Trouble finding the words:
Lexical differences affect how English and Chinese speakers communicate categories

Lilia Rissman, Qiawen Liu and Gary Lupyan

Background. Languages vary substantially in how they lexicalize the same concepts — for example, some languages have distinct lexical items for "niece" and "nephew" but others do not (Wallace & Atkins, 1960). We investigate the impact of such cross-linguistic differences on communication — whether having a conventional term for a concept facilitates communication of that concept. We focus on superordinates such as *beverages* and *vehicles* — nouns that convey broad categories of individuals. Languages have different inventories of superordinates (Kemmerer, 2019; Mihatsch, 2007). For example, ||Gana divides living things not into categories of *plant* and *animal* but into categories such as *kx'ooxo* ('living things which are edible') (Harrison, 2007). We ask whether the availability of a superordinate term leads speakers to communicate more effectively about that category than if a superordinate term were absent.

As humans, the range of ideas we want to convey is far wider than the discrete set of morphemes present in any one language. The expressive capacity of language comes largely through combinatoriality – composing morphemes into larger units and phrases that convey complex thoughts. Given this expressive capacity, it might be that communication is not hampered by the absence of a superordinate term. For example, although English has no superordinate that is translationally equivalent to ||Gana *kx'ooxo,* we can convey its meaning through the complex, ad hoc category description "living things which are edible" (see Barsalou, 1983 on ad hoc categories). From this perspective, languages are fully intertranslatable with each other — as articulated by Harnad (1996), "whether it does so analytically, synthetically, or even entirely holophrastically, a language must provide the resources for marking distinctly all the categories we distinguish." An alternative to this perspective is that an ad hoc description is more limited than a superordinate in its ability to convey a category, because the ad hoc description only approximates the meaning of the superordinate, or because speakers differ in their ability to construct ad hoc descriptions on the fly, or because speakers and receivers interpret ad hoc descriptions in different ways. We tested these two alternatives by comparing how English and Chinese speakers communicate about categories for which there is a superordinate in one language but not the other. These languages have been shown to lexicalize semantic space in strikingly different ways (e.g., Saji et al., 2011). At the same time, as members of industrialized societies in increasing contact, speakers of these languages share a range of category knowledge about artifacts, foods, and the natural world.

Method. Participants completed a referential matching game (a 'Director-Matcher' task). 77 American English speakers and 80 Chinese speakers played the Director role. Directors viewed a 3 x 3 grid with a noun in each cell. Three of these nouns (e.g., *beer, soda, juice*) were highlighted. Directors were instructed to write a clue that would enable another person to choose the highlighted and only the highlighted words from the same grid. 86 American English speakers and 124 Chinese speakers played the Matcher role. Matchers viewed the Director's clue along with the 3 x 3 grid (without highlighting) and selected the nouns corresponding to the clue. Matcher's and Director's grids had the same words, but in a different order.

We selected the words in the grid based on 10 English and 10 Chinese superordinate terms shown in Table 1. For each term, there is no direct translational equivalent in the other language. Each trial corresponded to a superordinate term and the nouns in the 3 x 3 grid for each trial were of the following types: 1) three Targets, which were typical members of the category denoted by the superordinate (e.g., *beer, soda, juice* for the term *beverages*), 2) two Lure Distractors, which were semantically similar to the Targets but were not members of the superordinate category (e.g., *vinegar, oil*), and 3) four Non-Lure Distractors, which were

semantically dissimilar to the Targets and Lures (e.g., *motorcycle, star, tree, sleet*). We constructed six grids for each superordinate term. The grids based on the English and the

Table 1. English and Chinese superordinates

English terms	Chinese terms (English gloss)
appetizers	nóngchǎnpǐn (agricultural products and livestock)
beverages	huàzhuāngpǐn (cosmetics and facial products)
crafts	dìxíng (terrain and water features)
crimes	jiājù (furniture and home décor)
drugs	fúshì (apparel, shoes, and jewelry)
pests	shuǐyù (bodies of water)
precipitation	shēngwù (living things)
skills	diànqì (electrical appliances and devices)
snacks	tiáowèi pǐn (food seasonings)
vehicles	fēngjǐng (scenic places to visit)

Chinese superordinates were translated into Chinese and English, respectively. We selected nouns that would be familiar to both English and Chinese speakers (e.g., *beer*; *píjiǔ* 'beer'). Directors and Matchers saw two grids per term, resulting in 40 trials per participant. Each Matcher was yoked to a single Director and saw all clues produced by that Director. All studies were conducted online.
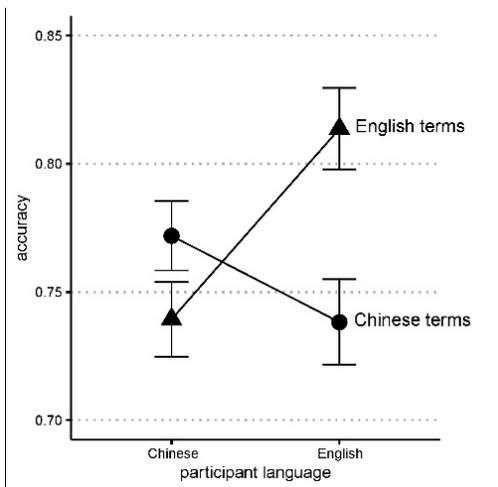

Figure 1. Mean Matcher accuracy across Language and Condition. Error bars show 95% confidence intervals.

Results and Discussion. We calculated Matcher accuracy, defined as the Hit rate per trial (correctly choosing a Target) minus the False Alarm rate per trial (incorrectly choosing a Distractor). Mean accuracy for each condition is shown in Figure 1. We modeled accuracy using linear mixed-effects regression and the *lme4* package for R (Bates, Maechler, Bolker, & Walker, 2014). Our model included Subject and Term random intercepts and Subject-by-Condition random slopes. With reaction time as a covariate, the main effects of Language and Condition were non-significant (b = -.05, SE = .06, p < .1; b = -.11, SE = .11, p > .1). That is, English speakers were not more accurate than Chinese speakers (or vice versa) and categories derived from English terms were not more difficult than categories derived from Chinese terms (or vice versa). We did, however, observe a significant interaction between Language and Condition: English speakers were more accurate for categories derived from English terms (b = .36, SE = .041, p < .001). This demonstrates that English speakers were more effective at conveying categories when there was an English superordinate term available, *mutatis mutandis* for Chinese speakers. These results dispute the view that languages are all mutually intertranslatable. Instead, when language users have to create ad hoc, non-conventional category descriptions, the descriptions they create appear to be less effective than conventional superordinates for conveying the same categories.

Barsalou (1983). Ad hoc categories. *Memory & cognition* | Bates et al. (2014). lme4: Linear mixed-effects models using S4 classes. | Harnad (1996). The origin of words: A psychophysical hypothesis. *Communicating meaning: evolution and development of language* | Harrison (2007). *When languages die: the extinction of the world's languages and the erosion of human knowledge*. | Kemmerer (2019). *Concepts in the brain: The View from cross-linguistic diversity*. | Mihatsch (2007). Taxonomic and meronomic superordinates with nominal coding. *Ontolinguistics: How Ontological Status Shapes the Linguistic Coding of Concepts.* | Saji et al.

(2011). Word learning does not end at fast-mapping… *Cognition*. | Wallace & Atkins (1960). The meaning of kinship terms. *American Anthropologist*.