Classifier categories reflect, but do not affect conceptual organization

Laura J. Speed[1], Jidong Chen[2], Falk Huettig[1,3], & Asifa Majid[4]

[1]Centre for Language Studies, Radboud University, Nijmegen, NL

[2]California State University, Fresno, USA

[3]Max Planck Institute for Psycholinguistics, Nijmegen, NL

[4]Department of Psychology, University of York, UK.

Journal of Experimental Psychology: Learning, Memory, and Cognition

Corresponding authors:

Laura J. Speed
Centre for Language Studies
Radboud University
l.speed@let.ru.nl

Asifa Majid
Department of Psychology
University of York
asifa.majid@york.a.uk

Abstract

Do we structure object-related conceptual information according to real-world sensorimotor experience, or can it also be shaped by linguistic information? This study investigates whether a feature of language coded in grammar—numeral classifiers—affects the conceptual representation of objects. We compared speakers of Mandarin (a classifier language) with speakers of Dutch (a language without classifiers) on how they judged object similarity in four studies. In the first three studies, participants had to rate how similar a target object was to four comparison objects, one of which shared a classifier with the target. Objects were presented as either words or pictures. Overall, the target object was always rated as most similar to the object with the shared classifier, but this was the case regardless of the language of the participant. In a final study employing a successive pile-sorting task, we also found that the underlying object concepts were similar for speakers of Mandarin and Dutch. Speakers of a non-classifier language are therefore sensitive to the same conceptual similarities that underlie classifier systems in a classifier language. Classifier systems may therefore reflect conceptual structure, rather than shape it.

Keywords: classifiers, linguistic relativity; language and thought; object concepts; Mandarin; Dutch

1. Introduction

What makes two objects conceptually similar? Object similarity can be based on visual similarities; for example, both a knife and a sword are *sharp*, *metallic* and *shiny*. Objects can also be similar in terms of their function or affordances: both a paper bag and basket can be picked up by a handle and used to carry objects. Objects that are similar are said to be represented close to each other in conceptual space (Malt, Sloman, Gennari, Shi, & Wang, 1999; Rosch & Mervis, 1975). In the present research, we explore another way in which objects may be conceptually similar—through a shared grammatical category. Can conceptual representations be structured according to higher-level linguistic information?

For many years, whether or not language affects thought—the question of linguistic relativity—has been fiercely debated. Evidence for at least some effect of language on thought has been demonstrated in various semantic domains (for review see Wolff & Holmes, 2011). Beyond the lexical level, there is evidence that grammatical features of a language can affect thought too. One such feature is grammatical gender. Grammatical gender divides nouns into classes based on the behavior of associated words, such as articles and adjectives (Corbett, 2006). In some languages, grammatical gender is associated with biological gender or sex, and all nouns are categorized as grammatically feminine or masculine regardless of whether or not they have a biological gender/sex. It has been suggested that gender assignment for some classes of objects (e.g., animals) may be semantically driven (Boroditsky & Schmidt, 2000; Sera, Elieff, Forbes, Burch, Rodríguez, & Dubois, 2002), but when tested systematically across multiple languages the similarity between grammatical gender assignment for superordinate classes of objects (e.g., artifacts, natural objects) is weak, if present at all (Foundalis, 2002). The grammatical gender assigned to specific objects likewise varies across languages: for example, *apple* is masculine in German (*der Apfel*), but feminine in French (*la pomme*).

Despite arbitrary gender assignment, evidence suggests that grammatical gender can affect the way people think about objects. For example, when asked to describe an object, Spanish and German speakers are more likely to ascribe stereotypically male qualities to objects with masculine grammatical gender, and stereotypically female qualities to objects with feminine grammatical gender (Boroditsky, Schmidt, & Phillips, 2003; but see Mickan, Schiefke, & Stefanowitsch, 2014). Grammatical gender can even affect the way people remember complex fragrances: fragrances are remembered better when the grammatical gender of ingredients in the fragrance match the gender of the person the fragrance is marketed towards (Speed & Majid, 2019).

Although there is evidence for grammatical gender affecting thought under some conditions, other paradigms show different effects (e.g., Vigliocco, Vinson, Indefrey, Levelt, & Hellwig, 2004), and the overall conclusion is still debated (see Bender, Beller, & Klauer, 2018). Effects of grammatical gender are particularly evident in tasks that use linguistic stimuli rather than pictures (Ramos & Roberson, 2011; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005), and when gender is emphasized by using either gender-marked articles (Imai, Schalk, Saalbach, & Okada, 2014) or having explicit reference to gender in the instructions (Bender, Beller, & Klauer, 2016; Cubelli, Paolieri, Lotto, & Job, 2011; Ramos & Roberson, 2011). See Samuel, Cole, and Eacott (2019) for a systematic review of these effects.

Gender is only one type of grammatical system that may have implications for object representations. Taking a broader perspective across the world's languages, we see that many languages employ different nominal classification systems (Seifart, 2010). In the current study we focus on one of these—i.e., numeral classifiers. Numeral classifiers are morphemes used to modify a noun when describing an object in terms of a specific quantity. For example, it would be ungrammatical to say "two snakes" in Mandarin Chinese; instead one must include

the classifier "*tiao*" (i.e., "two *tiao* snakes"). This is similar to how, in English it is ungrammatical to say "two sands", where "two *heaps* of sand" would be appropriate. Classifiers therefore act as a unitizer so that nouns can be counted, but they also provide semantic information about the referent (Imai & Saalbach, 2010; Lucy, 1996). Since they provide a unit to count or measure material, they may be more likely to reflect salient perceptual features of entities than other grammatical markers, such as grammatical gender. In this study we assess whether classifier categories affect conceptual organization, or instead merely reflect it.

Classifiers have a different organizational principle than nouns (Saalbach & Imai, 2012), thereby crosscutting taxonomic categories. They appear not to exhibit the same hierarchical structure of the noun lexicon (Imai & Saalbach, 2010). It is thought that all classifier systems distinguish an animate/human class (Adams & Conklin, 1973); alternatively some suggest classifier systems are primarily determined by shape (Olness, 1991). Other organizing classifications can include social status—such as kinship relations—nature/function, and orientation (Croft, 1994), whereas other systems may be more arbitrary (Gao & Malt, 2009).

Although classifier categories may reflect conceptual distinctions, there is often large variability in category size and coherence, with some classifier categories including a broad range of objects cutting across taxonomies (Gao & Malt, 2009; Saalbach & Imai, 2012) thereby appearing fairly heterogeneous (Srinivasan, 2010). For example, the classifier *tiao* in Mandarin Chinese can be used to describe long thin objects such as snakes or rivers, but can also be used with objects not strongly sharing such features, such as boats. Furthermore, the semantic distinctions can often seem idiosyncratic and arbitrary (Srinivasan, 2010). This raises the question of whether or not classifiers are meaningful to speakers of classifier

languages, and whether such a grammatical system can affect the way speakers think about objects.

Previous work suggests classifier systems influence the ontology of objects. Lucy and Gaskins (2003) compared object categorization in speakers of a classifier language, Yucatec Mayan, and speakers of a non-classifier language, English. In English, when talking about objects of a specific quantity, numerals are used for discrete objects (e.g., *two* candles), but a unitizer is required for non-discrete objects, e.g., two *clumps* of dirt. In comparison, reference to all objects specified with a numeral require a classifier in Yucatec, regardless of whether the object is discrete or non-discrete. Numeral classifiers in Yucatec typically provide crucial information about the shape of the object; without classifiers nouns are semantically unspecified with regards to shape and quantity, as if they were "unformed substances" (p. 471). In contrast, quantificational unit is part of the basic meaning of many English nouns, and shape is an integral feature of this meaning (for discrete objects, like candles).

To investigate whether the differences between the two languages affect object categorization, Lucy and Gaskins (2003) conducted a similarity matching task using real objects. Since the quantificational unit of English nouns is often shape, English speakers should be more likely to categorize objects in terms of shape than Yucatec speakers. In comparison, Yucatec nouns lack a quantificational unit and are more likely to highlight the material composition of objects. Yucatec speakers should therefore be more likely to categorize objects in terms of material. Speakers were presented with triads consisting of a target object and two alternate objects, one with the same shape as the target and one of the same material as the target. For example, speakers were shown a plastic comb with a handle as a target and asked whether it was more like a wooden comb with a handle or more like a plastic comb without a handle. Speakers of the two languages judged the similarity of objects differently: English speakers matched the same objects according to shape, whereas Yucatec

Mayans showed a preference for matching objects according to material substance. So the two groups conceptualize and classify objects differently, in line with the difference in the basic meaning of the languages' nouns: English speakers are more attuned to the shape of objects, whereas Yucatec speakers are more attuned to object material.

Other studies have examined the overall similarity of objects and have shown that speakers of a classifier language (Mandarin) judge pairs of objects that share a classifier as more similar than control objects that do not share a classifier (Saalbach & Imai, 2007, 2011; Schmitt & Zhang, 1998; Zhang & Schmitt, 1998). Moreover, when recalling a list of words, Chinese speakers were found to cluster their responses by classifier category unlike speakers of English (Zhang & Schmitt, 1998). This suggests that Chinese speakers' mental representation of objects is shaped by grammar (Zhang & Schmitt, 1998). However, on closer inspection classifiers may have only limited effects on judgments of similarity. It seems that differences observed between classifier and non-classifier languages in earlier studies may be more circumscribed than previously thought, and in many cases can be attributed to explicit use of linguistic stimuli, as outlined below.

The effect of a shared classifier on similarity ratings appears to be weak when contrasted with the effect of taxonomic or thematic relations (Saalbach & Imai, 2007, 2012). Furthermore, speakers of German—a non-classifier language—also judged objects that share a classifier in Mandarin to be more similar than objects without a shared classifier (although not to the same magnitude as speakers of Mandarin; Saalbach & Imai, 2007, 2011). This suggests that classifier categories may simply reflect the structure of conceptual information present in the world, rather than affecting the way object concepts are semantically represented (Saalbach & Imai, 2007).

Indeed, classifier categories are not entirely semantically arbitrary (compared to grammatical gender, say) and many objects within a classifier category share a common,

defining feature (cf. Allan, 1977; Lakoff, 1987). In support of this proposal, it has been noted that features of classifier categories mirror features salient during word learning in children (Clark, 1976). For example, classifier categories often rely on features such as shape, length, and animacy—the same features which appear in over-generalization errors in child speech. Conversely, no classifier system categorizes objects based on color, and nor do children make category errors on this basis. Classifier systems may therefore reflect the effect of thought on language, rather than language on thought. Experience using a classifier system to categorize similar pairs of objects may magnify their conceptual similarity, which leads to the observed classifier effects.

Other work suggests the effect of classifiers on similarity judgments may be even more limited in scope, depending on factors that affect the saliency of grammatical information. Language-specific effects of classifier category are not observed in tasks that engage fast, automatic cognitive processes such as word-picture priming (Saalbach & Imai, 2007, 2011; but see Srinivasan, 2010), and the presence of a classifier effect is affected by whether the classifier is explicitly presented or not (Gao & Malt, 2009; Huettig, Chen, Bowerman, & Majid, 2010). Similarly Tsang and Chambers (2011) found classifier effects only in restricted situations. Using an eye-tracking paradigm, they were able to separate grammatical and semantic effects of Cantonese shape classifiers. They found that classifiers activated shape semantics, indicated by looks to distractor objects congruent in shape with the classifier category, but only when the target object did not possess the prototypical shape. Grammatical information tended to be more important overall, demonstrated by increased looks to distractor objects with a shared classifier, regardless of shape. Finally, in a memory task Gao and Malt (2009) found no difference in object recall between Mandarin speakers and English speakers, suggesting classifier categories do not aid in memory. On the other hand, compared to English speakers, Mandarin speakers showed greater clustering in terms of

classifier categories in recall, and particularly for classifiers from “well-defined” categories, where objects sharing a classifier share at least one feature (such as shape or color). This suggests there may be some effect of classifier category on the organization of concepts.

Overall, the results from previous studies are mixed. Some studies have reported effects of classifier categories on the similarity of objects, but there may be factors constraining these effects. One critical factor could be the use of linguistic stimuli. Many studies that find classifier effects on similarity judgments have used words as stimuli, not pictures (Perniss, Vinson, Seifart, & Viliocco, 2012; Saalbach & Imai, 2007, 2012; Schmitt & Zhang, 1998; Zhang & Schmitt, 1998). It is known that when a noun is comprehended, corresponding grammatical information is activated (Meyer, Belke, Telling, & Humphreys, 2007; Pulvermüller & Shtyrov, 2003). It is therefore possible that previous effects are due to grammatical activation, rather than reflecting conceptual similarities between objects sharing a classifier (a similar proposal has been made for effects of grammatical gender; Ramos & Roberson, 2011; Samuel et al., 2019; Vigliocco et al., 2005). Specifically, if two objects are labeled by participants and this activates the same grammatical marker, they may be judged as more similar. This finding would also be in line with the thinking-for-speaking view (Slobin, 1987): language only affects thought when we are explicitly using language. A way to disentangle such possibilities is to conduct comparable experiments using linguistic and non-linguistic stimuli. We set out to conduct such an investigation comparing speakers of a classifier language with speakers of a non-classifier language.

In order to test whether object-related conceptual information is shaped by linguistic information or structured by shared real-world sensorimotor experience, we conducted a set of studies comparing speakers of language with numeral classifiers—i.e., Mandarin—with speakers of a language without numeral classifiers—i.e., Dutch. We used a similarity rating task in which participants were given a target object and had to rate the similarity of the object

to a set of four comparison objects. One of the comparison objects shared a classifier with the target; the other three objects did not. A rating task allows for a more fine-grained analysis of similarity than a forced-choice paradigm (cf., Lucy & Gaskins, 2003; Saalbach & Imai, 2007, Experiment 1) because it allows participants to indicate degree of similarity rather than forcing them to make a binary choice.

If classifier systems in a language affect the way speakers of that language conceptualize objects, then we would expect Mandarin speakers to judge a target object as more similar to an object with a shared classifier than to other objects without a classifier, but Dutch speakers would not show such a pattern. However, if classifier systems reflect real-world conceptual organization, then we would expect speakers of both Mandarin and Dutch to judge the target object as more similar to objects with a shared classifier than to objects without. Another possibility is that both Mandarin and Dutch speakers judge the target object to be more similar to an object with a shared classifier than other objects, but that this effect is larger in speakers of Mandarin. If classifier systems affect conceptual representations of objects, then we should see differences between the speakers of the two languages in tasks that use non-linguistic stimuli, such as pictures, as well as linguistic stimuli. If classifier effects reflect only thinking-for-speaking (Slobin, 1987), then effects should only be observed with linguistic stimuli.

We present four studies that assess the effect of classifiers on object similarity. In the first three studies, participants made ratings of similarity where all stimuli were pictures (Study 1), both pictures and words, (Study 2), and words only (Study 3). In a final study, we used a different procedure, where participants were asked to conduct a pile-sorting task of pictures to assess whether the conceptual structure of object concepts is affected by a classifier system (Study 4).

2. Study 1: Picture-picture similarity

2.1. Method

2.1.1. Participants

Twenty-five native speakers of Mandarin (15 female, mean age 23, range 18-27) and 24 native speakers of Dutch took part in the study. This sample size was chosen so as to be comparable to that of previous studies, whilst also increasing the number of items to gain more statistical power.[1] All participants were familiar with at least one other language. Mandarin speakers were all educated in the language Putonghua (i.e., standard Mandarin, the official language of China), and used this predominantly in daily life, but also came from different dialect backgrounds (e.g., Shandong, Henan, Shanxi, Jilin, Hubei, Jiangsu). Dutch speakers were multilingual with the languages English, German, Spanish or French, but had no familiarity with any classifier languages. The study received ethical approval from Radboud University.

2.1.2. Material

Line drawings depicting everyday objects familiar to both Mandarin speakers and Dutch speakers were used. Each trial contained one target object and four comparison objects, with one comparison object sharing a dominant classifier with the target. The three remaining comparison objects did not share a classifier with the target and were therefore considered distractors.

Classifiers were determined based on an initial norming study with a separate group of 12 Mandarin speakers. The speakers were asked to name 240 line drawings of concrete and imageable objects. In line with previous observations (e.g., Gao & Malt, 2009), over 100 different classifiers were produced, contrary to the claim that there are 50 classifiers in

[1] A post-hoc power analysis with G*Power for a medium-to-large effect size calculated a power of 0.79 for analysis by participants and 0.99 for analysis by items.

Mandarin (Chao, 1968). Classifier choice for the objects was not strict, with some nouns taking different classifiers. Nouns that had a clear dominant classifier (>80% agreement) were chosen for this study[2]. Distractors were chosen so that they did not share a classifier with the target. All objects (including the classifier match) were not semantically-related, or visually similar to the target, and did not have any phonological or morphological overlap with the target. We also checked that the target and the object with the shared classifier were not more likely to share a determiner in Dutch (*de* or *het*) than the target and the objects that do not share a classifier[3], and we checked that the target and the object with the shared classifier were not more likely to match in grammatical gender in German, Spanish, and French, than the target and distractor objects. Target objects, classifier match, and distractor objects are listed in Appendix A.

The target and comparison objects were presented on a piece of paper, with the comparison objects laid out in a two-by-two grid and the target object centered underneath the grid (see Figure 1). A response sheet also contained a two-by-two grid with empty boxes, in which the participant was to enter their similarity rating. The target object again appeared beneath this grid. There were 48 trials in total, with the position of each object counterbalanced across participants. The classifier match appeared equally often in each position.

---

[2] Classifiers were not selected according to the categories proposed by Gao and Malt, (2009), but a later re-examination showed there are 18 prototype classifiers, three arbitrary classifiers, and two well-defined classifiers.

[3] Based on the distribution of *de* and *het* for the objects, there was a 0.23 chance of choosing the object with the shared classifier based on a matching determiner with the target (with chance being 0.25).

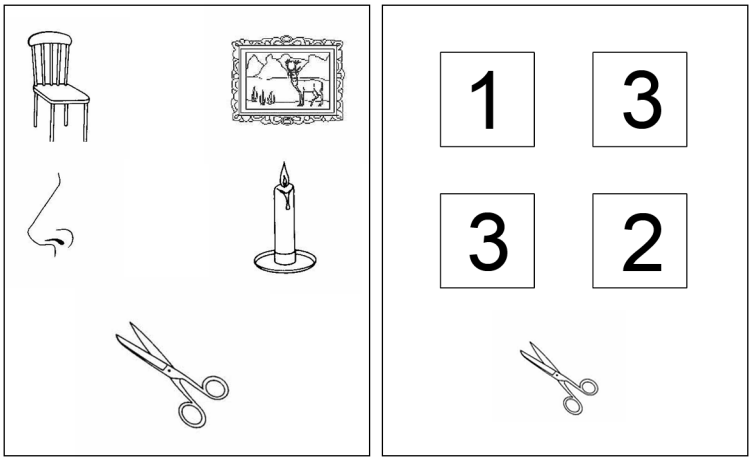

Figure 1. Example page from the stimulus booklet (left), with the target object (scissors), classifier match (chair) and comparison objects. The right picture depicts a page from the response booklet.

2.1.3. Procedure

Participants received instructions in their native language. They were asked to judge how similar the target object was to each comparison object, by placing a score between 1 (no similarity) and 10 (identical in similarity) in the corresponding boxes on their response sheet. There were four random orders of items.

2.2. Results

Ratings were z-transformed so that any differences in how rating scales were used were removed (cf. Saalbach & Imai, 2007), and averaged across distractors within a trial (raw means by item can be found in Appendix B). The transformed scores were then analyzed using a 2 by 2 mixed ANOVA with language (Mandarin vs. Dutch) as a between-subjects factor, and object (shared vs. not shared classifier) as a within-subjects factor. Analyses by subjects and by items were conducted.

There was a main effect of object type $F_1(1, 47) = 104.19$, $p < .001$, $\eta^2_p = .69$; F_2 (1, 47) = 18.49, $p < .001$, $\eta^2_p = .281$, with the object sharing a classifier with the target receiving higher similarity ratings than distractors. There was no significant main effect of language F_1

< 1; F_2 <1 and crucially, no significant language by object interaction F_1 < 1; F_2 < 1(see Figure 2).

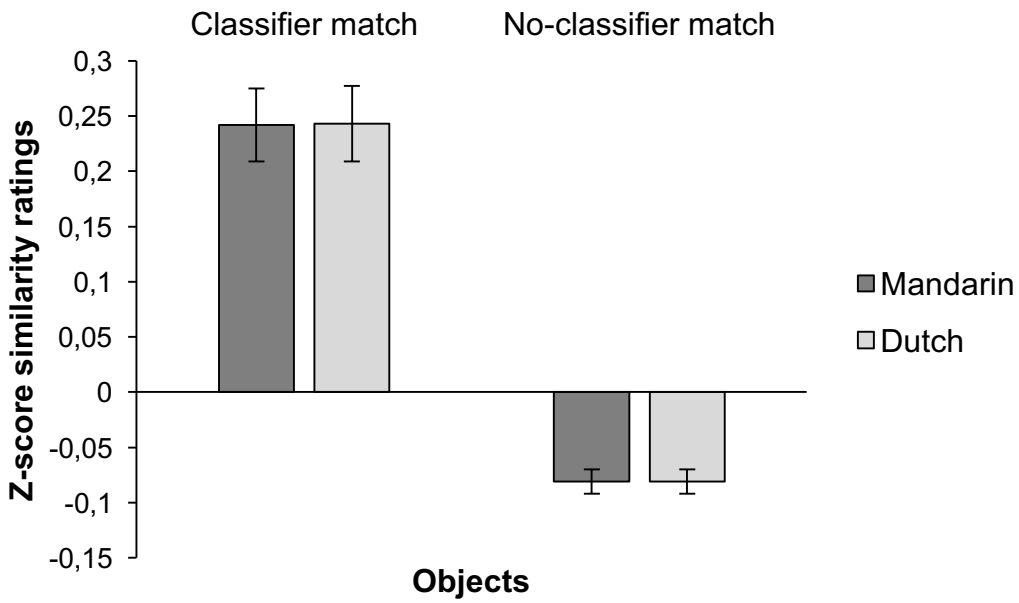

Figure 2. Picture-picture similarity judgments (z-score) for Mandarin and Dutch speakers (bars = 1 *SE*).

3. Study 2: Word-picture similarity

In Study 1 we found an effect of classifier pair similarity, such that objects sharing a classifier were judged as more similar than objects not sharing a classifier. However, there was no difference between speakers of Mandarin and Dutch. This suggests that classifiers do not affect object representations, but it does not rule out language-specific effects at the lexical level (i.e., when using language). So, in Study 2 we presented the target object as a word to encourage activation of the classifier category.

3.1. Method

3.1.1. Participants

Twenty-five native speakers of Mandarin (10 female, average age 24, range 19-30) and 24

native speakers of Dutch that did not take part in Study 1 were recruited. All further details were the same as Study 1.

3.1.2. Material

The stimuli were the same as Study 1 except that now the target was presented as a word (without the classifier) in the participants' native language. The comparison objects were presented as the same line drawings.

3.1.3. Procedure

As in Study 1, participants were asked to judge the similarity of the target object to the four comparison objects on a scale of 1 to 10 by writing their score in the corresponding box on a response sheet.

3.2. Results

The data was analyzed as in Study 1; raw means by item can be found in Appendix C. There was a main effect of object type $F_1(1, 47) = 82.64$, $p < .001$, $\eta^2_p = .64$; F_2 (1, 47 = 22.53, $p <$ .001, $\eta^2_p = .32$, with the object sharing a classifier with the target receiving higher similarity ratings than the objects not sharing a classifier. Once again, there was no main effect of language F_1 (1, 47) = 2.70, $p = .11$, $\eta^2_p = .054$; $F_2 < 1$, and no significant language by object interaction $F_1(1, 47) = 2.70$, $p = .11$ $\eta^2_p = .054$; $F_2(1, 47) = 3.04$, $p = .09$, $\eta^2_p = .061$ (see Figure 3).

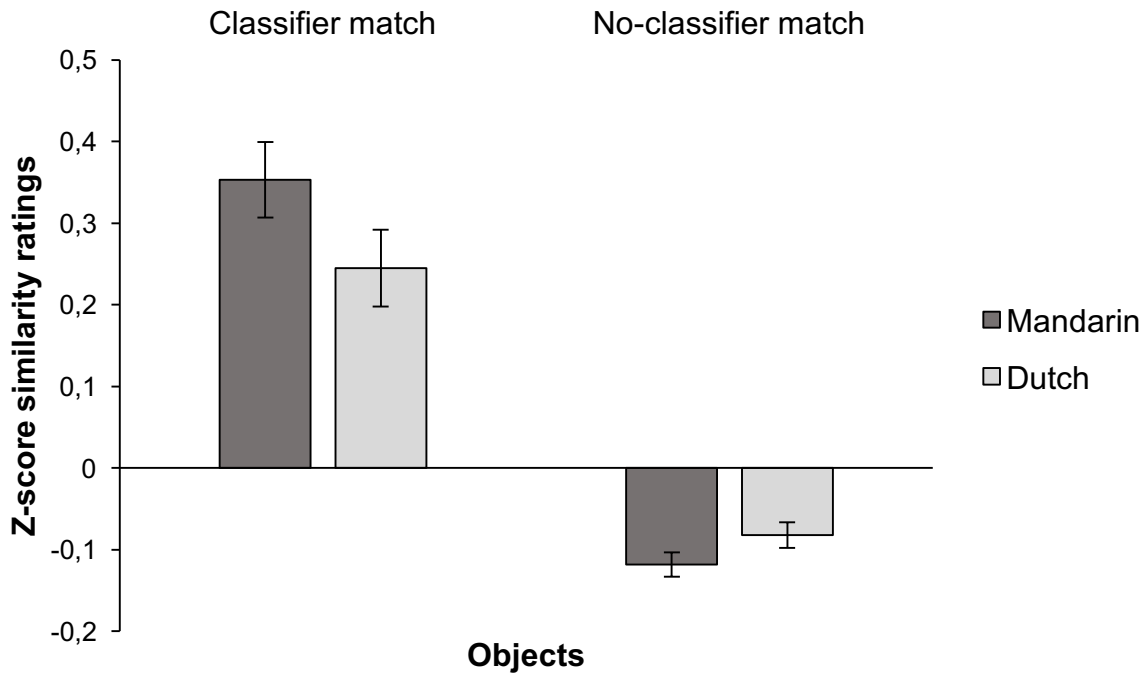

Figure 3. Word-picture similarity judgments (z-score) for Mandarin and Dutch speakers (bars = 1 *SE*).

4. Study 3: Word-word similarity

In Study 2, we again found a robust effect of shared classifier, and again found no difference between speakers of Mandarin and Dutch. In Study 3 we tested for a classifier effect in a linguistically maximal context, presenting target object and comparison objects as words.

4.1. Method

4.1.1. Participants

Twenty-five native speakers of Mandarin (10 female) and 24 native speakers of Dutch that did not take part in Study 1 or 2 were recruited (average age 22, range 19-38). All further details were the same as the previous experiments.

4.1.2. Stimuli

The stimuli were the same as the previous studies except participants were presented both target and comparison objects as words (Dutch nouns or Mandarin characters).

4.1.3. Procedure

As in Study 1 and 2, participants judged the similarity of the target object to the four comparison objects on a scale of 1 to 10.

4.2. Results

The data was analyzed as before; raw means by item can be found in Appendix D. There was a main effect of object type on similarity ratings $F_1(1, 47) = 161.83$, $p < .001$, $\eta^2_p = .78$; F_2 (1, 47) = 19.10, $p < .001$, $\eta^2_p = .29$, with the object sharing a classifier with the target receiving higher similarity ratings than the objects not sharing a classifier. Once again there was no main effect of language $F_1 < 1$; $F2 < 1$ or language by object type interaction $F_1 < 1$; $F_2 < 1$ (see Figure 4.)

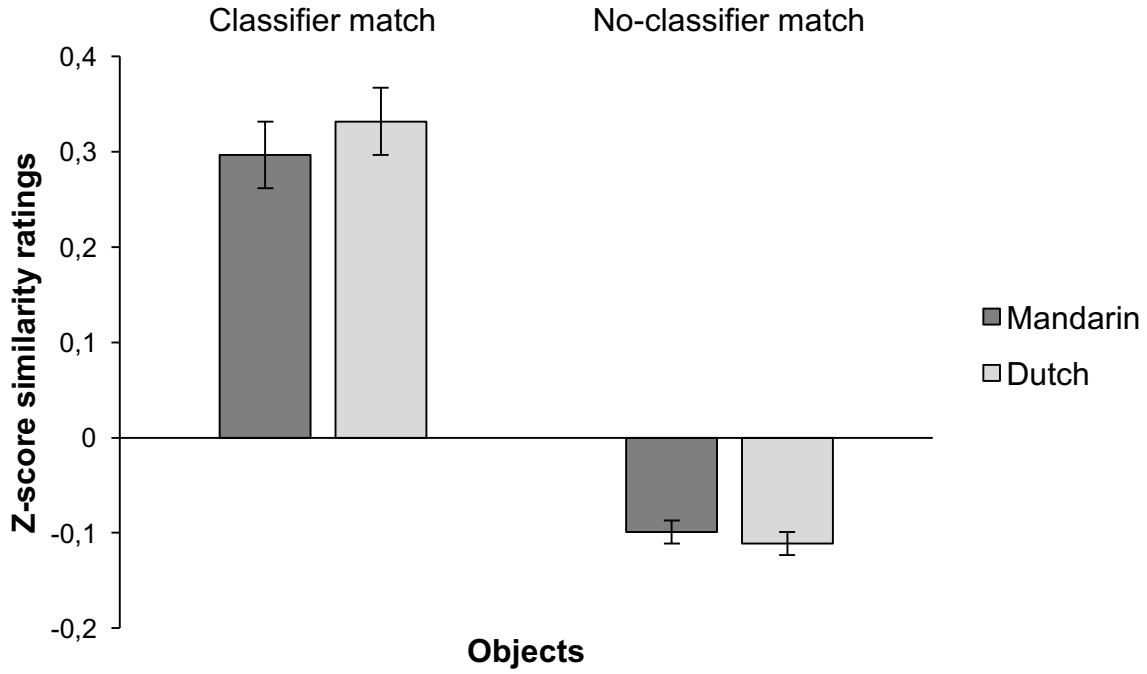

Figure 4. Word-word similarity judgments (z-score) for Mandarin and Dutch speakers (bars = 1 *SE*).

5. Intermediate Discussion: Studies 1–3

Across three experiments we found objects that share a classifier in Mandarin were judged as more similar than objects that do not share a classifier. We found this effect in speakers of Mandarin, but also in speakers of a non-classifier language, Dutch. Moreover, there was no difference in the magnitude of this effect between languages, contrary to previous studies that found a heightened effect in speakers of a classifier language compared to a non-classifier language (Saalbach & Imai, 2007; Schmitt & Zhang, 1998; Zhang & Schmitt, 1998).

Studies 1–3 show that objects that share a classifier are perceived as more similar to each other than objects that do not share a classifier. However, this does not give us a picture of the overall conceptual structure underlying objects in relation to classifier use. Therefore, as a final investigation of whether the conceptual organization of objects differed between Mandarin speakers and Dutch speakers—based on shared classifiers in Mandarin—we conducted a successive pile-sorting task using pictures of everyday objects. Instead of only judging similarity of pairs of objects (as in Studies 1–3) participants could compare similarity across all objects, and create clusters of objects in terms of their overall similarity. This allows us to view the similarity structure of the objects for Mandarin and Dutch speakers. This method has been used before to show that English, Chinese, and Spanish speakers' conceptual organization of bottles and jars is similar, whilst their naming of the same objects differs (Malt et al., 1999). Here we used a pile-sorting task for the first time to specifically assess the effect of classifier categories on the conceptual organization of a range of everyday objects.

6. Study 4: Successive pile-sorting task with pictures

6.1 Method

6.1.1. Participants

A different set of 23 Mandarin speakers (10 female) and 24 Dutch speakers participated in the study, none of whom had participated in the earlier studies.

6.1.2. Stimuli

Stimuli were selected from the larger set of stimuli described in Study 1. They were chosen so that they represented a wide class of objects, including food stuffs, artifacts and environmental objects (such as mountains and roads), and so that a sufficient number of exemplars were included within classifier categories. Objects shared classifiers within these object classes (e.g., food, artifacts) as well as across object classes (see Appendix E). The selected classifiers were high frequency and had high agreement across speakers in the norming task. The stimuli were forty line drawings of objects on 75mm by 75mm white laminated card. Forty was deemed a feasible number of objects for participants to sort so that the task did not become overwhelming or too time-consuming.

6.1.3. Procedure

A successive pile-sort was conducted in order to establish pairwise similarity between all stimuli (Boster, 1994). Pictures were laid out in a random order for participants who were told to sort them into groups based on how similar they were to each other. Participants were told they could use whatever characteristics they liked to make their piles, and they could make as many piles as they wished. After participants had finished their initial sort, they were then asked to join together the two piles that were the most similar. This was repeated successively until all piles were joined. Next, the original piles of the participant were reinstated and they were asked to divide one of their original piles into two. The participant then continued to divide all piles until all pictures were separated.

6.2. Results and discussion

Separate similarity matrices were constructed for each participant based on the rank order in which pictures were split apart (Boster, 1994). Aggregate matrices per language group were then computed by averaging individual matrices from Mandarin and Dutch participants. Separate cluster analyses were conducted on the Mandarin and Dutch matrices, using squared

Euclidean distance and the Ward method of linkage. Figures 5a and 5b show the outcomes from the sorting for Mandarin and Dutch speakers.

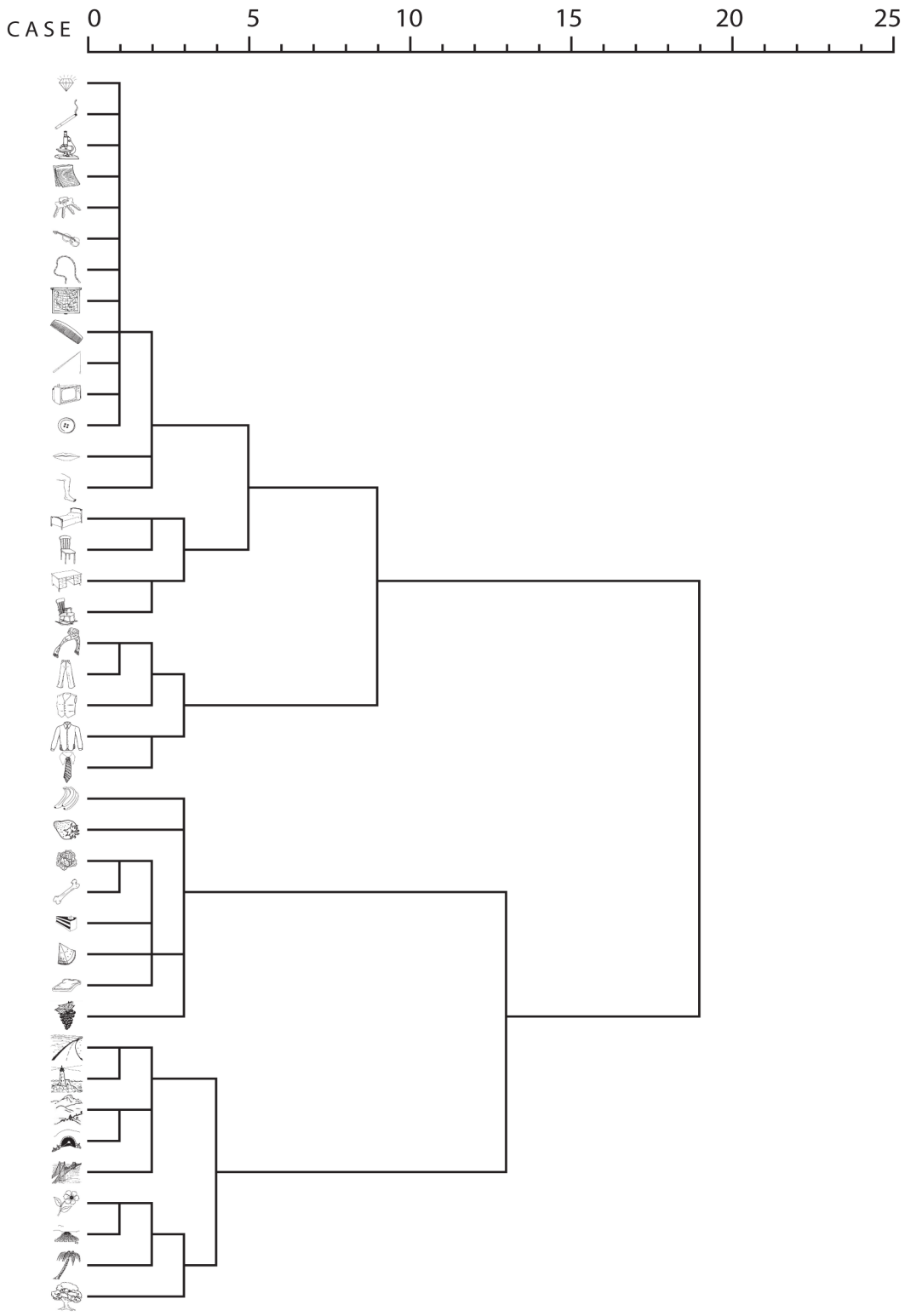

Figure 5a: Cluster analysis of pile-sorting by Mandarin speakers

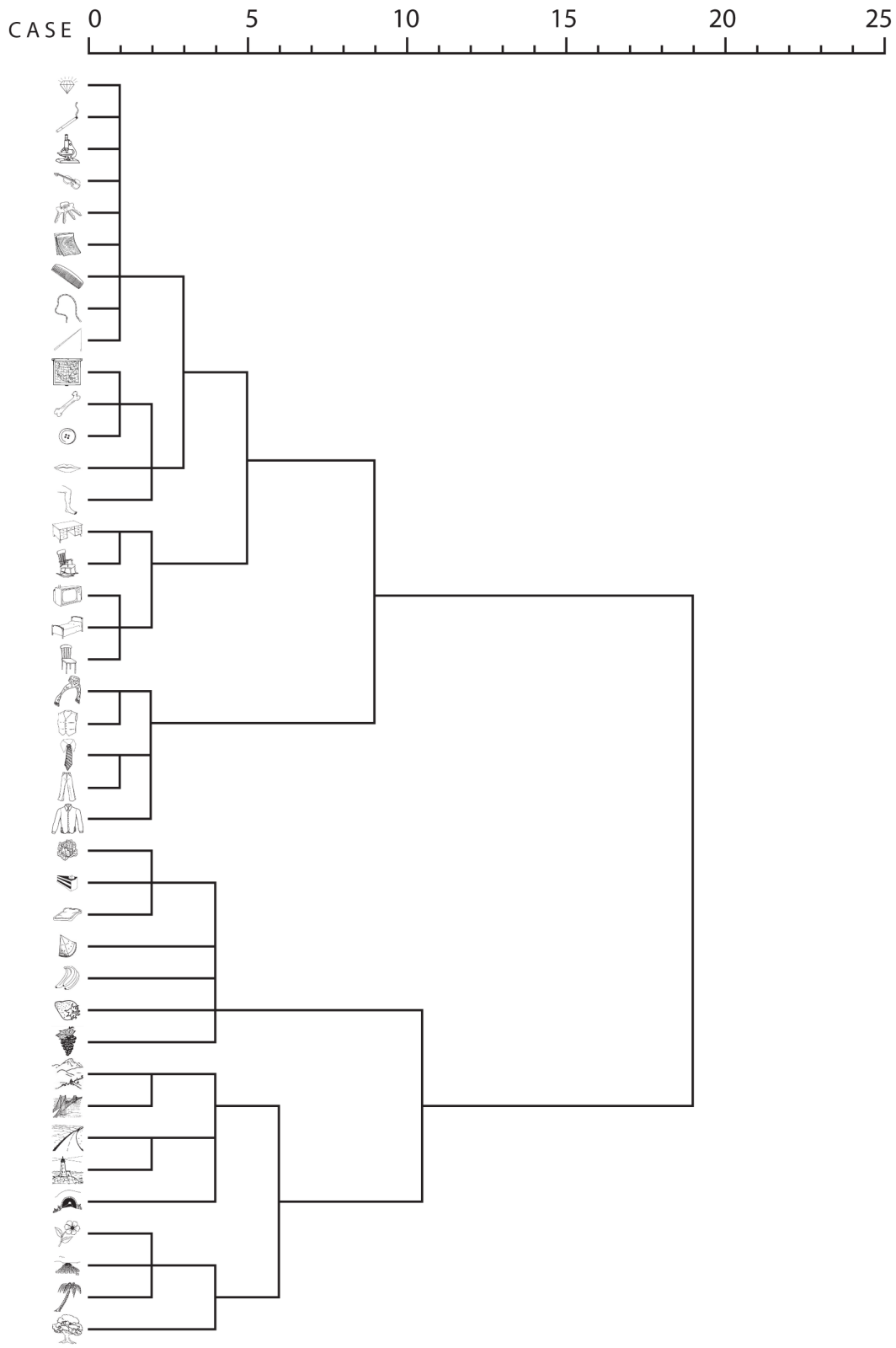

Figure 5b: Cluster analysis of pile-sorting by Dutch speakers

Both Mandarin and Dutch pile-sorting revealed a strong preference for grouping by taxonomic class. The main division of objects was into natural kinds vs. artifacts. Within these broad clusters further differentiation into plants, landscape objects, edible things, clothing, furniture and then a miscellaneous category of remaining objects was found. In order to evaluate how similar the groupings were for the two groups, we calculated the rank order of each stimulus pairs' proximity measure (which reflects the order in which pairs of objects combine in the trees in Figures 5a and 5b), following Baker (1974). These scores were used to calculate the rank correlation between Mandarin and Dutch groupings using Spearman's ρ, which showed a highly significant association $r_s = .89$, $p < .0001$. That is, Mandarin and Dutch speakers used the same strategy in grouping objects.

Nevertheless, we may still wish to ask whether Mandarin speakers are more likely to keep objects sharing a classifier in the same group in comparison to Dutch speakers. That is, within the taxonomic categories created, Mandarin speakers may still judge objects sharing a classifier to be more similar than Dutch speakers. In order to test this hypothesis directly, we classified each pair of objects as sharing a classifier or not and then calculated whether the proximities for objects was lower for Mandarin speakers than Dutch speakers (the lower the proximity the more similar pairs of objects are; the higher the more different). A 2 (language: Mandarin vs. Dutch) by 2 (classifier: shared vs. not shared) ANOVA revealed a main effect of language, $F(1, 778) = 6.54$, $p < .01$, $\eta^2p = .008$, and classifier, $F(1, 778) = 7.79$, $p < .005$, $\eta^2p = .01$, but crucially no interaction, $F(1, 778) = 2.92$, $p < .09$, $\eta^2p = .004$. That is, both Mandarin and Dutch speakers judged there to be more similarity between objects sharing a classifier than objects not sharing a classifier; but there was no added boost of perceived similarity for Mandarin speakers (see Figure 6).

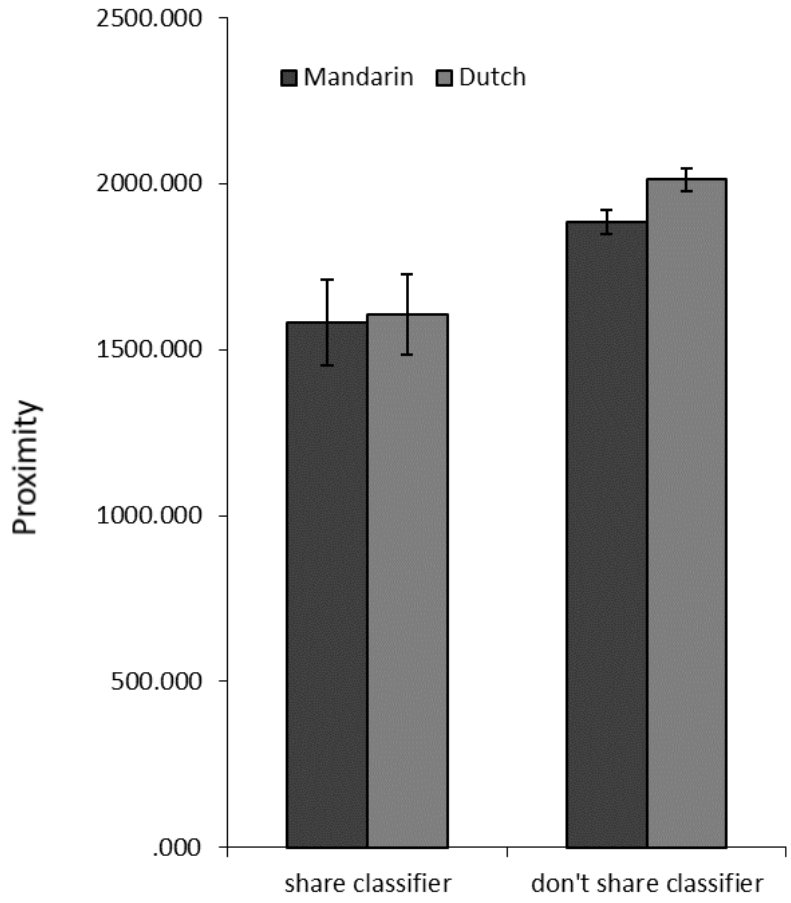

Figure 6: Comparison of Mandarin and Dutch proximity measures from the pile-sorting study.

7. General discussion

The present studies demonstrate speakers of a classifier language—Mandarin—and speakers of a non-classifier language—Dutch—represent object concepts similarly. Both groups of speakers judged two objects that share a classifier in Mandarin to be more similar than two objects that did not share a classifier, and contrary to previous studies (e.g., Saalbach & Imai, 2007, 2012), there was no difference between the languages. The same effect was observed using pictures, a combination of words and pictures, and words alone. Further, using a successive pile-sort task, we found the structure underlying object concepts was comparable for speakers of Dutch and Mandarin, and tended to follow taxonomic class. Thus, the most

parsimonious explanation is that classifier categories in Mandarin do not affect object conceptualization, but reflect it. As Malt (2019) notes, the way people understand the world may vary less across speakers of different languages than their linguistic differences would suggest—at least in this arena.

Our results are somewhat different to key previous findings. Similarity ratings have been used to suggest that speakers of non-classifier languages are sensitive to the organization of classifier categories, but that speakers of classifier languages are more sensitive. In line with our findings, speakers of both a classifier language and a non-classifier language rated pairs of objects with the same classifier as more similar than pairs of objects with different classifiers (Saalbach & Imai, 2007, 2012; Schmitt & Zhang, 1998; Zhang & Schmitt, 1998), suggesting classifiers reflect real-world structure. But, in previous studies this effect was shown to be larger in speakers of a classifier language (Saalbach & Imai, 2007, 2012; Schmitt & Zhang, 1998; Zhang & Schmitt, 1998), suggesting an effect specifically of classifiers, over and above real-world conceptual structure. In contrast, our studies found no such difference.

There are a number of potential factors that could explain why previous studies found differences in object conceptualization between speakers of a classifier language and speakers of a non-classifier language, but we did not. One factor we initially set out to test was the nature of the stimuli (i.e., pictures vs. words), but we found that neither pictures nor words revealed differences between Mandarin and Dutch speakers. It is possible that linguistic effects are only revealed when the classifier is explicitly presented. Using a visual world paradigm, Huettig et al. (2010) found that people looked more at objects that shared a classifier with a spoken word than at objects that did not share a classifier with the word, but only when the classifier was explicitly presented. Similarly, Gao and Malt (2009) found a heightened effect of classifiers in a condition when the classifier was explicitly presented: in a word recall task, Mandarin speakers were more likely than English speakers to recall words in

clusters of shared classifiers when the classifier was available. Notably, this effect was only observed for well-defined classifiers, i.e., classifiers that pick out a clear feature of objects, according to Gao and Malt (2009).

This resonates with what has been observed in the study of grammatical gender. Vigliocco et al. (2004) found effects of grammatical gender in speech substitution errors when nouns were produced with determiners that were marked for grammatical gender, but not when nouns were presented alone or with an indefinite determiner with no gender marking. Taken together, this is consistent with the proposal that grammatical information affects thought when it is explicitly presented; i.e., a thinking-for-speaking effect. However, a number of studies report classifier effects even when classifiers are not explicitly presented (e.g., Schmitt & Zhang, 1998; Zhang & Schmitt, 1998) suggesting additional factors are at play.

Effects of classifiers have been described as subtle and dependent on the specific task (Tsang & Chambers, 2011). It is possible that our studies did not make classifier information sufficiently salient to be applied in these tasks. In previous studies of similarity judgments, participants were asked to compare a target object with one or two comparison objects (e.g., Saalbach & Imai, 2007, 2011). In the present studies, however, participants were asked to compare a target object with four objects within the same trial. Although judgements were made between two objects at a time (as in previous studies), the simultaneous presence of three other comparison objects that did not share a classifier within the same trial may have reduced the salience of classifier categories overall. The relevant point here is the salience of the grammatical category within the context, rather than the number of objects to be compared per se. For example, Zhang and Schmitt (1998) and Schmitt and Zhang (1998) asked participants to judge the similarity between pairs of objects in sets of four, but where pairs of

objects matched in classifier. In that context classifier overlap becomes more salient, since every trial includes two pairs of items with a shared classifier.

The act of rating similarity may also reduce the salience of classifiers. Since classifier categories reflect shared real-world features, the similarity between the target object and the object with the shared classifier may be so strong that the additional shared classifier confers no further boost to judgments. This is in line with the findings of Saalbach and Imai (2007) who found that taxonomic relations between objects—rather than thematic or classifier relations—dominated performance across various tasks (i.e., categorization, similarity judgment, property induction, and word-picture priming). When Saalbach and Imai (2007) manipulated classifier relations orthogonally to taxonomic ones—thereby making classifiers more salient—they did find a weak effect of classifier. Although we chose objects that were not semantically related or visually similar to the target, there can be additional similarity between objects in terms of real-world features. For example, a *rocking chair* and *shovel* (compared to the distractors *zipper*, *badge*, and *vase*) both involve repetitive movement of the body, and a *leaf* and *snowflake* (compared with the distractors *lamp*, *pyramid*, and *violin*) are found in nature and fall to the ground.

If taxonomic relations maximize object similarity in the present tasks, it is possible that effects may only occur in more automatic, low-level tasks that requires less explicit rumination. Contrary to this proposal, Saalbach and Imai (2007, 2012) found no additional advantage for classifier similarity in a speeded word-picture matching task, although they did find effects in similarity judgments. However, using a different speeded task that prevented strategic responding, Srinivasan (2010) did find a larger effect of classifier category in Mandarin speakers compared to Russian and English speakers. Using a visual search task, Srinivasan asked participants to count the number of objects in a display containing distractor objects. The distractor objects either shared a Mandarin classifier or did not. Search

performance in Mandarin speakers was slower when the distractors shared a classifier with the target than when they did not, but this was not the case for the English and Russian speakers. So classifier information interfered with search performance, suggesting Mandarin speakers considered objects with the same classifier as more similar than objects with a different classifier.

The exact mechanisms at play here are unclear. The interference effect found by Srinivasan could be the result of lexical activation of the target and distractors (Meyer, Belke, Telling, & Humphreys, 2007) and their corresponding classifiers. An ERP study has previously shown that grammatical gender can be automatically activated during semantic categorization of pictures (Boutonnet, Athanasopoulos, & Thierry, 2012). Grammatical activation may not have occurred in the speeded word-picture matching task of Saalbach and Imai (2007, 2012) because pictures were only presented for 200ms; whereas Srinivasan's study had pictures presented for longer durations permitting the recruitment of linguistic encoding. Previous studies have shown that distractor objects that are homophones (Meyer et al., 2007) or phonological competitors (Görges, Oppermann, Jescheniak, & Schriefers, 2013) can interfere with visual search, but this does not necessarily imply that they are conceptually similar to the target. Similar conclusions have been made with regards to grammatical gender: Cubelli et al. (2011) found that semantic category judgments were facilitated for word pairs that were congruent in grammatical gender, but that this effect was absent under articulatory suppression. The authors concluded that effects of grammatical gender occur at the linguistic rather than the conceptual level. To summarize, the evidence seems to suggest that classifier effects are not observed in automatic tasks, unless the classifier is activated through linguistic encoding. Another possibility is that in the task of Srinivasan (2010), the act of counting led to stronger activation of classifier categories.

As a final consideration of the relevant factors for linguistic relativity effects in the domain of classifiers, we must consider the type of classifier. Gao and Malt (2009) propose that there are three types of classifier categories. First is the well-defined category. Within this category objects have a clear feature in common, for example *ke* classifies green growing objects with stems and leaves (Gao & Malt, 2009). Next is the prototype category in which a typical feature defines category membership, but where there is a gradient of typicality membership. For example, *tiao* is typically used to classify long and thin objects such as snakes or rope, but also includes less prototypical long, thin objects such as ships, news, and blankets (Gao & Malt, 2009). The final category is the arbitrary category in which there are no clear, typical defining features for membership. For example, *zun* classifies both large guns and statues of Buddah (Gao & Malt, 2009). Since objects within the well-defined category clearly share a feature, it could be predicted that speakers of a non-classifier language would judge objects in that category similarly to speakers of a classifier language. On the other hand, only speakers of a classifier language would judge objects within the arbitrary category as similar, because the only common feature is the grammatical one. Using this categorical distinction in a memory test, however, Gao and Malt (2009) found evidence for linguistic relativity for the well-defined category only. Our stimuli were not selected with these categories in mind, but on re-examination according to the categories set out by Gao and Malt (2009), the majority of the classifiers belong to the prototype category (18 out of 23, see Appendix F). So, it seems this category of classifiers is also not conducive for cross-linguistic effects of conceptualization. Future studies could systematically manipulate classifier category, rather than select classifiers according to high speaker agreement as in the current study.

Taking together all the evidence, we conclude that classifier categories reflect real-world conceptual structure, but when explicitly activated (e.g., due to their salience within a

task) they may boost judgments of conceptual similarity. This explanation is in line with a recent proposal that some effects of language on thought reflect simple statistical co-occurrences, rather than conceptual change (Samuel et al., 2019). This does not rule out the possibility that there are differences in how people think about objects, however. As described in the introduction, Lucy and Gaskins (2003) have shown that whether or not a language possesses numeral classifiers affects whether shape versus material substance is more essential in object concepts. The current tasks were designed to assess overall similarity between objects due to a shared classifier, and not to specifically address the importance of specific conceptual features in categorization (i.e., shape versus material). It could be possible that effects of classifier categories are more nuanced than what can be shown with a coarse measure of similarity.

A similar conclusion about the role of explicit grammatical activation could be made for grammatical gender, with some arguing that effects only occur at the lexical level (Cubelli et al., 2011; Ramos & Roberson, 2011; Vigliocco et al., 2005). With regards to the wider issue of linguistic relativity, however, it may be useful to consider the distinction Lucy (2016) makes between structure-centered approaches and domain-centered approaches. Structure-centered approaches to linguistic relativity focus on differences in grammatical structure (e.g., number marking, gender), whereas domain-centered approaches focus on linguistic differences within a semantic domain (e.g., color, space). Possibly structure-centered approaches produce more variable outcomes for linguistic relativity in part because they lead to a more heterogenous set of predictions for putative effects on thinking. It is also worth considering whether the present results generalize to other languages with a classifier system. Comparing again with grammatical gender, Vigliocco et al. (2005) found effects of grammatical gender in Italian, but not German. The authors pointed to the fact that German has three gender classes compared to two in Italian as an explanation for the difference (see

also Sera et al., 2002). Similarly, there is variability across languages in terms of the semantics, size, coherence, and use of classifier categories—for example, Mandarin has a higher frequency of classifier use than Japanese (Saalbach & Imai, 2012). In the future, combining typological and experimental studies will be critical to assess the existence and scope of classifier effects on thought.

8. Conclusion

Earlier studies have suggested that speaking a classifier language affects object conceptualization. In the present set of studies, however, we failed to find any difference between speakers of a classifier (Mandarin) and non-classifier (Dutch) language, suggesting any effects of language on thought in this arena are more restricted than the previous literature may lead us to conclude. Based on the present results, we conclude that classifier categories do not affect the overall similarity of object concepts. Instead, our study shows that the broad organization of object concepts remains comparable cross-linguistically, even in the face of attested reports to the contrary. Our study had comparable numbers of participants and a much larger set of stimuli than previous studies suggesting lack of effects in the current paper are not due to power. Instead, previous studies that found effects of linguistic relativity are likely to have made classifiers particularly salient in their tasks, or honed in on exactly those items that maximize differences. Our results show that classifier categories do not affect conceptualization of object similarity, even when linguistic stimuli are used. To conclude, classifier categories reflect, but do not affect the categories we experience in real-world interactions.

Acknowledgements

This work was funded by the Max Planck Gesellschaft, and the Ammodo KNAW Award awarded to AM. Thanks to Erica Renckens, Ludy Cilissen, Thijmen Sietsma, and Paul Hömke

for help preparing materials and collecting the Dutch data. We would also like to acknowledge the important contribution of Melissa Bowerman in the formative phase of this project.

Appendix A. Classifier, target, match and distractor objects used in Study 1-3.

Classifier	**Target**	**Match**	**Object 2**	**Object 3**	**Object 4**
ba	banjo	padlock	house	binoculars	bell
ba	comb	pick	pool	record	train
ba	key	Chinese fan	horseshoe	crown	submarine
ba	rocking chair	shovel	zipper	badge	vase
ba	scissors	chair	picture	nose	candle
ba	umbrella	knife	peacock	harp	statue
chuan	bananas	necklace	carpet	tweezers	elephant
chuan	grapes	keys	gun	rain	speaker
ding	hat	tent	palace	sewing machine	pen
dui	hay	fire	vacuum	arrow	light switch
duo	clouds	flower	step	spoon	lightning
fu	chess	handcuffs	cable	typewriter	mushroom
fu	saddle	glasses	tear	feather	crackers
gen	cigarette	cane	tape recorder	board	rake
gen	fishing pole	branch	bee	motorcycle	rule
gen	match	sausage	shirt	tractor	cage
jia	plane	ladder	earring	movie	potato
jian	vest	present	donkey	fountain	mousetrap

ke (棵)	celery	tree	potted plant	lipstick	organ
ke (颗)	diamond	button	truck	matchbox	mug
ke (棵)	palm tree	lettuce	spatula	church	bathtub
ke (颗)	strawberry	satellite	razor	duck	mop
ke (颗)	tooth	screw	envelope	frog	crib
kuai	bread	napkin	band aid	coat	well
kuai	eraser	piece of cake	refrigerator	cell phone	stethoscope
kuai	watermelon slice	field	lawnmower	popcorn	camel
mei	medal	ring	sweater	broom	fence
mian	mirror	flag	sheep	stairs	salt
pan	tape	spaghetti	pumpkin	skis	saxophone
pian	leaf	snowflake	lamp	pyramid	violin
shuang	chopsticks	shoes	globe	book	sailboat
tai	TV	microscope	rocket	hat	bricks
tiao	chain	dolphin	bomb	nut	helicopter
tiao	fish	towel	cookie	stroller	sword
tiao	jump rope	tunnel	butter	igloo	paperclip
tiao	leg	scarf	film roll	pig	cannon
tiao	pants	snake	tank	saw	tomato
tiao	road	tie	seesaw	lion	safe
tiao	tail	bracelet	car	suit	check
zhang	desk	paper	hammer	horse	magnet
zhang	map	bed	eye	bicycle	onion
zhang	poker card	spider web	log	curtains	pot

zhang	stamp	lips	toaster	boat	screwdriver
zhi	cat	pencil	bus	belt	accordion
zhi	chicken	ear	sandwich	bone	peanut
zhi	sock	rabbit	needle	package	lighthouse
zuo	castle	slide	cactus	rolling pin	spike
zuo	mountain	bridge	hinge	peas	axe

Appendix B. Mean Similarity Ratings Study 1

	Dutch		**Mandarin**	
Item	**Match**	**Non-match**	**Match**	**Non-match**
banana	2.50	2.57	2.48	2.88
banjo	2.21	3.54	3.20	3.17
bread	3.13	2.46	3.32	2.99
castle	3.88	2.36	4.84	3.44
cat	2.00	3.13	1.56	2.40
celery	5.58	3.44	5.64	3.68
chain	2.29	4.35	3.80	3.36
chess	3.58	2.10	3.72	3.08
chicken	3.71	5.88	3.52	3.17
chopsticks	3.38	3.11	3.92	2.57
cigar	4.96	2.24	4.28	2.81
cloud	4.71	4.40	4.40	3.55
comb	2.79	2.40	2.56	2.72
desk	5.46	3.07	4.08	3.09

diamond	2.54	2.22	3.12	2.60
fish	2.33	2.35	2.60	2.79
fishing pole	5.04	2.92	4.68	3.12
grapes	2.17	2.13	4.48	2.68
gum	2.13	2.51	4.52	3.97
hat	4.04	3.29	5.44	3.15
hay	5.38	2.39	6.68	2.56
jump rope	2.63	2.67	4.24	3.60
key	1.71	4.03	3.52	3.59
leaf	5.58	2.40	5.24	2.93
leg	3.67	3.22	3.76	3.75
map	3.00	3.49	3.72	4.05
match	2.83	2.74	3.28	2.41
medal	6.33	2.46	4.80	3.09
mirror	1.96	2.18	3.08	2.89
mountain	5.46	3.10	5.20	3.13
palm tree	5.13	2.61	5.20	3.56
pants	2.88	2.33	2.48	2.89
plane	4.04	2.13	4.84	2.53
poker	2.17	2.49	2.52	3.11
road	3.25	2.51	4.28	2.88
rocking chair	2.25	2.60	2.52	2.11
saddle	2.38	2.71	3.32	2.73

scissors	2.00	2.01	2.24	2.73
sock	3.08	3.01	3.36	2.96
stamp	3.88	2.14	3.04	2.75
strawberry	1.50	2.26	3.04	2.57
tail	3.21	2.61	3.40	2.49
tape	2.71	3.72	2.96	3.61
tooth	3.79	2.54	4.72	3.33
TV	5.00	3.18	3.64	3.57
umbrella	2.38	2.93	2.20	4.08
vest	4.00	2.17	4.20	2.91
watermelon	4.54	3.69	4.80	3.27

Appendix C. Mean Similarity Ratings Study 2

	Dutch		**Mandarin**	
Item	**Match**	**Non-match**	**Match**	**Non-match**
banana	1.83	2.44	3.32	3.57
banjo	2.04	3.29	2.80	3.67
bread	3.17	2.24	5.40	3.64
castle	3.38	2.06	5.04	2.63
cat	1.58	2.75	2.36	2.83
celery	5.17	3.14	5.52	3.99
chain	1.83	3.72	3.36	3.63
chess	2.46	1.99	3.20	3.63
chicken	2.79	4.39	2.88	2.97

chopsticks	2.17	2.50	4.40	2.71
cigar	4.29	1.94	5.40	4.32
cloud	3.83	4.04	4.32	3.67
comb	3.00	2.07	4.08	2.99
desk	4.96	2.58	4.60	3.00
diamond	3.42	1.78	5.40	3.55
fish	2.29	2.28	2.64	3.20
fishing pole	5.25	2.78	5.44	3.19
grapes	2.21	1.78	4.72	3.56
gum	1.92	2.33	5.88	4.11
hat	3.08	2.76	6.36	2.96
hay	4.96	2.03	5.92	2.33
jump rope	1.92	2.07	3.56	3.76
key	1.96	3.28	5.00	3.75
leaf	5.04	2.31	5.76	3.08
leg	2.50	3.00	4.44	3.59
map	2.17	3.29	3.84	3.81
match	2.42	2.44	3.36	3.05
medal	6.08	1.92	5.80	3.00
mirror	1.83	2.03	4.80	2.64
mountain	5.54	2.56	6.44	3.45
palm tree	5.04	2.39	5.12	2.92
pants	2.25	1.69	3.48	2.61
plane	3.50	2.00	4.92	3.24

poker	2.38	2.35	4.04	3.20
road	2.75	2.38	3.96	2.92
rocking chair	2.33	1.76	2.64	2.61
saddle	1.83	2.36	3.84	2.93
scissors	1.75	1.83	2.72	2.85
sock	2.71	2.58	3.80	2.93
stamp	3.04	1.97	3.16	3.63
strawberry	1.71	1.83	3.44	2.69
tail	2.58	1.74	3.92	2.24
tape	2.08	3.10	4.76	3.13
tooth	3.33	2.06	5.08	3.25
TV	4.58	2.51	4.68	3.45
umbrella	2.08	2.81	3.28	4.13
vest	3.71	1.88	2.76	2.93
watermelon	4.92	3.36	5.04	3.16

Appendix D. Mean Similarity Ratings Study 3

	Dutch		**Mandarin**	
Item	**Match**	**Non-match**	**Match**	**Non-match**
banana	1.54	2.01	2.12	2.99
banjo	1.63	2.40	2.48	3.53
bread	4.04	1.65	4.12	2.33
castle	2.67	1.99	4.00	2.43

cat	1.25	2.13	2.48	2.83
celery	4.63	2.71	5.60	3.27
chain	1.83	2.33	2.20	3.61
chess	1.33	1.92	2.44	2.37
chicken	2.33	3.63	3.48	3.89
chopsticks	1.38	1.69	2.76	2.25
cigar	3.33	1.58	3.72	2.44
cloud	3.38	3.07	5.40	3.35
comb	2.75	1.51	3.92	2.25
desk	4.75	2.04	5.80	2.68
diamond	3.21	1.72	4.76	2.64
fish	1.79	2.13	2.36	3.21
fishing pole	5.67	2.24	6.08	3.08
grapes	1.79	1.78	2.00	2.72
gum	1.50	1.69	3.28	3.17
hat	3.88	2.40	5.28	3.64
hay	3.79	1.74	6.68	2.37
jump rope	2.04	1.64	2.68	2.57
key	1.67	2.49	2.36	3.13
leaf	4.50	2.01	4.68	2.81
leg	2.29	2.31	2.36	3.87
map	2.08	2.93	3.72	3.31
match	2.63	1.68	4.00	2.80
medal	4.88	1.75	4.68	2.01

mirror	2.17	1.68	3.32	2.41
mountain	4.50	1.88	4.48	3.61
palm tree	4.17	1.78	4.24	2.60
pants	1.75	1.42	2.52	2.24
plane	3.00	1.76	4.40	2.60
poker	1.79	1.75	3.36	2.92
road	1.58	1.94	3.56	2.77
rocking chair	2.38	1.94	2.84	2.09
saddle	2.13	2.04	2.60	2.51
scissors	2.13	2.08	2.88	2.76
sock	2.04	2.06	2.28	3.60
stamp	3.71	1.57	2.68	2.83
strawberry	1.50	1.57	2.16	1.67
tail	2.67	1.65	3.12	2.17
tape	1.71	2.53	2.44	2.99
tooth	2.54	1.64	3.80	2.25
TV	4.83	2.13	4.92	3.17
umbrella	1.63	2.50	2.92	2.80
vest	3.38	1.24	3.88	2.29
watermelon	2.83	2.31	4.68	2.64

Appendix E. Items for successive pile-sort in Study 4.

Classifier	**Object**
ba	rocking chair

ba	comb
ba	chair
ba	violin
chuan	grapes
chuan	bananas
chuan	keys
duo	flower
gen	bone
gen	fishing pole
gen	cigar
gen	jump rope
jian	shirt
jian	vest
ke (棵)	palm tree
ke (棵)	tree
ke (棵)	lettuce
ke (颗)	button
ke (颗)	strawberry
ke (颗)	diamond
kuai	field
kuai	bread piece
kuai	cake piece
kuai	watermelon slice

tai	TV
tai	microscope
tiao	leg
tiao	scarf
tiao	tie
tiao	pants
tiao	road
tiao	tunnel
zhang	bed
zhang	desk
zhang	map
zhang	lips
zhang	paper
zuo	mountain
zuo	bridge
zuo	lighthouse

Appendix F. Classifier categories and definitions according to classifier types defined in Gao & Malt (2009). Our additions are in bold.

Well-defined

duo	flowers, white cloud
ke 棵	all plants with stems and leaves (the whole plant), such as tree, grass, corn, cabbage

Prototype

ba	[a handle] things that have a handle, such as umbrella, pistol, teapot, knife, screwdriver, scissors, pliers, hammer, spoon, broom, violin, chair, key, ruler
chuan	**things strung together or a sequence of events/sounds, e.g. pearls, keys, fire crackers, peppers, fruits**
ding	[crown of the head, top] something that has a top, such as cap, hat, straw hat, tent, mosquito netting **(prototype because cannot be used with umbrella and mountain)**
dui	**a pile of things**
gen	[root (of a plant), indicating a stick-shape object] stick, chopstick, straw, candle, finger, hair, needle, thread, rope, nerve, pencil
jia	[a frame, stand] things that have a frame, such as airplane, space shuttle, helicopter, ladder, eye glasses, machine, piano, accordion, electronic keyboard, camera
jian	[room] any rooms, including bedroom, living-room, kitchen, bathroom, study, office, classroom, workshop
ke 颗	[something small and roundish in shape] pearl, soy bean, button, tooth, mine, bullet, bomb, star, (man-made) satellite
kuai	[a lump-shape thing] soap, candy, cake, meat, stone, wrist watch, cloth, handkerchief, lawn, farming field, white/dark cloud
mei	coin, badge, medal, stamp, missile
mian	[surface] mirror, silk banner, flag, wall, big drum
pan	[a plate] magnetic audio tape, video tape, mosquito-repellent incense (coiled in a shape of a plate), grinding stone, chess match

pian	[a flat, thin piece, slice, or a stretch of land] bread, meat, tree leaf, snow flake, farming field, desert, forest, white/dark cloud
tai	[fetus] boy, girl, twins, also used for animals, such as piglets, puppies, etc.
tiao	[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
zhang	[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
zuo	[seat, stand, pedestal, base] bell, stone tablet, pagoda, bridge, house, temple, building, factory, church, grave, reservoir, forest, mountain, village, city
shuang	**a pair of things such as chopsticks, shoes**
Arbitrary	
fu	[the width of cloth (a bolt of)] picture, painting, ad, poster, map
zhi 支	[tree branch, twig] tree branch, match, pencil, pen, cigarette, arrow, gun
zhi 只	[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

References

Adams, K. L., & Conklin, N. F. (1973). Toward a theory of natural classification. In *Papers from the Ninth Regional Meeting, Chicago Linguistic Society* (pp. 1–10). Chicago: Chicago Linguistic Society.

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

Baker, F. B. (1974). Stability of two hierarchical grouping techniques case 1: Sensitivity to data errors. *Journal of the American Statistical Association*, *69*(346), 440–445.

Bender, A., Beller, S., & Klauer, K. C. (2016). Lady Liberty and Godfather Death as candidates for linguistic relativity? Scrutinizing the gender congruency effect on personified allegories with explicit and implicit measures. *The Quarterly Journal of Experimental Psychology*, *69*(1), 48–64. https://doi.org/10.1080/17470218.2015.1021701

Bender, A., Beller, S., & Klauer, K. C. (2018). Gender congruency from a neutral point of view: The roles of gender classes and conceptual connotations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(10), 1580–1608. https://doi.org/10.1037/xlm0000534

Boroditsky, L., & Schmidt, L. A. (2000). Sex, syntax, and semantics. *22nd Annual Meeting of the Cognitive Science Society*. Presented at the Austin, TX. Austin, TX: Cognitive Science Society.

Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, syntax, and semantics. In *Language in mind: Advances in the study of language and thought* (pp. 61–80). Cambridge, MA: MIT Press.

Boster, J. (1994). The successive pile sort. *Field Methods*, *6*(2), 11–12. https://doi.org/10.1177/1525822X9400600205

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79. https://doi.org/10.1016/j.brainres.2012.08.044

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

Clark, E. V. (1976). Universal categories: On the semantics of classifiers and children's early word meanings. In *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday* (Vol. 3). Saratoga, CA: Anma Libri.

Corbett, G. G. (2006). Gender, grammatical. *The Encyclopedia of Language and Linguistics. 2nd Edition. Oxford: Elsevier*, 749–756.

Croft, W. (1994). Semantic universals in classifier systems. *Word*, *45*(2), 145–171. https://doi.org/10.1080/00437956.1994.11435922

Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The effect of grammatical gender on object categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 449–460. https://doi.org/10.1037/a0021965

Foundalis, H. E. (2002). Evolution of gender in Indo-European languages. *Proceedings of the 24th Annual Conference of the Cognitive Science Society, Fairfax, VA*, 304–309.

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, *24*(7–8), 1124–1179. https://doi.org/10.1080/01690960802018323

Görges, F., Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2013). Activation of phonological competitors in visual search. *Acta Psychologica*, *143*(2), 168–175. https://doi.org/10.1016/j.actpsy.2013.03.006

Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze

behavior, but only during linguistic processing. *Journal of Cognition and Culture*, *10*(1), 39–58. https://doi.org/10.1163/156853710X497167

Imai, M., & Saalbach, H. (2010). Categories in mind and categories in language: Do classifier categories influence conceptual structures? In *Words and the mind: How words capture human experience* (pp. 138–164). New York: Oxford University Press.

Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2014). All giraffes have female-specific properties: Influence of grammatical gender on deductive reasoning about sex-specific properties in German speakers. *Cognitive Science*, *38*(3), 514–536. https://doi.org/10.1111/cogs.12074

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

Lucy, J. (1996). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. A. (2016). Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning*, *66*(3), 487–515. https://doi.org/10.1111/lang.12195

Lucy, J. A., & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In *Language in mind: Advances in the study of language and thought.* (pp. 465–492). Cambridge, MA: MIT Press.

Malt, B. C. (2019). Words, thoughts, and brains. *Cognitive Neuropsychology*, 1–13. https://doi.org/10.1080/02643294.2019.1599335

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*(2), 230–262. https://doi.org/10.1006/jmla.1998.2593

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14*(4), 710–716. https://doi.org/10.3758/BF03196826

Mickan, A., Schiefke, M., & Stefanowitsch, A. (2014). Key is a llave is a Schlüssel: A failure to replicate an experiment from Boroditsky et al. 2003. *Yearbook of the German Cognitive Linguistics Association*, *2*(1). https://doi.org/10.1515/gcla-2014-0004

Olness, G. S. (1991). Semantic correlates of three morphosyntactic classifier types: A cross linguistic survey. *Trabajo de Investigación No Publicado, University of Oregon.*

Perniss, P., Vinson, D., Seifart, F., & Vigliocco, G. (2012). Speaking of shape: The effects of language-specific encoding on semantic representations. *Language and Cognition*, *4*(3), 223–242. https://doi.org/10.1515/langcog-2012-0012

Pulvermüller, F., & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *NeuroImage*, *20*(1), 159–172. https://doi.org/10.1016/S1053-8119(03)00261-1

Ramos, S., & Roberson, D. (2011). What constrains grammatical gender effects on semantic judgements? Evidence from Portuguese. *Journal of Cognitive Psychology*, *23*(1), 102–111. https://doi.org/10.1080/20445911.2011.466795

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

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General*, *136*(3), 485–501. https://doi.org/10.1037/0096-3445.136.3.485

Saalbach, H., & Imai, M. (2012). The relation between linguistic categories and cognition: The case of numeral classifiers. *Language and Cognitive Processes*, 1–48. https://doi.org/10.1080/01690965.2010.546585

Samuel, S., Cole, G., & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-019-01652-3

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

Seifart, F. (2010). Nominal classification. *Language and Linguistics Compass*, *4*(8), 719–736. https://doi.org/10.1111/j.1749-818X.2010.00194.x

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, 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*(3), 377–397. https://doi.org/10.1037//0096-3445.131.3.377

Slobin, D. (1987). Thinking for speaking. *Annual Meeting of the Berkeley Linguistics Society*, *13*, 435–445.

Speed, L. J., & Majid, A. (2019). Linguistic features of fragrances: The role of grammatical gender and gender associations. *Attention, Perception, & Psychophysics*, *81*, 2063–2077. https://doi.org/10.3758/s13414-019-01729-0

Srinivasan, M. (2010). Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese. *Language and Cognition*, *2*(02), 177–190.

Tsang, C., & Chambers, C. G. (2011). Appearances aren't everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1065–1080. https://doi.org/10.1037/a0023601

Vigliocco, G., Vinson, D. P., Indefrey, P., Levelt, W. J. M., & Hellwig, F. (2004). Role of grammatical gender and semantics in German word production. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *30*(2), 483–497. https://doi.org/10.1037/0278-7393.30.2.483

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*(4), 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

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*(4), 375–385.