WHAT 50 MILLION DRAWINGS CAN TELL US ABOUT SHARED MEANING

MOLLY LEWIS[*1,2] and GARY LUPYAN[1]

[*]Corresponding Author: mollyllewis@gmail.com
[1]Department of Psychology, University of Wisconsin-Madison
[2]Computation Institute, University of Chicago

A foundational assumption of linguistic communication is that conversants have similar underlying concepts (Brennan & Clark, 1996; Wierzbicka, 2012). On this view, the ability of one person to understand another when she says "the tree" depends on the word activating the same concept in both people. One approach to verifying this assumption is to rely on definitions, but this reasoning is circular—how can we be sure the words *in* our definitions are the same? Here, we investigate the assumption of shared linguistic concepts by studying concepts represented in the visual modality—drawings—and examining predictors of their variability. Specifically, we ask whether people who are geographically closer and inhabit a similar linguistic environment produce more similar drawings.

We analyzed a dataset of 50 million drawings (of mostly concrete artifacts such as "tree") from 15 million participants (QuickDraw: quickdraw.withgoogle.com/data). Although all drawings were elicited in English, the participants spanned the globe and, we can assume, represent a wide variety of cultural and linguistic experience. Such drawings only capture a part of meaning—people know much more about trees than what they look like—and therefore offer a conservative estimate of diversity.

We quantified similarity of drawings in two ways: (1) Hausdorff Distance (HD), which quantifies image similarity as the minimum Euclidean distance between two sets of points (Huttenlocher, Klanderman, & Rucklidge, 1993); (2) the internal weights (layer FC2) for each of our drawings from a neural net model trained on ImageNet (Deng, 2009), with similarity corresponding to the cosine distance (CD) on weights between pairs of images.

Initial analysis included 1500 image pairs of two categories—"bread" and "tree"—from participants located in 72 countries. We validated our similarity measures using human judgements. We selected 20 pairs from each HD decile for each category (Fig. 1a), and asked participants to rate the similarity of the objects in the drawings using a 1 (almost identical) to 7 (completely different) Likert scale. Each participant ($N = 100$) rated 50 pairs drawings from the same category.
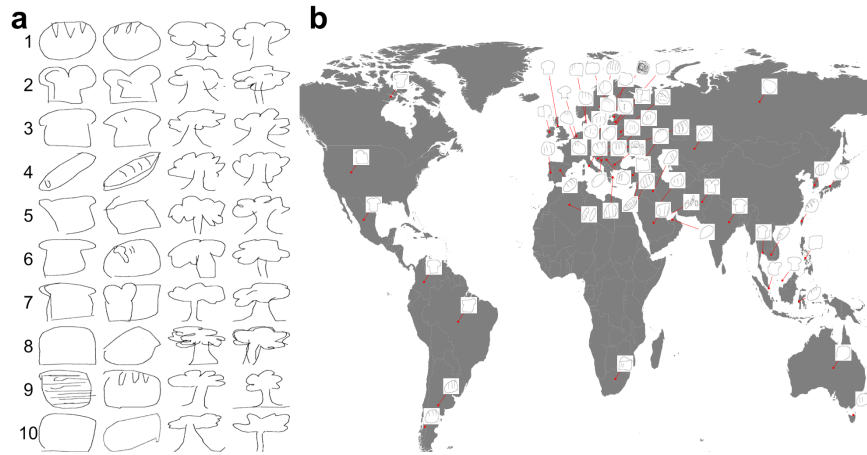
Figure 1. (a) Sample drawing pairs ("bread" and "tree") used in the human similarity judgement experiment. Each row corresponds to a Hausdorff decile. (b) The prototypical "bread" drawing for each country, calculated as the drawing with the smallest average pairwise distance to other drawings from the same country.

Human judgements of similarity were highly correlated with HD (r = .29, $p <$ .0001) as well as CD (r = 0.20, $p <$ .0001). In a mixed effect model with HD and CD as fixed effects, the two measures were simultaneously predictive of human similarity judgements (HD: β = .35; t = 12.39; CD: β = -.26; t = -9.95) and thus appeared to capture different aspects of visual similarity.

With our automated similarity measures validated, we next examined predictors of variability in drawing similarity. We sampled 42,900 pairs of drawings across countries for "bread," "tree" and 15 additional items, and then calculated the HD for each pair. We quantified geographic distance as the distance in meters between the centroid of each pair of countries. We quantified linguistic distance in two ways: (1) vocabulary overlap (ASJP database; Bakker, et al., 2009, Dediu, in press); (2) grammatical similarity based on features values from the WALS typological database (Dediu, in press). The best fitting model revealed an effect of all three distance measures on picture similarity, and pointed to an interaction between grammatical similarity, vocabulary overlap and geographical distance: For languages that differed more in terms of their grammar, countries with greater overlap in vocabulary (β = .002; t = 3.27) or smaller geographic distance tended to have more similar drawings (β = -.002; t = -2.17).

These data reveal systematic cross-cultural variability in semantics, and suggest that speakers' physical and linguistic proximity may contribute to convergence on shared semantics.

References

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., Egorov, D., Mailhammer, R., Grant, A., & Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology*, *13*(1), 169–181.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Dediu, D. (in press). Making genealogical language classifications available for phylogenetic analysis: Newick trees, unified identifiers, and branch length. *Language Dynamics and Change*.

Deng, J. (2009). Imagenet: A large-scale hierarchical image database. 248–255.

Huttenlocher, D. P., Klanderman, G., & Rucklidge, W. (1993). Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*, 850–863.

Wierzbicka, A. (2012). Understanding others requires shared concepts. *Pragmatics & Cognition*, *20*(2), 356–379.