# Culture, communicative need, and the efficiency of semantic categories

**Shan Gao (shangaocog@gmail.com)**
Ningbo, Zhejiang, China

**Terry Regier (terry.regier@berkeley.edu)**
Department of Linguistics, Cognitive Science Program, UC Berkeley
Berkeley, CA 94720 USA

## Abstract

It has been proposed that a drive for efficient communication shapes systems of semantic categories across languages. Recent work in this vein has increasingly emphasized communicative need: how often a particular object or idea will need to be referenced. Many studies assume for simplicity that the distribution of need across referents is the same for different cultures, and that this need distribution can be reliably inferred from corpora. In contrast, we elicited culture-specific estimates of communicative need from native speakers of English and Chinese. We compared those need distributions to each other and to a corpus-based need distribution, and we assessed the efficiency of the English and Chinese naming systems for the semantic domain of household containers under different need distributions. Our results suggest that languages reflect culture-specific need patterns, and that subjective estimates are sometimes superior to corpus data as a measure of need.

**Keywords:** efficient communication; communicative need; language and culture; semantic variation; semantic universals

## Introduction

An influential proposal holds that language is shaped by pressure for efficient communication. On this view, languages take the forms they do in part because of the functional need to be simultaneously informative and simple. This is an idea with a long history (e.g. von der Gabelentz, 1901, cited by Haspelmath, 1999; Zipf, 1949; Hopper & Traugott, 2003), and a growing body of recent supporting evidence (see Gibson et al., 2019 for a review). One aspect of language that has been analyzed in these terms is semantic categories, such as word meanings. Specifically, it has been proposed that the typology of attested semantic category systems across languages reflects a variety of ways of efficiently trading off the competing forces of informativeness and simplicity against each other. Support for this view has been found in a number of semantic domains, including color, kinship, numeral systems, and others; Kemp et al. (2018) provide a review.

Recent literature on this topic has increasingly emphasized *communicative need*: how often a particular object or idea will need to be referenced (e.g. Gibson et al., 2017; Zaslavsky, Kemp, et al., 2019; Hawkins, 2019; Conway et al., 2020; Karjus et al., 2021; Twomey et al., 2021). Communicative need is a critical construct in theories of efficient communication because it modulates how important it is to be precise in referring to a given object. A drive for informativeness will reward narrow, precise semantic categories — but a drive for simplicity will seek to avoid a proliferation of such categories. An efficient compromise is to have narrow, precise categories in high-frequency (high-need) parts of semantic space, and broader, imprecise categories in low-frequency (low-need) parts of space. This way, the category system is kept relatively simple, while the expected accuracy, aggregated over the space as a whole, is kept reasonably high.

Patterns of communicative need may show commonalities across cultures, corresponding to universal tendencies in human nature. In line with this, some studies of efficiency in semantic categories have assumed a universal need distribution, based either on corpus counts from high-resource languages combined with naming data (e.g. Kemp & Regier, 2012; Xu et al., 2016; Zaslavsky, Kemp, et al., 2019; Xu et al., 2020), naming data alone (Zaslavsky et al., 2018; Zaslavsky, Regier, et al., 2019), or the statistics of the environment (e.g. Gibson et al., 2017). However there is also reason to expect cross-cultural variation in need, since cultural emphases differ. The efficiency view predicts that when need varies across cultures, semantic categories should vary accordingly. This is a classic idea (Boas, 1911) that is supported by recent evidence (e.g. Regier et al., 2016; Gibson et al., 2017; Floyd et al., 2018; Winter et al., 2018; Twomey et al., 2021).

We wished to ascertain how broadly culture-specific need shapes semantic categories, and the reliability of corpus counts as an estimate of communicative need. To that end, we considered the domain of household containers, for which previous research has assumed a single need distribution across languages, inferred either from corpus counts (Xu et al., 2016) or from naming data (Zaslavsky, Regier, et al., 2019). In contrast, we elicited estimates of communicative need for this domain on a per-culture basis from native speakers of two languages. In what follows, we first describe earlier studies that provide a backdrop for ours, and then describe our own study. Our results suggest that subjective estimates may sometimes be superior to corpus data as a measure of communicative need, that communicative need varies across languages, and that variation in need can help to explain variation in systems of semantic categories across languages.

## Prior studies

Malt et al. (1999) presented images of simple household containers, such as those in Fig. 1, to native speakers of English, Chinese, and Spanish. They asked their participants to sort these objects into piles by similarity, and to name them. They



In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.* 

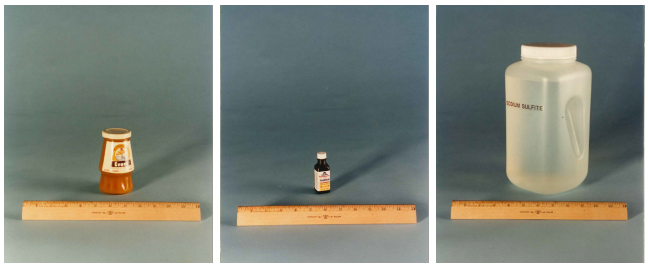

Figure 1: Sample stimuli from Malt et al. (1999). These three containers are named 'jar', 'bottle', and 'jug' (from left to right) in English, whereas all are named 'píng' in Chinese.

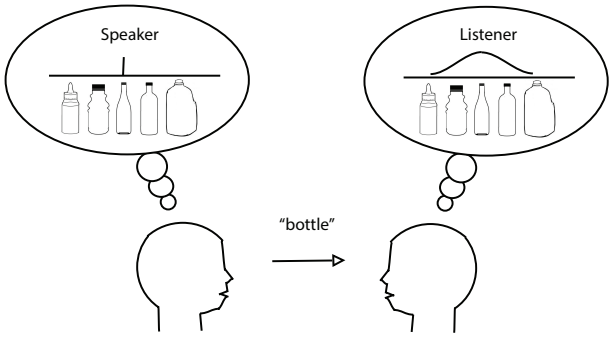


Figure 2: A communicative scenario, from Xu et al. (2016).

found that the naming systems of these three languages for this domain were quite different, but that the similarity structure of the domain was very similar across these languages. Xu et al. (2016) built on these findings, and assessed the efficiency of the English, Chinese, and Spanish naming systems. We lay out their approach in some detail here because our study adopts large parts of it.

In assessing the efficiency of a language's naming system, Xu et al. (2016) assumed a simple communicative scenario, illustrated in Fig. 2. Here, a speaker has a particular object in mind and attempts to communicate it to a listener using the English word 'bottle'. The listener then attempts to mentally reconstruct which referent they think the speaker must have meant — but because the word 'bottle' is semantically broad, they cannot do so exactly. Instead, the listener's mental reconstruction takes the form of a probability distribution over a range of referents all named 'bottle'. This means that some information has been lost in communication, and Xu et al. formalized that information loss as follows.[1] First, they assumed that the probability mass $L(i|w)$ that the listener allocates to object $i$ upon hearing word $w$ is based on the similarity of object $i$ to all objects $j$ that are named by $w$:

$$L(i|w) \propto \sum_{j \in w} sim(i,j) \tag{1}$$

where $sim(i,j)$ are similarities empirically grounded in the sorting data of Malt et al. (1999), pooled across languages. They assumed that each object has a unique name (cf. Zaslavsky et al., 2018), and they modeled the cost of communicating about an object $i$ using its name $w$ by surprisal:

$$C(i) = \log_2 \frac{1}{L(i|w)} \tag{2}$$

Finally, they modeled the overall communicative cost of a semantic system as the expected cost of communicating about objects in the domain:

$$E[C] = \sum_i C(i)N(i) \tag{3}$$

where $i$ ranges over objects in the domain, and $N(i)$ is the *need probability*, or communicative need, for object $i$. Xu et al. estimated need probabilities $N(i)$ using the Google Ngram American English corpus (Michel et al., 2011) for the year 1999, the year of publication of Malt et al. (1999), using a method described below. They then took a semantic system to be informative to the extent that it had low communicative cost $E[C]$; they took the complexity of a system to be the number of lexical categories it contained; and they took a system to be efficient if it was more informative (had lower $E[C]$) than most logically possible hypothetical systems of the same complexity. They created hypothetical systems of the same complexity as a given target naming system through a random process of probabilistic chaining described on pp. 2090-2091 of their paper, by which a name is extended from one exemplar to similar exemplars. They argued that such chained hypothetical systems provide a conservative comparison case because such systems respect the similarity structure of the domain, unlike many other logically possible systems.[2]

Xu et al. used this formalization and the Malt et al. data to assess the efficiency of the naming systems of English, Chinese, and Spanish. Fig. 3 shows our replication of their results for English and Chinese, using the same data and methods they used. Our results qualitatively match theirs.[3] It can be seen that on this analysis, assuming a single need distribution based on an English-language corpus, English is more clearly efficient than Chinese: the English system shows lower communicative cost than almost all hypothetical systems of its complexity, whereas this is not as true of Chinese. We focus on Chinese and English, and not Spanish, because Chinese and English are the native languages of the authors of this paper, and the first author noticed that the Malt et al. stimulus set intuitively seemed Western in emphasis. This observation suggested that a culture-specific need distribution may be appropriate — and it also presented an opportunity to compare corpus-based and subjective need estimates.

[1] Our formal notation differs slightly from theirs.

[2] A complementary approach is to determine optimally efficient systems, and compare empirical systems to them (e.g. Zaslavsky et al., 2018; Zaslavsky, Regier, et al., 2019). We followed Xu et al.'s approach to maximize comparability with their findings.

[3] Quantitatively, we obtain the same communicative costs for the two attested systems as Xu et al. did, but somewhat different distributions for the hypothetical systems. Xu et al.'s description of their chaining algorithm leaves some room for interpretation, and although we tried to follow them exactly, it is possible there are some minor differences; we attribute the difference in outcome to that.

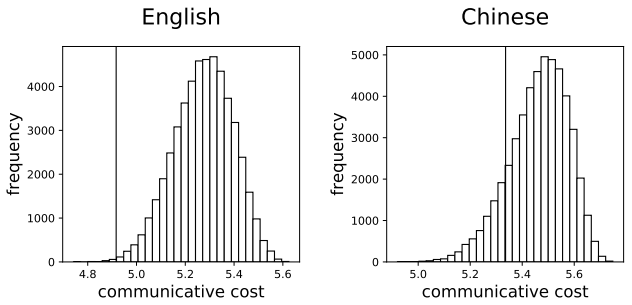


Figure 3: Replication of the findings of Xu et al. 2016 for English (left) and Chinese (right), using need probabilities derived from an English-language corpus. In each panel, the vertical line denotes the communicative cost of the attested system, and the histogram denotes communicative costs of 50,000 hypothetical systems of the same complexity as that attested system. On this analysis, English ($p < .005$) is more clearly efficient than is Chinese ($p = 0.14$).

## Methods

To test these ideas, we collected subjective estimates of communicative need from native speakers of English and of Chinese, for the objects in the Malt et al. stimulus set. We also asked the same participants to name each of the objects in their native language.

**Materials.** We used the images of 60 household containers that were originally used by Malt et al. (1999) and later used by Xu et al. (2016). Sample stimuli are shown above in Fig. 1.

**Participants.** We recruited 25 participants from the US, and 25 from mainland China. Participants were recruited through online platforms (US: Prolific; China: WeChat) and were compensated for their participation. We included prescreening criteria to exclude potential participants who were not adult (18 years old) native English speakers born in the US (for US participants), or not adult (above 18 years old) native Mandarin Chinese speakers born in mainland China (for Chinese participants). We also used two further exclusion criteria. (1) To double-check participants' language background, we included a picture free description task and excluded participants whose responses to this task indicated that they were not native speakers of English (for US participants) or Mandarin Chinese (for Chinese participants), as judged by the authors who are native speakers of these two languages. (2) We also excluded participants who, when asked to name the container objects, provided responses that were not container names for more than 10% of the containers. After exclusions, 16 US and 23 Chinese participants remained.

**Procedure.** Participants completed an online survey, administered in English for US participants and in Mandarin Chinese for mainland China participants. The survey began with a short demographic questionnaire, in which participants were asked about their birth country/region, intercultural experience, gender, and age; some of this information was used for participant exclusion as described above. Next, participants were shown a perceptually rich picture of an indoor garden and asked to describe it in detail using a minimum of 4-5 full sentences; this was also used for participant exclusion as described above. Participants then began the main part of the online survey, in which we collected subjective estimates of communicative need, and object names. To control for order effects, the order of the 60 container stimuli was randomized. For each stimulus picture, participants were asked two questions, which we present here in English but which were presented to the participants in their native language (English or Chinese). First, to elicit subjective estimates of communicative need, we asked: "In your everyday life, how often do you refer to the container shown in the picture above? Please focus on the container itself, not what it contains or what its label says; the ruler in the picture is there to simply show you the size of the container. Rate the frequency that you refer to the container in the picture on a scale from 1 (I don't refer to this container at all) to 7 (I refer to this container every day)." Second, to elicit object names, we asked: "What do you call the container in the picture above? Please name the container itself, not what it contains. Give whatever name that seems the best or most natural to you; it can be one word or more than one word." We coded naming responses by determining the head noun of each response; e.g., if a participant named a particular object a "small plastic bottle," we would code that response as "bottle." Then, for each object, we identified the most commonly used (or modal) head noun for that object among participants in each language and used that noun as the name for that object in that language. We encountered one instance in which there was a tie in frequency between two labels for an object. We conducted all analyses below with both ways of labeling that object and found that our results are qualitatively robust to which way that tie is broken, and so we present results based on one of those two labelings.

**Other data.** We also used the pile-sort data of Malt et al. (1999). Our treatment and use of the pile-sort data followed that of Xu et al.: we considered pile-sort data from English and Chinese, the two languages for which such data were retrievable, and aggregated pile-sort responses across the two languages, motivated by Malt et al.'s finding that these responses were very similar across languages. We then took the similarity $sim(i, j)$ of two objects $i$ and $j$ (see Eqn. 1 above) to be the proportion of participants who placed those two objects in the same pile. We also used the Google Ngram American English corpus (Michel et al., 2011) for the year 1999 to estimate need as Xu et al. did. Table S1 of the Supplementary Material for Xu et al. (2016) provides, for each of the 60 objects in the stimulus set, an English phrase describing it, e.g. "tupperware container", "baby bottle", "applesauce jar". We obtained English corpus-based need probabilities for the 60 objects in the stimulus set by finding the corpus frequency for the phrase associated with each object, assigning that frequency to that object, and then normalizing across objects.[4]

[4]This yielded results that aligned closely with those reported by Xu et al., and so we assume this is the approach they used, and

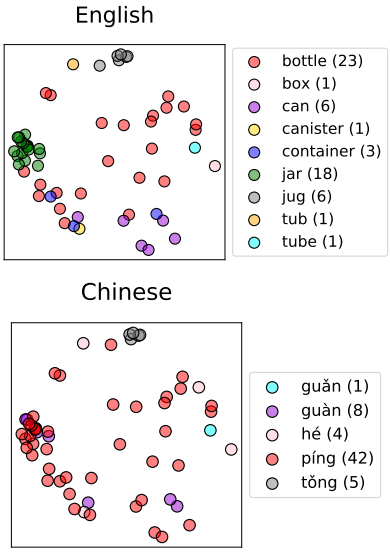


Figure 4: Container naming systems in English (top) and Chinese (bottom), from our study. Objects are represented by small circles, arranged in a single similarity space obtained via MDS from pile-sort data. The proximity between two circles roughly corresponds to the perceived similarity of those two objects. Colors denote linguistic labels. In each legend, labels are shown together with the number (in parentheses) of objects for which that label was the modal head noun.

In the following analyses, we first explore the naming data, then the communicative need data, and then conduct efficiency analyses that bring together these two sorts of data.

## Analysis 1: Naming

The naming systems we obtained for English and Chinese are presented in Fig. 4, in a similarity space derived from the pile-sort data.[5] It can be seen that the Chinese naming system over these objects is dominated by a single broad category, 'píng', whereas the English system is more fine-grained (recall Fig. 1). It can also be seen that Chinese has fewer categories in this domain than English does — thus, the Chinese naming system for this domain has lower complexity than the English one. That difference in complexity cannot explain the apparent weaker efficiency of Chinese relative to English, highlighted above in our replication of Xu et al.'s analyses. That is because we, following Xu et al., assessed

we adopt it here. We also explored another approach, in which we first divided the ngram corpus frequency allocated to each object by the number of objects that received that name, and then normalized. Our results were qualitatively the same under the two approaches. Finally, we also explored the use of corpus counts from the year 2019 rather than 1999, and explored corpus counts based on both singular and plural forms of the target phrases; these variants yielded only small changes in the resulting need distribution and so we do not consider them further here.

[5]These naming systems are broadly similar to those obtained by Malt et al. (1999), with some differences. We defer for future work a detailed comparison between Malt et al.'s naming data and ours.

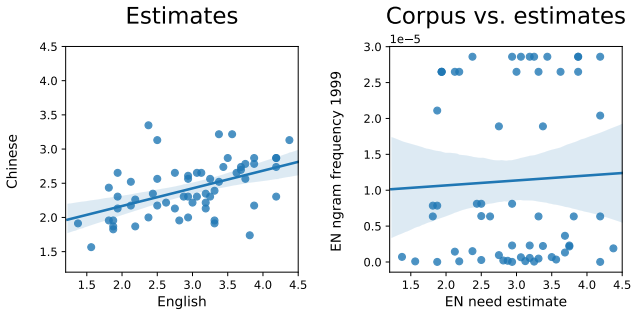


Figure 5: Communicative need. In each scatter plot, points correspond to objects in the stimulus set, plotted according to two need-relevant quantities. The linear regression fit for these points is shown as a blue line, with the 95% confidence interval as a light blue region. Left: Subjective need estimates from English vs. Chinese. Right: English subjective need estimates vs. English corpus-based frequency.

the efficiency of a system while holding complexity constant: by comparing the communicative cost of a given naming system to the costs of hypothetical systems of the same complexity as the original. However the lower complexity of Chinese does have a different important theoretical connection, to the question of communicative need. If category systems are shaped by a drive for efficiency, we would expect to find low-complexity systems with broad categories in domains that are referenced only infrequently in a given language (Kuschel & Monberg, 1974; Kemp et al., 2018, sec. 4.2). The reason is that broad categories lead to information loss in communication when their names are used without modifiers that narrow the semantic range; but if such broad, lossy categories were to appear in low-frequency parts of semantic space, that cost would be incurred only rarely. This reasoning leads us to expect that Chinese may have lower need generally for (this part of) the domain of household containers than English, in line with the author intuition mentioned above. We test this prediction below by examining subjective need ratings collected from speakers of the two languages.

## Analysis 2: Need

For each object in the stimulus set, and for each language (Chinese, English), we estimated the subjective need for that object in that language by taking the average subjective need elicited for that object across speakers of the language. We then compared the resulting English and Chinese subjective need estimates; these two quantities are shown in Fig. 5 (left panel). There are several points to highlight here. First, there is a relatively clear linear relation between need estimates from the two languages ($R^2 = 0.24, F(1,58) = 18.17, p < 0.001$), which can be thought of as capturing cross-culturally shared aspects of communicative need (see also Zaslavsky et al., 2021). Second, there is also substantial scatter around the line, which can be thought of as capturing culture-specific aspects of need for particular objects. Third, need estimates

for Chinese in this domain tend to be lower than those for English (paired $t(59) = 6.33, p < .001$), consistent with the prediction made above based on greater semantic breadth in Chinese, and with prior author intuition mentioned above.

We next asked whether the subjective need estimates we obtained in our study are similar to corpus-based need. Xu et al. found that the Google Ngram corpus for Chinese was too sparse to be useful, so we have corpus-based need only for English, and we compared that to our English subjective need estimates. Fig. 5 (right panel) shows that there is not a clear linear relation between English subjective need estimates and English corpus-based need for year 1999 ($R^2 = 0.002, F(1,58) = 0.12, p = 0.74$). We conclude from this that at most one or the other of these two quantities can be a reasonable measure of communicative need in English — not both since they appear to be incompatible. This in turn means that if we find a reason to trust the subjective estimates over corpus-based need, Xu et al.'s earlier efficiency findings could be called into question, since they assumed corpus-based need.

Are there grounds for trusting subjective need estimates over corpus-based need in this case? There are several indications that seem to point in that direction. We have already seen that the cross-language comparison of subjective need shows a plausible mixture of shared and language-specific patterns, and that it helps to explain the broader semantic categories found in Chinese. Further insight can be gained by considering need estimates for specific objects. A jar of belt cleaner, which appears unusual to both authors (native speakers of Mandarin Chinese and U.S. English), received very low need estimates from speakers of both Chinese and English. There is only one item in the stimulus set that is clearly Chinese in origin — a large plastic bottle of soy sauce with Chinese writing on the label — and it was one of relatively few objects to receive a higher need estimate from Chinese speakers than from English speakers. Conversely, a water bottle with a plastic straw was rated high-need by English speakers and low-need by Chinese speakers, again in line with author intuition. Considerations of space rule out a fuller exploration along these lines, but there seems to be at least some alignment of the subjective need estimates with intuition. This contrasts with the (English) corpus-based need frequencies, for which the top five highest-need items were five different glass jars, which all received the same high rating because all were labeled "glass jar" for the purpose of the corpus search. This highlights a limitation of corpus searches: the search terms may be too semantically broad, in which case distinctions are not made among stimulus items that should be distinguished, and at the same time frequency is likely to be picked up from other items outside the stimulus set. On balance, the evidence reviewed here favors subjective need estimates over corpus-based need, at least in this case.

However, subjective need estimates should not necessarily be taken at face value (e.g. Alderson, 2007). Need estimates are estimates of frequency, and it is known that for many naturally occurring frequency distributions, frequency follows Zipf's law (Zipf, 1949; see Piantadosi, 2014 for a review), according to which $f(i) \propto 1/r(i)$, where $f(i)$ is the frequency of item $i$, and $r(i)$ is the frequency rank of item $i$. Inspection of the subjective need estimates we collected revealed that they drop off with rank much more gradually than this. For that reason, we considered the possibility that the subjective estimates may understate the variation in actual need but that the ranks of those estimates might be accurate, and that actual need frequency could be inferred from those ranks using Zipf's law. We applied this transformation to the need estimates, and we call the resulting estimates *transformed* need estimates, and the raw estimates shown in Fig. 5 *non-transformed* need estimates. For each language, English and Chinese, we produced two need probability distributions — transformed and non-transformed — by normalizing the corresponding need estimates. Other approaches to transforming such need estimates are also possible (Stevens, 1966; Shapiro, 1969), but we leave exploration of those possibilities for future work.

## Analysis 3: Efficiency

As we have seen, Xu et al.'s efficiency analyses, which assumed corpus-based need derived from an English language corpus, suggested that the English and Chinese naming systems in this domain are both efficient, but Chinese more weakly efficient than English. However we have also seen reason to doubt whether a need distribution based on an English language corpus is an appropriate choice. Here, we reassess the efficiency of the English and Chinese naming systems, now using need probability distributions derived from subjective estimates of need. In each analysis, the communicative cost of a given naming system was compared to the distribution of costs obtained from 50,000 hypothetical systems of the same complexity as that naming system, obtained using the probabilistic chaining algorithm described on pp. 2090-2091 of Xu et al. (2016). For reasons of space we only present results obtained using transformed need; results using non-transformed need were qualitatively similar.

Fig. 6 shows the results of these analyses. The top row shows outcomes obtained using native-language need: that is, the English naming system was assessed using a need distribution based on English subjective need estimates, and the Chinese naming system was assessed using a need distribution based on Chinese subjective need estimates. It can be seen that under native-language need, both languages are clearly efficient, in contrast with the findings of Xu et al. For comparison, the bottom row shows outcomes obtained when each language is analyzed using the other language's need distribution. Here, we see that Chinese appears only weakly efficient when assessed using English need estimates, mirroring the findings of Xu et al. for those circumstances. Unexpectedly, English appears clearly efficient under a Chinese need distribution, even a little more so than under an English one, an outcome for which we do not (yet) have a ready ex-

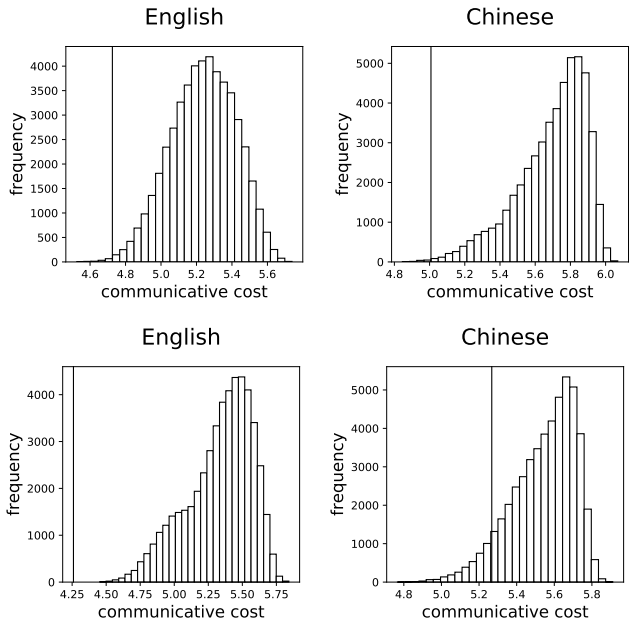


Figure 6: Efficiency analyses under subjective need estimates, transformed. Top row: Native-language need distribution (English: $p < .005$; Chinese: $p < .005$). Bottom row: Other language's need distribution (English: $p < .001$; Chinese: $p = .07$).

planation. Still, despite this unanticipated aspect of our results, we do find that each language is clearly efficient when assessed using a need distribution based on subjective need estimates from speakers of that language.

## Discussion

We have seen that corpus-based estimates of communicative need are not always reliable, and that native-speaker subjective judgments of communicative need can provide a more reliable, fine-grained, and culturally specific measure. We have also seen, using such culturally specific subjective need estimates, that the category system of a language is clearly efficient when assessed using a native-language need distribution. These findings underscore the importance of considering cultural variation in communicative need, and the usefulness of directly eliciting need estimates from native speakers.

Corpus statistics may be more useful for estimating communicative need in some semantic domains than in others. In a domain like number, many languages have a distinct name for each integer, and so it is a simple matter to search for those names and to assess need for the corresponding integers on that basis (Xu et al., 2020). However, in a domain like household containers, there is often no fixed conventional name for individual items in the domain, for which one may search. Instead, there is a tension between selecting multi-word search phrases that are highly specific and therefore likely to encounter data sparsity problems, vs. selecting search terms that are somewhat more general (e.g. "glass jar") but that may correspond to multiple items in the domain, providing only coarse-grained information about need. This problem has been addressed by using general search terms, obtaining corpus frequencies for them, and then inferring need for individual items in the domain from those corpus statistics and naming data in a principled way (see e.g. Zaslavsky, Kemp, et al., 2019). However even this approach has a limitation that is especially relevant in the domain of household containers. The corpus frequencies for many natural search terms — such as "glass jar" — will reflect references not only to items in a given stimulus set, but also to other items well outside it, meaning that they overestimate communicative need for the items in the set. This is arguably not as serious a problem for a domain like color (Zaslavsky, Kemp, et al., 2019), in which the standard stimulus set is a representative and systematic sample of the domain, but that is not the case for the domain we have considered here. In general, estimating communicative need is a challenging and domain-dependent task, and one of our aims has been to add an additional means of approaching it: subjective need estimates, which allow a fine-grained and culture-specific measurement of need.

This study is an initial exploration of these ideas, and it leaves open a number of important questions. To what extent do our findings generalize to other languages, cultures, and semantic domains? Should subjective need estimates be transformed, and if so what is the most appropriate transform? Might need estimates collected at different points in time from a single language help to explain patterns of language change over time? We leave these questions for future research. For now, our present study has both practical and theoretical implications. Practically, subjective need estimates can in principle be collected not just for high-resource languages like English and Chinese, but also from speakers of under-documented and low-resource languages. Assessments of communicative need could become a standard part of linguistic fieldwork in such languages, such that assessments of need for such languages could proceed prior to the completion of large-scale corpora. Our findings suggest that such need assessments may be possible, and would be useful. Theoretically, this study and others like it touch on an important topic that has not always been treated with the seriousness it deserves. Boas (1911) claimed that the semantic categories in the lexicon of a language "must to a certain extent depend upon the chief interests of a people" (p. 22). This claim was popularized by Whorf (1956) specifically in connection with words for snow, which led to distortions and exaggerations, in popular culture, of the original claim — which eventually led to the topic largely being dismissed by scholars (Martin, 1986; Pullum, 1991). We hope that our study, like some others (Krupnik & Müller-Wille, 2010; Regier et al., 2016; Floyd et al., 2018; Winter et al., 2018; Kemp et al., 2019; Twomey et al., 2021), can help to re-normalize this topic, and restore it to the status of a simple, fundamental, and testable claim, as Boas appears to have originally intended.

## Acknowledgments

We thank Yang Xu, Charles Kemp, and Noga Zaslavsky for helpful comments on an earlier version of this paper. Author contributions: SG and TR designed the research; SG performed the research; SG analyzed the data; and SG and TR wrote the paper.

## References

Alderson, J. C. (2007). Judging the frequency of English words. *Applied Linguistics*, *28*(3), 383–409.

Boas, F. (1911). Introduction. In *Handbook of American Indian Languages, Vol.1* (p. 1-83). Government Print Office (Smithsonian Institution, Bureau of American Ethnology, Bulletin 40).

Conway, B. R., Ratnasingam, S., Jara-Ettinger, J., Futrell, R., & Gibson, E. (2020). Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. *Cognition*, *195*, 104086.

Floyd, S., Roque, L. S., & Majid, A. (2018). Smell is coded in grammar and frequent in discourse: Cha'palaa olfactory language in cross-linguistic perspective. *Journal of Linguistic Anthropology*, *28*(2), 175-196.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785-10790.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389-407.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*, *18*, 180-205.

Hawkins, R. D. (2019). *Coordinating on meaning in communication* (Doctoral dissertation). Department of Psychology, Stanford University.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization. second edition*. Cambridge, UK: Cambridge University Press.

Karjus, A., Blythe, R. A., Kirby, S., Wang, T., & Smith, K. (2021). Conceptual similarity and communicative need shape colexification: An experimental study. *Cognitive Science*, *45*(9), e13035.

Kemp, C., Gaby, A., & Regier, T. (2019). Season naming and the local environment. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 539–545).

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*(1).

Krupnik, I., & Müller-Wille, L. (2010). Franz Boas and Inuktitut terminology for ice and snow: From the emergence of the field to the "Great Eskimo Vocabulary Hoax". In I. Krupnik, C. Aporta, S. Gearheard, G. Laidler, & L. K. Holm (Eds.), *Siku: Knowing our ice. Documenting Inuit sea-ice knowledge and use* (pp. 377–400). Springer.

Kuschel, R., & Monberg, T. (1974). 'We don't talk much about colour here': A study of colour semantics on Bellona Island. *Man*, *9*(2), 213–242.

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

Martin, L. (1986). "Eskimo words for snow": A case study in the genesis and decay of an anthropological example. *American Anthropologist*, *88*, 418–423.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, G. B., ... others (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin and Review*, *21*(5), 1112—1130.

Pullum, G. (1991). The great Eskimo vocabulary hoax. In *The great Eskimo vocabulary hoax and other irreverent essays on the study of language* (pp. 159–171). University of Chicago Press. (Chapter 19)

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLoS ONE*, *11*(4), e0151138.

Shapiro, B. J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 248–251.

Stevens, S. S. (1966). A metric for the social consensus. *Science*, *151*(3710), 530–541.

Twomey, C. R., Roberts, G., Brainard, D. H., & Plotkin, J. B. (2021). What we talk about when we talk about colors. *Proceedings of the National Academy of Sciences*, *118*(39).

von der Gabelentz, G. (1901). *Die Sprachwissenschaft, ihre Aufgaben, Methoden, und bisherige Ergebnisse*. Leipzig: Weigel.

Whorf, B. L. (1956). Science and linguistics. In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (pp. 207–219). MIT Press.

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213-220.

Xu, Y., Liu, E., & Regier, T. (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind*, *4*, 57–70.

Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, *40*, 2081–2094.

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2019). Communicative need in color naming. *Cognitive Neuropsychology*.

Zaslavsky, N., Maldonado, M., & Culbertson, J. (2021). Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society.*

Zaslavsky, N., Regier, T., Tishby, N., & Kemp, C. (2019). Semantic categories of artifacts and animals reflect efficient coding. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 1254–1260).

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.