



Cognitive Science (2015) 1–14

Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12312

# Historical Semantic Chaining and Efficient Communication: The Case of Container Names

Yang Xu,<sup>a</sup> Terry Regier,<sup>b</sup> Barbara C. Malt<sup>c</sup>

<sup>a</sup>*Department of Linguistics, University of California, Berkeley*

<sup>b</sup>*Department of Linguistics, Cognitive Science Program, University of California, Berkeley*

<sup>c</sup>*Department of Psychology, Lehigh University*

Received 18 March 2015; received in revised form 12 August 2015; accepted 19 August 2015

---

## Abstract

Semantic categories in the world's languages often reflect a historical process of *chaining*: A name for one referent is extended to a conceptually related referent, and from there on to other referents, producing a chain of exemplars that all bear the same name. The beginning and end points of such a chain might in principle be rather dissimilar. There is also evidence supporting a contrasting picture: Languages tend to support efficient, informative communication, often through semantic categories in which all exemplars are similar. Here, we explore this tension through computational analyses of existing cross-language naming and sorting data from the domain of household containers. We find (a) formal evidence for historical semantic chaining, and (b) evidence that systems of categories in this domain nonetheless support near-optimally efficient communication. Our results demonstrate that semantic chaining is compatible with efficient communication, and they suggest that chaining may be constrained by the functional need for efficient communication.

*Keywords:* Semantic variation; Artifact categories; Semantic chaining; Historical semantics; Semantic universals; Efficient communication

---

## 1. Introduction

Languages vary widely in the ways they partition human experience into categories. For example, some languages use a single color term to cover both green and blue (Berlin & Kay, 1969), and some languages have a spatial term that captures the notion of being located in water (Levinson & Meira, 2003, p. 496), which is not captured by any single basic spatial term in English. Yet at the same time, many logically possible

---

Correspondence should be sent to Yang Xu, Department of Linguistics, University of California, Berkeley, CA 94720-2650. E-mail: yang\_xu\_ch@berkeley.edu

semantic categories are not attested, and similar categories appear in unrelated languages (e.g., Malt & Majid, 2013). What explains this pattern of wide but constrained variation?

An existing proposal holds that this variation may be explained by the functional need for *efficient communication*—that is, the need to communicate precisely, using minimal cognitive resources. On this account, the different semantic category systems that we see across languages constitute different structural means to this same functional end. This idea has accounted for cross-language variation in semantic domains, including color (Regier, Kay, & Khetarpal, 2007), kinship (Kemp & Regier, 2012), space (Khetarpal, Neveu, Majid, Michael, & Regier, 2013), and number (Xu & Regier, 2014). It also coheres naturally with a recent focus on efficient communication as an explanation for other aspects of language (e.g., Fedzechkina, Jaeger, & Newport, 2012; Piantadosi, Tily, & Gibson, 2011; Smith, Tamariz, & Kirby, 2013). Importantly for our present purposes, in several of the above studies of semantic categories (Khetarpal et al., 2013; Regier et al., 2007), efficient communication is shown to be supported by coherent categories in which all exemplars tend to be similar to each other and hence tightly clustered.

This proposal appears to conflict with a well-established and influential idea: that semantic categories reflect a historical process of *chaining*, whereby a name for one referent is extended to a related referent, and from there on to further referents, resulting in a chained structure in which the later items in the chain may have little similarity to the early ones (Brugman, 1988; Heit, 1992; Lakoff, 1987; see also Bybee, Perkins, & Pagliuca, 1994; Hopper & Traugott, 2003). In particular, it has been suggested that such semantic chaining over historical time may explain the extensions of English container names such as *bottle* and *jar*: containers are human-made artifacts that are invented in different forms at different points in time in response to technological innovations and current cultural needs, and the extensions of container names may reflect this historical process (Malt, Sloman, Gennari, Shi, & Wang, 1999).

If the meanings of container names must flex to accommodate rapid changes to the domain itself, the result may be that they are extended along dimensions that are not fully predictable (see review by Malt, 2010). As a result, the words in this domain can denote a wide range of objects that do not necessarily seem to be the most natural perceptual or conceptual grouping. For example, Malt et al. (1999) found that the extensions for container names include exemplars that are dissimilar to exemplars within the category on average, but are very similar to certain individual exemplars, consistent with the idea of chaining. For instance, a child's drink container in the shape of a bear, made of plastic, and emptied via a straw—which does not resemble a prototypical box—may nonetheless be called a *juice box*. Sloman, Malt, and Fridman (2001) found that a computational model that captures chain-like structures accounted well for container naming data from English. These analyses examined the data without reference to historical information, and thus did not directly assess whether the data are consistent with chaining over historical time. But they do appear to challenge the proposal that semantic systems support communicative efficiency through categories in which exemplars tend to be similar to each other.

There is another consideration that suggests that names for containers may pattern differently from those in other domains. In this paper, we consider the communicative efficiency of simple container nouns such as *box* or *bottle*. Yet very often in discourse, such container names appear with a preceding modifier of some sort (e.g., *juice box*, *aspirin bottle*). Given this usage pattern, communicative efficiency need not require that a simple noun such as *box* be highly informative by itself, because the noun often appears with a disambiguating modifier. Thus, it is possible that head nouns for containers, such as *box* or *bottle*, may not support efficient communication, even if the regularly appearing modified forms *juice box* or *aspirin bottle* do.

Two important questions are left open. First, is there evidence for a genuinely historical process of chaining in modern semantic systems such as those of container names? And second, if so, does chaining in this domain in fact prevent efficient communication? Or are the lexical categories of this domain, like those of other domains, shaped by the need for efficient communication, despite semantic chaining? The studies we present address these questions.

In what follows, we summarize the theory of efficient communication, and demonstrate that chaining is in principle a challenge to this theory. We also briefly describe the cross-language data on which we rely. We then present two studies based on those data. The first study tests for historical chaining in the naming of containers, and the second study tests whether container naming across languages is communicatively efficient, in the sense of having tightly clustered, coherent semantic category membership. To preview our results, we find evidence for historical chaining, yet we also find that despite this chaining, the container naming systems of three languages all support near-optimally informative communication. We speculate that semantic chaining may be constrained by the need for categories to be informative.

## 2. Formal presentation of theory

In this section, we present the theory of efficient communication in formal terms. We then demonstrate that chaining can in principle lead to inefficient communication.

Consider the communicative scenario of Fig. 1. Here, the speaker has a target object in mind—in this case, a specific kind of bottle—and wishes to communicate that referent to the listener. To that end, the speaker utters the word *bottle*. Given that utterance, the listener then attempts to *mentally reconstruct* the speaker's intended meaning. Because the word *bottle* covers a range of possible objects, the listener's representation is inexact and is shown as a probability distribution extending over that range. We take a communicative system to be *informative* to the extent that it supports accurate mental reconstruction by the listener of the speaker's intended meaning; that is, reconstruction that is as exact as possible. While the listener need not always reconstruct exactly the referent the speaker has in mind for communication to succeed (Connell & Lynott, 2014; Ferreira, Bailey, & Ferraro, 2002), if the speaker utters

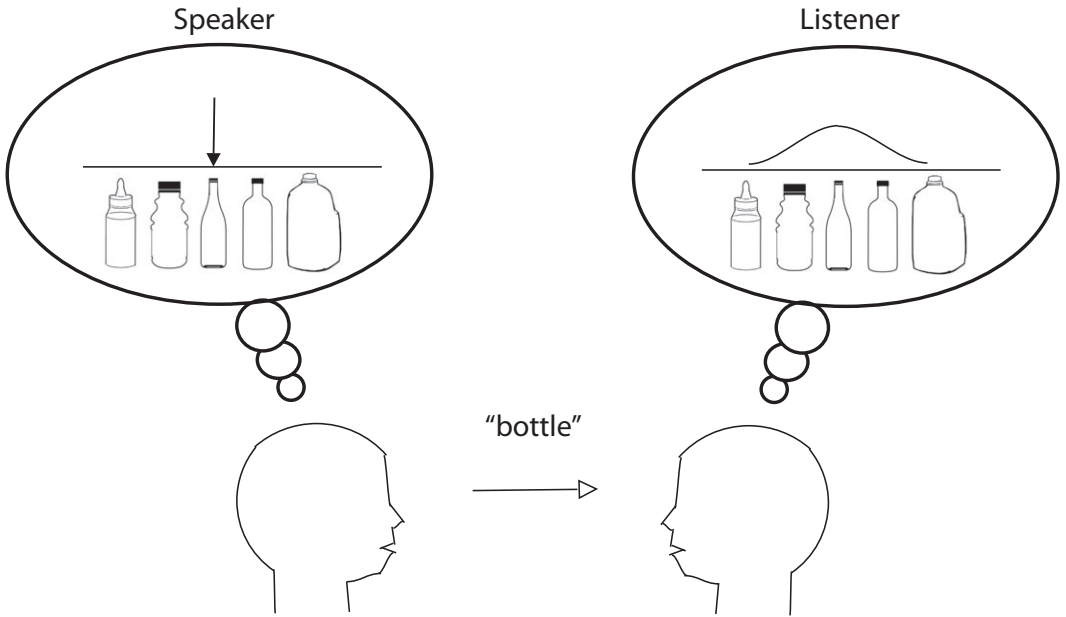


Fig. 1. A simple communicative scenario.

*bottle* having in mind a baby bottle and the listener thinks of a Coke or wine bottle, unintended consequences may result.

We model the mental representations of both speaker and listener as probability distributions. Unlike the listener's distribution, the speaker's distribution  $S$  is certain: It consists of a point mass centered on the target, capturing our assumption that the speaker is certain of the meaning she wishes to convey. Following Regier, Kemp, and Kay's (2015) analysis of color naming, we take the listener distribution  $L(i)$  to be based on the similarity (assessed empirically) of exemplar  $i$  to all exemplars in the category named by the word  $w$ :

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

We then take the unit *communicative cost*  $C(i)$  of communicating about a target object  $i$  using a particular communicative system to be a measure of the discrepancy between the listener distribution  $L$  and the speaker distribution  $S$ : specifically the Kullback–Leibler divergence  $D_{KL}(S||L)$  between these two distributions. In the case of speaker certainty, this reduces to surprisal (Tribus, 1961):

$$C(i) = D_{KL}(S||L) = \sum_j S(j) \log_2 \frac{S(j)}{L(j)} = \log_2 \frac{1}{L(i)} \quad (2)$$

Finally, we take the overall communicative cost of a system to be the expected communicative cost incurred in communicating about the domain (Kemp & Regier, 2012). This is the sum of the unit costs of all possible targets in the domain, each weighted by its relative frequency of occurrence in usage, or need probability  $N(i)$  (assessed empirically):

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

We take a communicative system to be *informative* to the extent that it exhibits low communicative cost  $E[C]$ . We take the *complexity* of a system to be the number of lexical categories in the system. Finally, we take a system of categories to be *communicatively efficient* to the extent that it is more informative than most logically possible hypothetical systems of the same complexity.

### 2.1. Chaining and inefficient communication

Semantic chaining can give rise to inefficient communication, as illustrated in Fig. 2. Panel (a) of this figure shows two artificial category systems that partition the same set of eight objects (shown as black dots). The category system in the left half of the panel divides these objects into two non-chained (or clustered) categories. The system in the right half of the panel divides the same set of objects into two chained categories. The complexity of the two systems is the same (two categories in each system), but they differ in informativeness. Panel (b) shows the communicative costs of these two systems, compared with the costs of all possible partitions of the eight objects into two groups of size 4.<sup>1</sup> It can be seen that relative to these hypothetical systems, the non-chained system from panel (a) is optimally informative for this level of complexity, and thus communicatively efficient, whereas the chained system is not. This demonstrates that semantic chaining has the potential to yield inefficient communication, as formalized here.

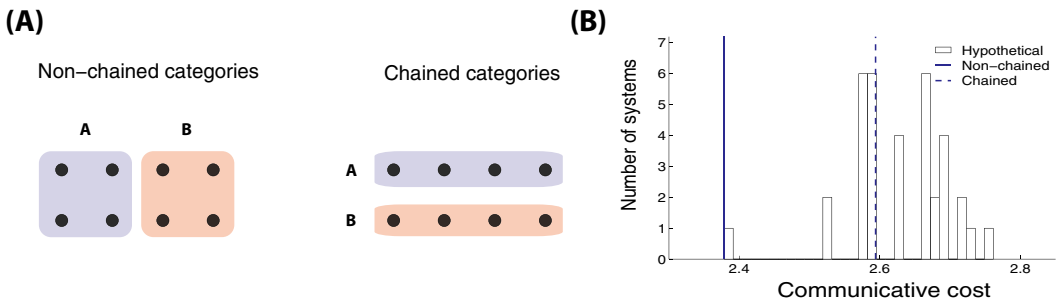


Fig. 2. Chaining and communicative inefficiency. (a) Non-chained and chained systems of equal complexity (two categories each). (b) Communicative costs of these systems and hypothetical systems of equal complexity.

### 3. Data

We reanalyze data from Malt et al. (1999) on the naming and perceived similarity of household containers. The stimulus set for this research consisted of photographs of 60 household containers, representing a wide range of bottles, jars, and similar containers. Fig. 3 shows a sample of these objects.

We used two types of data that had been collected relative to this stimulus set. The first type is linguistic. Native speakers of American English, Mandarin Chinese, and Argentinian Spanish were instructed to name each container stimulus in their native language, giving whatever name they felt was best or most natural.<sup>2</sup> We took the most frequently produced (modal) name for each object as its name in the given language. The second type of data is from a pile-sorting task. We used these data to construct a similarity space of the 60 container objects in our stimulus set. In particular, we used pile-sorting by English and Chinese speakers from the study by Malt et al. (1999) (the groups for which data were retrievable), and we focused on sorting based on overall similarity of the containers (i.e., participants were allowed to sort freely, using both the physical attributes and functions of the containers, as they saw fit). Since the pile sorts correlated highly across groups (Malt et al., 1999), we aggregated pile-sorting responses from participants across the two groups, and took the similarity  $sim(i,j)$  of any two objects  $i$  and  $j$  to be the proportion of all participants who sorted those two objects into the same

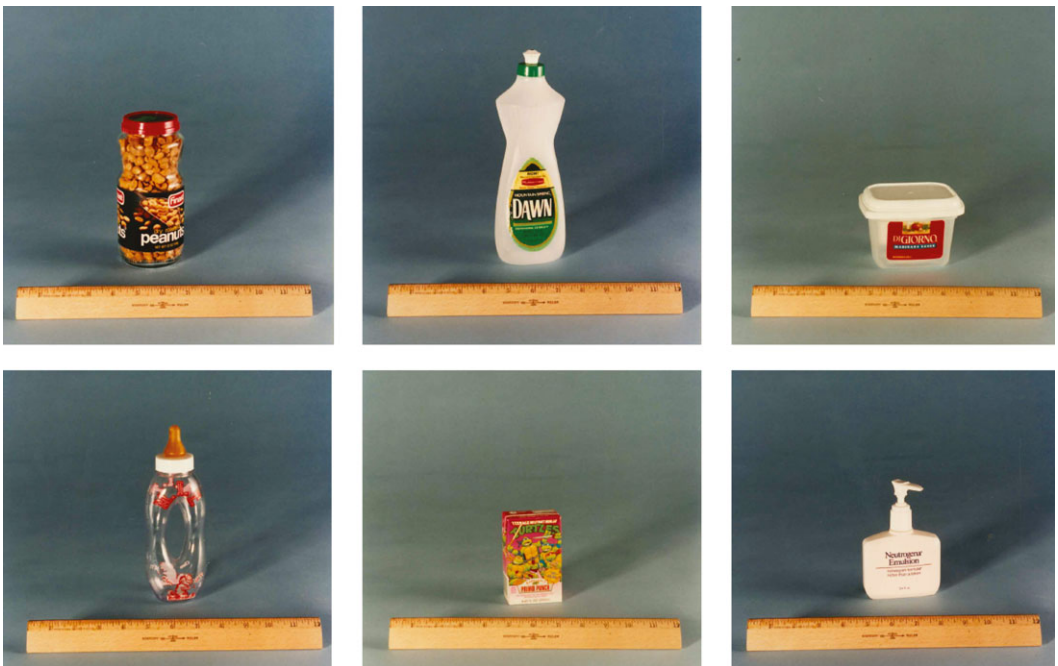


Fig. 3. Sample stimuli from Malt et al. (1999).

pile. If two objects are always co-sorted together across participants, their similarity would be 1. If two objects are never co-sorted, their similarity would be 0. We used this  $60 \times 60$  similarity matrix as a proxy for a universal conceptual space. Malt et al. (1999) showed a high degree of correlation of these similarity matrices between each pairing of the three language groups in our study (see also Ameel, Storms, Malt, & Sloman, 2005, and Malt et al., 2014, for high sorting correlations across language groups for other stimulus sets). These naming and similarity measures were used in our analyses below.

#### 4. Study 1: Historical chaining

In our first study, we asked whether these data provide evidence for historical chaining. That is, has the current extension of the names been developed through a chain of uses expanding over historical time? We test for historical chaining over all three languages in the dataset.

We considered three categorization models, specified in Table 1: a chaining model, a clustering model, and a majority vote model. The chaining model is a nearest-neighbor (or 1-nearest-neighbor) model, which assigns a target item to the category that includes the exemplar most similar to that target item; this is the model that was explored by Sloman et al. (2001). The clustering model is based on Eq. 1 and assigns a target item to the category whose exemplars exhibit the greatest similarity to the target overall. The majority vote model is a baseline model that assigns a target item to the category that has the most exemplars, without reference to any intrinsic relations among exemplars. Similarities  $sim(i, j)$  were determined by the pile-sort data. The category (word)  $w$  for each container object was determined by the modal head noun that was used to label it in the naming data; for example, the category for *juice bottle* would be *bottle* in English.

Each model was tested against the data in a manner that recapitulates the addition of new exemplars to categories over historical time. We began by time-stamping the name of each container item, providing an estimate of when that item appeared in history. We obtained these time stamps from a large historical corpus, the Google Ngram American English corpus (Michel et al., 2011), using a computational procedure described in the Supplementary Material. Following this procedure, we then simulated the sequential emergence of all 60 exemplars in history and asked which of the three models specified above best accounted for categorizations found in the naming data.

Table 1

Summary of models. In the rules below,  $i$  is the target exemplar,  $j$  is any exemplar other than  $i$ ,  $w$  is a lexical category,  $sim(\cdot, \cdot)$  is the similarity between two exemplars, and  $|w|$  is the size of category  $w$

Model	Categorization Rule
Chaining	$i \rightarrow$ the category $w$ of $\arg \max_j sim(i, j)$
Clustering	$i \rightarrow \arg \max_w \sum_{j \in w} sim(i, j)$
Majority vote	$i \rightarrow \arg \max_w  w $

For each model, this predictive analysis proceeded as follows. We first seeded the model with the object that first appeared in history, based on appearance of the corresponding phrase in the corpus; this object was “glass bottle.” As each remaining object became available (based on appearance of its name in the corpus), we used the model to predict its category membership, based only on items already encountered and holding out the category labels for the current item and all upcoming items. The category label was always the head noun associated with that object (e.g., *box* rather than *juice box*). Whenever the model mispredicted the category membership of an item because the object’s name was not yet represented by already-considered exemplars, we introduced that label and thus expanded the repertoire of categories that the model had to choose from among beyond that point in time. We then asked which model best predicted the data, when presented in historical order.

Fig. 4 summarizes the predictive accuracies of these models. The results show that in general, the chaining model accounts for the data better than the two alternative models in all three languages. Specifically, we found the chaining model to outperform the clustering model by about 8%, 3%, and 10% for English, Spanish, and Chinese, respectively (and to outperform the majority vote model by 23%, 35%, and 62% for the three languages). We observed that although the chaining model makes largely similar predictions to the clustering model (the two models overlap in approximately 65% of their predictions within each of the three languages), they are different in interesting ways. Concretely, in the case of English, we observed that the following historical chain of exemplars (dates included in parentheses) was predicted accurately by the chaining model but not by the clustering model: *olive jar* (1903) → *jelly jar* (1925) → *peanut butter jar*

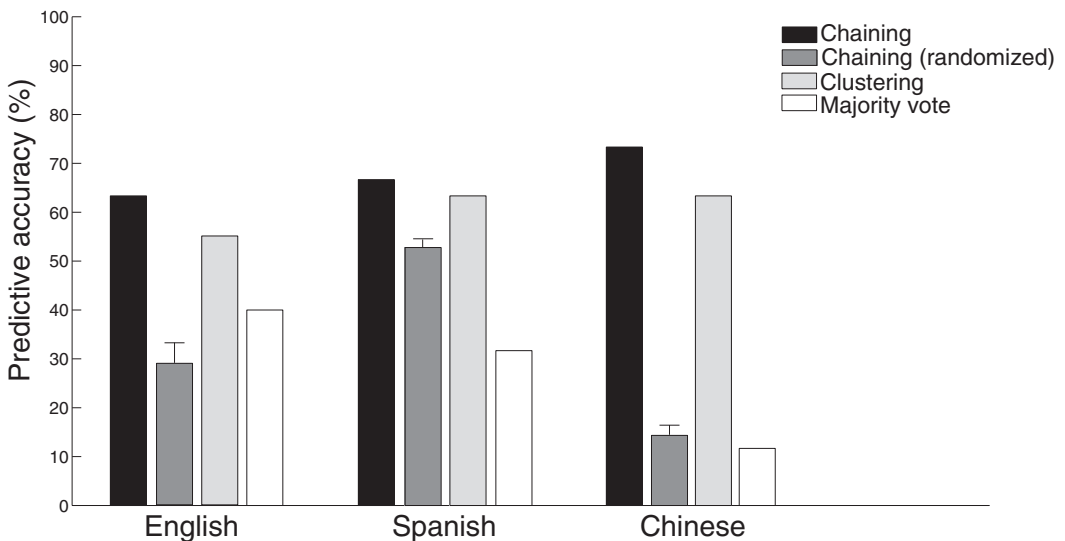


Fig. 4. Summary of historical analysis of chaining. Error bars (standard error) are shown only for the case of randomized chaining, as that is the only case in which multiple simulations were run.



(1945) → *peanut jar* (1963) → *applesauce jar* (1970). Conversely, the clustering model—but not the chaining model—was correct in predicting the category labels of a small number of exemplars such as *film container* (1912) and *squeeze bottle* (1942).

We sought to determine whether the chaining model’s superiority over the clustering model on our data is due to an exceptionally good match between the chaining model and the structure of our data. If so, we would expect the superiority of the chaining model over the clustering model to be greater when assessed relative to our data than when assessed relative to hypothetical variants of our data. A statistical analysis described in the Supplementary Material supports this prediction. We also wished to test whether the good fit of the chaining model is dependent on the historical time-stamps provided by our corpus searches. Would the same chaining model with different time stamps perform as well? To test this, we re-ran the chaining model, but with the temporal sequence of exemplars randomized while keeping the category labels the same. We ran 100,000 such randomized sequences for each language. For all three languages, the mean prediction of the chaining model on the randomized historical sequences was significantly worse than with the real historical sequence ( $p < .01$  via standard permutation tests), indicating that the success of this model does reflect the actual historical emergence of these items.

These results demonstrate that the chaining model accounts well for the diachronic development of the extension of container names. This outcome supports the proposal that historical semantic chaining is involved in the formation of container lexical categories.

Fig. 5 illustrates the historical semantic chaining in the development of the English category *bottle*, on our analysis. We observe that chaining in this real-world category is

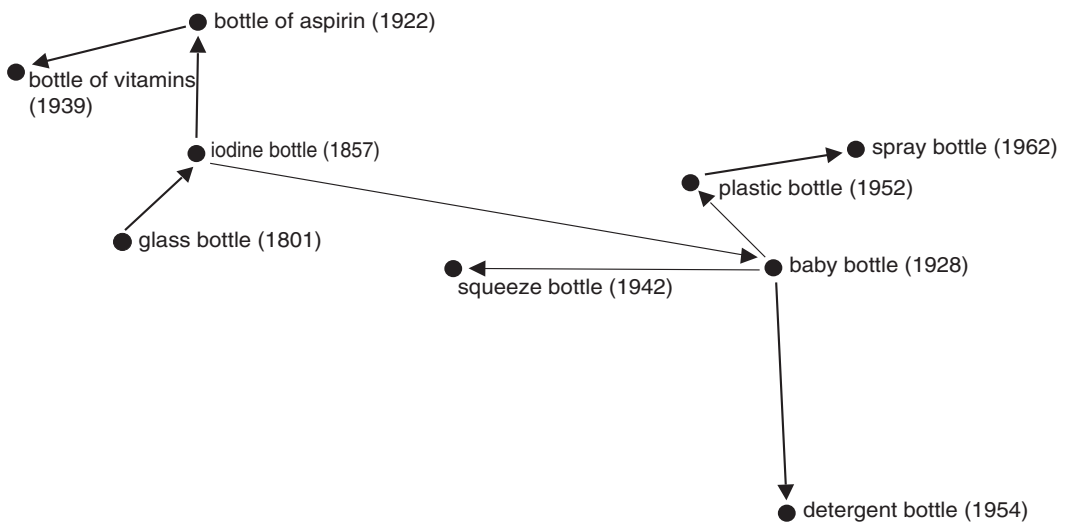


Fig. 5. Semantic chaining in the English *bottle* category. Container names are annotated with dates identified from the corpus searches. Spatial proximity between two items roughly corresponds to their judged similarity. Arrows indicate the trajectory traced by the chaining model.

interestingly different from the idealized case we considered earlier in Fig. 2. Instead of forming a long chain, “natural” semantic chaining in this instance takes the form of short chains grounded in hub exemplars (e.g., *iodine bottle* and *baby bottle*), which form local clusters within the category. Whether such natural chaining structures support efficient communication is a question we will test in the next study.

## 5. Study 2: Chaining and efficient communication

We have seen that semantic chaining has the potential to yield inefficient communication, and that the container naming data of Malt et al. (1999) show evidence of semantic chaining over time. Left unaddressed is whether this natural semantic chaining in fact prevents efficient, informative communication.

The data of Malt et al. (1999) have not previously been analyzed with respect to whether the lexical categories support efficient communication about containers. We conduct that analysis here, separately for each of the three languages (English, Spanish, and Chinese), using the computational formulation of efficient communication specified above. As before, we took the category system of each language to be determined by the modal head noun that was used to label it in the naming data, and we took similarities  $sim(i,j)$  to be determined by the pile-sort data. We estimated the need probability  $N(i)$  for each item  $i$  using frequencies of container names (modifier + head noun) from the Google Ngram American English corpus; specifically, we took frequencies at year 1999, which matches the year of publication of the original work by Malt et al. (1999). We then applied Eqs. 1–3 to obtain the communicative cost for the container naming system in each language.

We take a category system to be near-optimally efficient if it is more informative than most hypothetical systems with the same number of categories. Thus, in order to assess the communicative efficiency of the English, Spanish, and Chinese container naming systems, we need to compare the communicative cost of each to the costs of a large set of hypothetical systems with the same number of categories as those reported by Malt et al. (1999) for this stimulus set (English: 7 categories; Spanish: 15 categories; Chinese: 5 categories). We also constrained the size of each category in a hypothetical system to be equivalent to that in the corresponding attested target system; this constraint ensures that attested and hypothetical categories are identical in number and size, and differ only in the exemplars that are assigned to those categories.

Concretely, for each target language (English, Spanish, Chinese), we constructed hypothetical comparison systems through simulated chaining in similarity space (cf. Khetarpal et al., 2013), as follows. We began by randomly choosing an initial exemplar and assigning an arbitrary category to it. We then extended that name to a new exemplar, which was selected by sampling exemplars in proportion to their similarity with the existing exemplar. We then repeated this chaining process, where the probability of a category name being chosen for expansion was proportional to the number of remaining (as-yet-unlabeled) exemplars in that category. This process continued until we had assigned each

exemplar to a category, such that each category had the same number of exemplars as in the attested target system. This procedure effectively generated a hypothetical chained system in the same similarity space as an attested system. Our use of chained systems as hypothetical competitors provides a conservative test, because it excludes from consideration unnatural-seeming but logically possible hypothetical systems with disconnected (non-contiguous) categories. Recall that in Fig. 2, although chained systems were not highly informative, they were more informative than many other hypothetical systems that are excluded from consideration here.

For each of the three target languages, we created 100,000 such hypothetical chained systems. We then compared the communicative cost of the attested target language to the costs of the hypothetical systems using the formulation presented earlier. The results are shown in Fig. 6.

Each of the three attested systems is significantly less costly than its corresponding class of hypothetical chained systems (English:  $p < .001$ ; Spanish:  $p < .0001$ ; Chinese:  $p < .03$ ). We conclude from these results that although these systems do exhibit semantic chaining, each of them is nevertheless highly informative—and thus tightly clustered—relative to a large class of comparable hypothetical chained systems.

## 6. Discussion

In this paper, we have presented two related contributions. First, we have provided what is, to our knowledge, the first computational demonstration of historical semantic chaining in a modern semantic system, using a large corpus of historical text. Second, we have shown that names for household containers in English, Spanish, and Chinese all support highly informative communication, despite the presence of historical chaining in this domain and the potential of chaining to prevent informative communication.

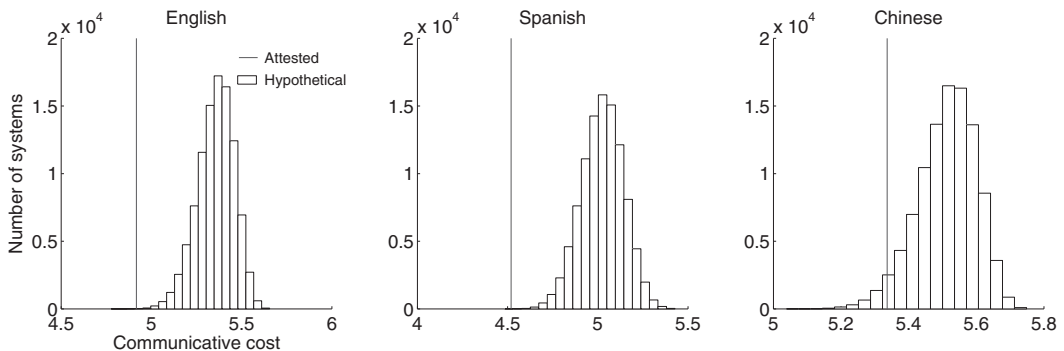


Fig. 6. Efficiency analysis of container naming systems in English, Spanish, and Chinese. In each case, the attested system exhibits low communicative cost (high informativeness) relative to a large set of hypothetical systems of equal complexity.

Our results concern simple nouns for containers, such as *bottle* or *box*, even though these often appear in discourse in modifier-noun combinations such as *aspirin bottle* or *juice box*. The fact that these nouns often appear with a modifier that narrows the intended semantic range might be taken to imply that bare nouns for containers do not by themselves support highly informative communication in the sense identified here, because much of the informativeness is often supplied by the modifier. But our results suggest that these bare nouns are nonetheless near-optimally informative. What are we to make of this? A relevant consideration is that we have shown that systems of bare container nouns across languages are near-optimally informative relative to other possible systems *of the same complexity*—that is, with the same number of separate container words. It is possible that such a system could be nearly as informative as possible for its level of complexity, yet still be inadequately semantically fine-grained to support the needs of daily communication, which often require the addition of a modifier. Our results suggest this interpretation. On this view, the finding of interest is that a system of general categories (bare nouns)—the informativeness of which is often enhanced by the use of a modifier in discourse—nonetheless appears to bear the stamp of communicative pressure when considered without these modifiers.

Our findings leave a number of questions open. How general is the phenomenon of historical chaining, and the nature of it that we have suggested here? Does chaining appear in similar “hub exemplars plus short chains” form in other domains? How general is our finding that historical chaining may be constrained by communicative forces? Future studies can address these questions by applying analyses similar to ours to other domains, as well as new models and analyses, to explore the linguistic packaging of meaning across languages and across time.

## Acknowledgments

This research was supported by NSF award SBE-1041707 to the Spatial Intelligence and Learning Center (SILC). We thank Charles Kemp for earlier ideas on chaining and Rob Kass for the concept of change-point detection.

## Notes

1. In computing cost, we assumed that the distance between horizontally or vertically neighboring objects in the grid is 1, that the similarity between any two objects  $i, j$  is  $sim(i, j) = \exp(-distance(i, j))$ , and that need probability  $N(i)$  is uniform across all objects  $i$ .
2. Naming data were collected from 28 native speakers of English, all students at Lehigh University in the United States; 51 native speakers of Spanish, all from Comahue National University in Argentina; and 50 native speakers of Mandarin

Chinese, 10 of whom were students at Lehigh and 40 of whom were students at Shanghai University in China.

## References

- Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language*, *53*, 60–80.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Brugman, C. M. (1988). *The story of over: Polysemy, semantics, and the structure of the lexicon*. New York: Garland.
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*, 390–406.
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*, 17897–17902.
- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, *24*, 341–380.
- Hopper, P., & Traugott, E. (2003). *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.
- Khetarpal, N., Neveu, G., Majid, A., Michael, L., & Regier, T. (2013). Spatial terms across languages support near-optimal communication: Evidence from Peruvian Amazonia, and computational analyses. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 764–769). Austin, TX: Cognitive Science Society.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago.
- Levinson, S. C., & Meira, S. (2003). “Natural concepts” in the spatial topological domain-adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, *79*, 485–516.
- Malt, B. C. (2010). Naming artifacts: Patterns and processes. *Psychology of learning and motivation*, *52*, 1–38.
- Malt, B. C., & Majid, A. (2013). How thought is mapped into words. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 583–597.
- 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.
- Malt, B. C., Ameel, E., Imai, M., Gennari, S. P., Saji, N., & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language*, *74*, 107–123.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D. Novig, P., Orwant, J., Pinker, S., Nowak, M. A. & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*, 3526–3529.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436–1441.

- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & O. Grady (Eds.), *The handbook of language emergence* (pp. 237–263). Hoboken, NJ: Wiley-Blackwell.
- Sloman, S. A., Malt, B. C., & Fridman, A. (2001). Categorization versus similarity: The case of container names. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 73–86). New York: Oxford University Press.
- Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 1348–1353). Austin, TX: Cognitive Science Society.
- Tribus, M. (1961). *Thermodynamics and thermostatics: An introduction to energy, information and states of matter, with engineering applications*. Princeton: D. Van Nostrand Company Inc.
- Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual meeting of the Cognitive Science Society* (pp. 1802–1807). Austin, TX: Cognitive Science Society.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Supplementary material