Do classifier categories affect or reflect object concepts?

Laura J. Speed[1] **Jidong Chen[2]** **Falk Huettig[3]** **Asifa Majid[1,3,4]**
(l.speed@let.ru.nl) (jchen@csufresno.edu) (Falk.Huettig@mpi.nl) (asifa.majid@let.ru.nl)
[1]Centre for Language Studies, Radboud University, Nijmegen, NL
[2]California State University, Fresno, USA
[3]Max Planck Institute for Psycholinguistics, Nijmegen, NL
[4]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, NL

Abstract

We conceptualize objects based on sensorimotor information gleaned from real-world experience. To what extent is conceptual information structured according to higher-level linguistic features? We investigate whether classifiers, a grammatical category, shape the conceptual representations of objects. In three experiments native Mandarin speakers (a classifier language) and native Dutch speakers (a language without classifiers) judged the similarity of a target object with four objects (presented as words or pictures). One object shared a classifier with the target, the other objects did not. Overall, the target object was judged as more similar to the object with the shared classifier than distractor objects in both Dutch and Mandarin speakers, with no difference between the two languages. Thus, even speakers of a non-classifier language are sensitive to object similarities underlying classifier systems, and using a classifier system does not exaggerate these similarities. This suggests that classifier systems reflect, rather than affect, conceptual structure.

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

Introduction

When asked to describe the similarity between a knife and a sword, one might describe their visual properties: they are both "*sharp*", "*metallic*", "*shiny*", etc. Likewise for the similarities between a saw and a pair of scissors, there are numerous similarities. The present research investigates whether the addition of a shared grammatical category between nouns would serve to increase conceptual similarity of objects too. Do conceptual representations reflect universal sensory and motor regularities? Or can they also be structured according to higher-level linguistic information?

For decades psychologists, linguists, and anthropologists have debated whether or not the language we speak can affect the way we think about the world. Evidence for some effect of language on thought has been provided for many domains, such as color (e.g., Davidoff, Davies, & Roberson, 1999; Gilbert, Regier, Kay, & Ivry, 2006; Winawer, Witthoft, Frank, Wu, Wade, & Boroditsky, 2007), spatial cognition (e.g., Levinson, 2003; Majid, Bowerman, Kita, Haun, & Levinson, 2004), and time (e.g., Boroditsky, 2001; Boroditsky, Fuhrman, McCormick, 2010), to name but a few.

One linguistic domain in which effects of language on thought have been extensively investigated is grammatical gender. Grammatical gender divides nouns into classes according to the behavior of associated words (e.g., articles, adjectives; cf. Corbett, 2006). In some languages, nouns are grammatically classified according to sex, i.e., masculine or feminine. For example *apple* is masculine in German, *der Apfel*, but feminine in French, *la pomme*. Grammatical gender tends to be semantically arbitrary for objects without a natural gender; however, it has been shown to affect how speakers of such languages think about objects. For example, Spanish and German speakers are more likely to ascribe male qualities to grammatically masculine objects and female qualities to grammatically feminine objects (Boroditsky, Schmidt, & Phillips, 2003).

We may ask, however, whether such linguistic features are, in fact, changing the conceptual structure of objects, or whether they are merely being used as strategic devices during language processing. One way to test this is to assess the effect of such grammatical systems using tasks that recruit language to a greater or lesser extent. Phillips and Boroditsky (2003) provided some evidence that grammatical gender affects object concepts for German and Spanish speakers. They found effects of grammatical gender on judgments of similarity between people and objects even when the task was completed in English (a language with no gender system), when performing a non-linguistic task (using pictures), and during a verbal interference task (suggesting real conceptual change). On the other hand, Bender, Beller, and Klauer (2011) suggest grammatical gender is only available as a syntactic property, and does not change conceptual representations. Lexical decisions to nouns were faster when they had been preceded by words matching in grammatical gender but not for words matching in semantic gender (natural gender). That is, congruence in the syntactic features (grammatical gender) facilitated responses, but congruence between grammatical gender and semantic gender (which would reflect conceptual similarity) did not affect response time. Recently, Bender, Beller, and Klauer (2016) have suggested grammatical gender effects may in fact be due to "personification" (with nouns being associated to personified allegories such as *Lady Liberty*), rather than grammatical information, per se.

Here we look at a different grammatical category associated with nouns, i.e., numeral classifiers. In languages with numeral classifiers, their use is obligatory when a specific quantity is indicated, with certain quantifiers, and with demonstratives. Numeral classifiers are said to refer to

1

a specific feature (e.g., material, shape, size) of the entity associated with the corresponding noun (cf., Allan, 1977). For example, the Mandarin classifier *tiao2* refers to long, rope-like objects such as legs, snakes, and rivers. Although classifier categories have underlying conceptual meanings, there can be large variability with respect to category size and coherence, with some classifiers covering a broad range of objects and taxonomic categories (e.g., Saalbach & Imai, 2012). So, although *tiao2* is used with "long" things, it can also be used for less prototypically long things, such as dogs, underwear and a piece of news.

It appears that speaking a language with a classifier system can affect how speakers of a language conceptualize objects. Lucy and Gaskins (2003), for example, compared speakers of Yucatec Maya (an indigenous language of Mexico that contains numeral classifiers) with English speakers. Speakers of the two languages judged the similarity of objects differently: Yucatec Mayans had a preference for matching objects according to material, whereas English speakers matched the same objects according to shape. Zhang and Schmitt (1998) and Saalbach and Imai (2007) found Mandarin speakers rated pairs of nouns that shared a classifier as more similar than nouns that did not share a classifier, but speakers of non-classifier languages (English and German) rated both pairs similarly.

But, do classifier systems really *affect* the conceptual organization of object concepts, or do they *reflect* conceptual organization instead? Classifiers are usually not arbitrarily related to features of objects. They pick out common features of the entities they classify (cf., Allen, 1977; Lakoff, 1987). There is evidence to suggest there is some conceptual salience to the features classifiers denote. For example, Clark (1976) noted the parallel between the conceptual features frequently found in classifier languages and the conceptual features salient to children in early word learning. Classifiers often denote features such as "*long*", "*round*", or "*animate*"; the exact features on which children base their over-generalization errors. No language has been found with a classifier system that distinguishes referents on the basis of color, and similarly children do not use color as a basis of their over-generalizations.

Thus, instead of classifiers influencing the organization of object concepts, classifiers could reflect the organization of object concepts, thus providing evidence for the effect of thought on language. Classifier categories may be structured around natural similarities in the world. If so, speakers of non-classifier languages should perceive relations between objects sharing classifiers similarly to those speaking classifier languages. Consistent with this, Saalbach and Imai (2005) found speakers of German, as well as Mandarin, judged objects sharing a classifier as more similar than pairs of objects that did not share classifiers.

If classifiers reflect universally recognized similarities between objects, how does this reconcile with previous findings of linguistic relativity (e.g., Zhang & Schmitt, 1998; Saalbach & Imai, 2007)? The answer may lie in the experimental details. The majority of previous experiments investigating effects of classifier systems present people with words from the participants' language; e.g., Mandarin speakers are presented with names of objects in Mandarin, and English speakers with names in English. So, in one sense, it is not surprising speakers of classifier languages are sensitive to the grammatical information associated with nouns. If we think of current psycholinguistic models of language use, then when a noun is activated grammatical information forming part of the lemma level representation would also be activated. Linguistic relativity effects could be explained by the activation of the relevant grammatical feature associated with the word form, rather than differences in the conceptual representation of objects, per se (cf., Kousta, Vinson, & Vigliocco, 2008). To test whether classifiers, in fact, affect the structure of object concepts or whether they serve only to facilitate processing of words with shared grammatical features, parallel experiments with linguistic and non-linguistic stimuli need to be conducted.

We investigated one classifier language – Mandarin – and one language without classifiers – Dutch. Participants were presented with a target object and had to rate its similarity to four other objects. Using a rating task means more fine-grained differences can be detected, compared to, for example a forced-choice task (e.g., Lucy & Gaskins, 2003; Saalbach & Imai, 2007, Experiment 1), where participants have to choose one object over another. One of the four objects used a noun that possessed the same classifier as the target object, the other three objects did not. If classifiers affect the way objects are thought of, we could expect speakers of Mandarin would judge the target object and classifier object as more similar than the other objects, but Dutch speakers would not, because having a shared classifier will increase perceived similarity. If, on the other hand, classifiers reflect real-world similarities amongst objects, then we would expect both Mandarin and Dutch speakers to judge the target object and classifier object as more similar than the distractor objects. A third possibility is that both Mandarin and Dutch speakers judge the classifier object as more similar to the target than the distractors, because of real-world similarities that the classifier system is built upon, but that this effect is greater in Mandarin than Dutch speakers (cf., Saalbach & Imai, 2007, Experiment 2 and 3). Further, if classifier systems affect object representations, rather than lexical representations, then we should see differences between Mandarin and Dutch speakers on judgments of pictures, as well as words. In the following three experiments we assess the effect of classifiers on similarity judgments with pictures and words.

Experiment 1. Picture-picture similarity

Method
Participants
25 native Mandarin speakers and 24 native Dutch speakers participated in the experiment (average age 23, range 18-

2

33). All speakers were familiar with other languages. Mandarin speakers came from different dialect backgrounds (e.g., Shandong, Henan, Shanxi, Jilin, Hubei, Jiangsu), but were all educated in Putonghua (standard Mandarin) and mainly use this language daily, both at work and at home. Dutch speakers were all multilingual with English, German, Spanish and French. None of the Dutch participants were familiar with Mandarin, or any other language with classifiers.
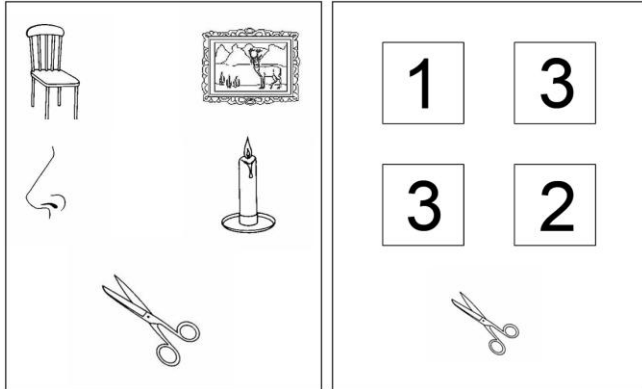

Figure 1. Example page of the stimulus booklet (left) with target (scissors), classifier match (chair) and comparison objects, with response sheet (right).

Stimuli
Line drawings of everyday objects, familiar to both groups of participants, were used. Each trial consisted of a target object and four comparison objects. The target object shared a classifier with one of the comparison objects. A norming study was first conducted in which a separate set of twelve native Mandarin speakers named 240 line drawings of concrete and imageable objects. Classifier choice for a noun was not strict (i.e., nouns could take different classifiers), but there was a clear dominant classifier for each object. Based on the norming results, we chose noun pairs that shared the same dominant classifier. The remaining three comparison objects were "distractors", and did not share a classifier with the target. On each trial, pictures of the four comparison objects were presented on a piece of paper, laid out in a two-by-two grid. The target object was centered under the comparison objects (see Figure 1). Participants were given a separate response sheet, consisting of four empty boxes in the same configuration as the comparison objects, with a picture of the target object at the bottom of the sheet. The experiment consisted of 48 trials. The classifier match appeared equally often in each of the positions within the gird, and the position of an each item was counter-balanced across participants.

Procedure
Participants were instructed, in their native language, to judge how similar each target object was to each of the other objects. They were asked to indicate their judgments in the corresponding boxes in their answer sheet, with 0 indicating "no similarity" and 10 "identical in similarity".

Results
Previous studies have shown Chinese participants give overall higher similarity ratings than Westerners (Saalbach & Imai, 2007). In order to control for any differences in how the similarity scale was used across individuals, we transformed each participant's rating scores into standardized z-scores. The transformed data were then analyzed with a language (Mandarin vs. Dutch) by object (classifier match vs. distractor 1 vs. distractor 2 vs. distractor 3) mixed ANOVA, treating participants and items as random effects.

There was a main effect of object type on similarity ratings $F_1(3, 141) = 40.05$, $p < .0001$, $\eta_{2p} = .46$; $F_2(3, 144) = 5.72$, $p < .0001$, $\eta_{2p} = .11$. Simple planned comparisons revealed the classifier match received higher similarity ratings than any of the other distractors: classifier vs. distractor-1 $F_1(1, 47) = 67.64$, $p < .001$, $\eta_{2p} = .6$; $F_2(1, 48) = 17.02$, $p < .001$, $\eta_{2p} = .26$, classifier vs. distractor-2 $F_1(1, 47) = 57.4$, $p < .001$, $\eta_{2p} = .557$; $F_2(1, 48) = 12.37$, $p < .001$, $\eta_{2p} = .21$, classifier vs. distractor-3 $F_1(1, 47) = 105.18$, $p < .001$, $\eta_{2p} = .69$; $F_2(1, 48) = 11.05$, $p = .002$, $\eta_{2p} = .19$. There were no significant differences between the distractor objects.

There was no significant main effect of language $F_1 < 1$; $F_2 < 1$. Crucially, there was no significant language by object type interaction $F_1(3, 141) = 1.4$, $p = .25$, $\eta_{2p} = .03$; $F_2 < 1$ (see Figure 2).
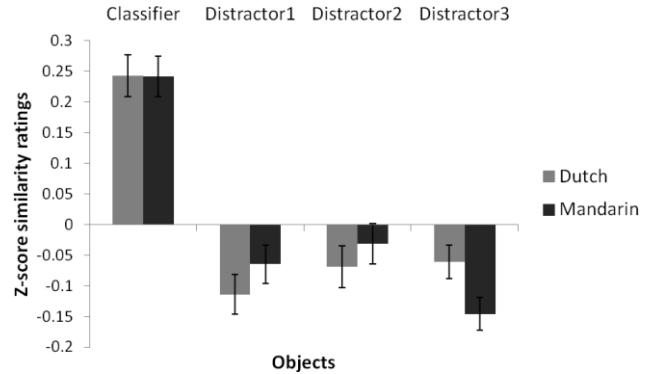

Figure 2. Mandarin and Dutch speakers' picture-picture similarity judgments (z-score) Experiment 1 (bars = 1SE).

Experiment 2. Word-picture similarity

We found a robust effect of classifier pair similarity, but no indication of a difference between Mandarin and Dutch speakers in Experiment 1. This suggests object representations are not affected by classifiers, but does not rule out lexical-level language-specific effects. In Experiment 2 we presented the target object as a word, in order to encourage activation of the classifier categorization.

Participants
A different set of 25 Mandarin speakers and 24 Dutch speakers (average age 23 years, range 18-28) participated in

3

the experiment. The demographic characteristics and recruitment procedure was the same as in Experiment 1.

Stimuli

The stimuli were as in Experiment 1 but now the target was presented as a word (without the classifier) in the native language and the comparison objects were presented as pictures.

Procedure

Participants were asked to judge similarity of the target object to the comparison objects on a scale from 1 (no similarity) to 10 (identical in similarity).

Results

Data was analyzed as in Experiment 1. There was a main effect of object type on similarity ratings $F_1(3, 141) = 52.92$, $p < .001$, $\eta_{2p} = .53$; $F_2(3, 144) = 8.66$, $p < .001$, $\eta_{2p} = .15$. Simple planned comparisons revealed the classifier match received higher similarity ratings than any of the other distractors: classifier vs. distractor-1 $F_1(1, 47) = 70.79$, $p < .001$, $\eta_{2p} = .6$; $F_2(1, 48) = 20.57$, $p < .001$, $\eta_{2p} = .30$, classifier vs. distractor-2 $F_1(1, 47) = 73.05$, $p < .001$, $\eta_{2p} = .61$; $F_2(1, 48) = 20.61$, $p < .001$, $\eta_{2p} = .30$, classifier vs. distractor-3 $F_1(1, 47) = 73.71$, $p < .001$, $\eta_{2p} = .61$; $F_2(1, 48) = 11.54$, $p < .001$, $\eta_{2p} = .19$. There were no significant differences between the distractor objects.

There was again no main effect of language $F_1 < 1$; $F_2 < 1$ and no significant language by object type interaction $F_1(3, 141) = 1.54$, $p = .21$ $\eta_{2p} = .03$; $F_2 < 1$ (see Figure 3).
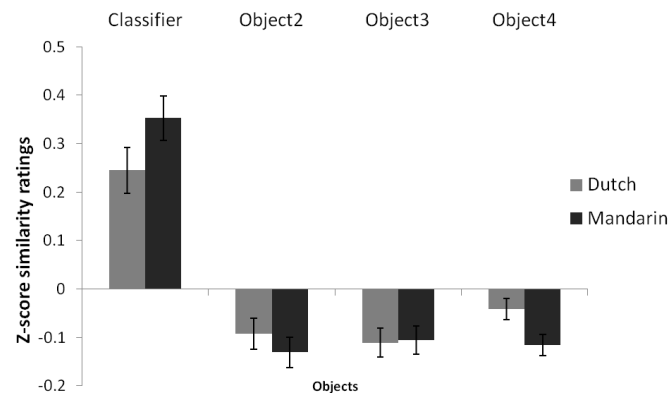

Figure 3. Mandarin and Dutch speakers' word-picture similarity judgments (*z*-score) Experiment 3 (bars = 1*SE*)

Experiment 3. Word-word similarity

Again, in Experiment 2 we find a robust classifier effect, but no difference between Dutch and Mandarin speakers. As a final test of a classifier effect we maximized the linguistic context by presenting both the target and the comparison objects as words.

Participants

A different set of 25 Mandarin speakers and 24 Dutch speakers participated in the experiment (average age 23

years, range 18-28). The demographic characteristics and recruitment procedure was as in Experiment 1 and 2.

Stimuli

The stimuli were as in Experiment 1 and 2, but participants were presented with words instead of pictures. Mandarin participants were presented with nouns in Mandarin characters and Dutch participants with the Dutch nouns.

Procedure

Participants were asked to judge similarity of the target object to the comparison objects on a scale from 1 (no similarity) to 10 (identical in similarity).

Results

Data was analyzed as in Experiment 1 and 2. There was a main effect of object type on similarity ratings $F_1(3, 141) = 52.92$, $p < .001$, $\eta_{2p} = .53$; $F_2(3, 144) = 9.6$, $p < .001$, $\eta_{2p} = .17$. Simple planned comparisons revealed the classifier match received higher similarity ratings than any of the other distractors: classifier vs. distractor-1 $F_1(1, 47) = 145.22$, $p < .001$, $\eta_{2p} = .6$; $F_2(1, 48) = 27.25$, $p < .001$, $\eta_{2p} = .36$, classifier vs. distractor-2 $F_1(1, 47) = 201.19$, $p < .001$, $\eta_{2p} = .81$; $F_2(1, 48) = 16.95$, $p < .001$, $\eta_{2p} = .26$, classifier vs. distractor-3 $F_1(1,47) = 80.56$, $p < .001$, $\eta_{2p} = .63$; $F_2(1, 48) = 7.99$, $p = .007$, $\eta_{2p} = .19$. There were no significant differences between the distractor objects.

There was again no main effect of language $F_1(3, 141) = 1.54$, p = .21, $\eta_{2p} = .03$; $F_2 < 1$ and crucially no significant language by object type interaction $F_1(3, 141) = 1.54$, $p = .21$ $\eta_{2p} = .03$; $F_2 < 1$ (see Figure 3).
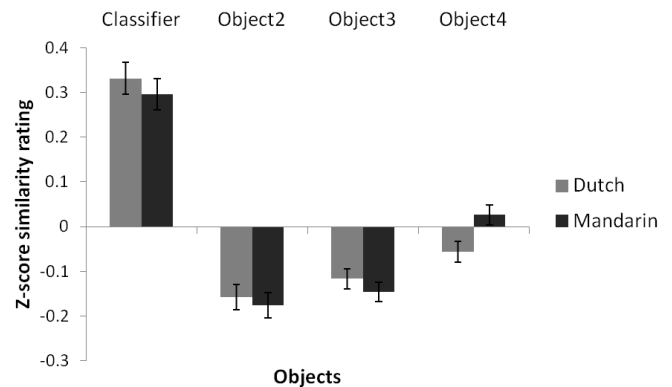

Figure 4. Mandarin and Dutch speakers' word-word similarity judgments (*z*-score) Experiment 2 (bars = 1*SE*).

Discussion

Across three experiments we found objects were judged as more similar to a target object when they shared a classifier in Mandarin compared to when they did not. However, this effect was observed both in a language that uses classifiers – i.e., Mandarin – as well as a language that does not have a classifier system – i.e., Dutch. Moreover, the magnitude of this effect did not differ between Mandarin and Dutch

4

speakers. From these results we conclude classifiers do not impact overall object similarity.

Why do some studies find differences in similarity ratings between languages with and without classifiers, but we do not? Our study focused on overall similarity of objects, and used stimuli (e.g., visual pictures) which emphasized shape. We did not use real objects, unlike Lucy and Gaskins (2003), who found what distinguished Yucatec from English speakers was attention to shape vs. material. It is possible, therefore, that we missed a critical ontological distinction. Nevertheless, other studies have focused on pictorial or linguistic stimuli and have reported linguistic relativity effects (Zhang & Scmitt, 1998; Saalbach & Imai, 2007). So this cannot be the whole story.

It is possible a shared classifier is more salient when there are fewer objects to judge. For example, when comparing only two objects at a time, with no distractors, a greater number of features of a word (such as grammatical category) can be attended to. However, when there are four objects to compare simultaneously, it is likely that only the most salient similarities are attended to. It is also possible that we failed to find an effect of language on classifier similarity judgments because we did not present the words with their classifiers, but as bare nouns. Huettig, Chen, Bowerman, and Majid (2010) found people look towards objects sharing a classifier with a spoken word more than visual distractors, but only when the classifier was explicitly used, and Gao and Malt (2009) found heightened classifier effects when a classifier was present in a sentence compared to when it was not. Similarly, Vigliocco, Vinson, Indefrey, Levelt, and Hellwig (2004) found grammatical gender effects in speech substitution errors only when nouns were produced with determiners marked for gender, but not for single noun phrases or noun phrases with indefinite determiners not marked for gender. Thus, grammatical information may need to be explicit and salient to affect categorization.

It is also possible that using similarity judgments as a measurement is not sensitive enough to reveal differences between Dutch and Mandarin speakers. That is, since classifiers are mostly built on shared real-world features, the similarity between the target object and the classifier object in terms of these features may be quite obvious for explicit similarity judgments. That is, in some sense, the judgments could be at ceiling level. Perhaps a more low-level, automatic task, closer to real-world language use, could reveal the linguistic advantage of a classifier system. However, Saalbach and Imai (2011) found the opposite results: Chinese speakers showed an enhanced classifier similarity effect compared to German speakers in a similarity rating task, but not a speeded word-picture matching task.

Similarly, Gao and Malt (2009) propose there are three classifier categories: "well-defined" categories in which there is a clear feature that all objects sharing the classifier possess; "prototype" categories in which there is a typical feature, but also a gradient of typicality in category membership; and "arbitrary" categories for which there are no typical features defining membership. One might predict that objects belonging to the "well-defined" classifier category could easily be grouped in terms of similarity by speakers of a non-classifier language, but only speakers of that specific classifier language could group the "arbitrary" category correctly. Thus, it is possible that the classifiers used in the present study fit more into the "well-defined" category. It could be predicted that an advantage for speakers of a classifier language could be found if the classifier pairs used shared "arbitrary" classifiers. However, Goa and Malt (2009) only found an advantage for the "well-defined" category amongst Mandarin speakers when testing recall of nouns with shared classifiers.

So, does language influence thought? Our results suggest that a grammatical feature, classifiers, does not have an influence on the way that objects are categorized, as measured by global similarity judgments. However, it is likely that other forms of linguistic information do affect the way objects are thought about. For example, effects of grammatical gender on object categorization appear to be well-attested. As noted above, classifiers are said to highlight only one, or a few, features of an object (e.g., "shape"; cf. Allen, 1977), and not information relevant to the entire concept. This contrasts with grammatical gender, which could affect the way all features of an object are conceptualized. For example, Spanish speakers described the word *key*, a noun with female grammatical gender, with female characteristics "*golden, intricate, little, lovely, shiny and tiny*" (Boroditsky, Schmidt & Phillips, 2003). Thus, gender attributes have knock-on effects onto other features: e.g., size, hedonics, texture, visual appearance (perhaps through connotative meaning; cf. Osgood, Suci, & Tannenbaum, 1957).

Previous research on the cognitive consequences of speaking a classifier language have been mixed, with the suggestion that effects are modest. In three experiments we failed to find language-specific heightened sensitivity of object similarity for speakers of a classifier language compared to speakers of a non-classifier language, using pictures and words. We, therefore, conclude that classifier systems do not affect overall conceptual representations of objects. Previous evidence could be the result of strategic or explicit use of classifier information that does not reflect typical categorization processes. Alternatively, classifiers may affect conceptualization of specific object features, which cannot be observed with overall similarity judgments (e.g., Lucy & Gaskins, 2003).

Both speakers of a classifier language and a non-classifier language judged objects sharing a classifier as more similar than objects that did not share a classifier. Thus, classifier systems do not impact overall object similarity, but instead likely reflect the way the world is organized.

Acknowledgements

References

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

Bender, A., Beller, S., & Klauer, K. C. (2011). Grammatical gender in German: A case for linguistic relativity? *The Quarterly Journal of Experimental Psychology*, *64*(9), 1821–1835.

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.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English Speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2010). Do English and Mandarin speakers think about time differently? *Cognition, 118*(2), 123–129.

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

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

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

Clark, E. V. (1976). Universal categories: On the semantics of classifiers and children's early word Meanings. In A., Juilland (Ed.) *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday. Vol* 3, (pp. 449-62) Saratog, CA: Anma Libri.

Davidoff, J., Davies, I. & Roberson, D. (1999) Colour categories of a stone-age tribe. *Nature, 398*, 203-204.

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.

Gilbert, A. L., Regier, T., Kay. P. & Ivry R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, 103*(2), 489–494.

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.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 843–858.

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

Levinson, Stephen C. (2003). *Space in language and cognition*. Cambridge University Press: Cambridge.

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

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108-114.

Osgood, C.E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In *Proceedings of the 25th annual meeting of the Cognitive Science Society*. Lawrence Erlbaum Associates Mahwah, NJ.

Saalbach, H., & Imai, M. (2005). Do classifier categories structure our concepts? In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society.* Lawrence Erlbaum Associates Mahwah, NJ.*y*

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

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.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

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.