain animal, is medium/large in size, is a livestock, has head/head characteristics (horn, nose, etc.)
Ba	can be held/operated by hand, is often a tool (to repair things with), is a physical object, has a handle
Tiao	is slender/medium/very long, can be soft/flexible, can be hard/inflexible, is rectangular in shape
Jia	has legs/supporting structure/can be popped up), is a solid/physical object, can be an aircraft, is large in size
Tai	has or is related to a flat surface, is a piece of mechanical machinery, has a regular square shape (like a table), can stand steadily/solidly

APPENDIX B (*Continued*)

Arbitrary	
Bi	is related to money/finance, is in a big quantity, is business or business deal/activity, involves math/numbers
Bu	is literary/artistic work, is a vehicle/ground transportation tool, is a movie, has a beginning and end/complete
Zun	is an inanimate object, is heavy, sits/stands upright, is an image/imitation of an animate object,
Dao	is long and thin, is a certain procedure, has/marks (clear-cut) borderlines, is a light/flash of lightning or the like
Dun	is an action taking a period of time to complete, is a meal, has more than one component, its action needs to be repeated
Jü	is a dead person/corpse, is a dead animal body, is rectangular in shape, is medium in size
Zhuang	is an event that has taken place, is an event that is serious/not trivial, is a case in law, is something abstract/not concrete
Zong	is business deals/transactions, is a case (as in law), is a matter that requires investigation, is an intangible event