COMMUNITIES OF DIFFERENT SIZE CREATE DIFFERENT CATEGORIZATION SYSTEMS

SHIRI LEV-ARI[*1]

[*]Corresponding Author: shiri.lev-ari@rhul.ac.uk
[1]Dept. of Psychology, Royal Holloway University of London, UK

Categorization is the foundation of many cognitive functions, and the way in which we categorize the world is informed by the language we speak. Languages, however, differ in the granularity of the categories they encode, and the source of these cross-linguistic differences is poorly understood. Prior research has shown that larger communities have larger phonological inventories in both human (Hay & Bauer, 2007) and non-human animals (e.g., Freeberg, 2006; McComb & Semple, 2005) as well as create more systematic languages (Lupyan & Dale, 2010; Raviv, Meyer & Lev-Ari, 2019). This paper tests whether community size and density can also influence the granularity and structure of semantic categories, a domain that at the interface between language and cognition. As such, this study will not only add to our understanding of why and how languages evolved to have the forms they have, but it will also have implication for linguistic and non-linguistic cognitive performance, as the way a language encodes a category has implications for memory, attention, and even low level perception (e.g., Roberson, Davidoff, Davies & Shapiro, 2005; Winawer et al., 2007).

The positive association between community size and the size of the phonological inventory suggests that larger communities are likely to develop more granular categories, that is, categories with more sub-divisions than those of smaller communities. Additionally, larger communities' tendency to develop more systematic symbols suggests that they might also develop categorization systems with more informative structure that leads to better performance. In addition, we tested whether community density plays a similar role since sparsity, similarly to size, can influence input variability and diffusion mechanisms.

The proposal that community structure can influence the community's categorization system was tested with simulations in which populations of either 100 or 200 agents communicated about a 20x20 meaning space for 50,000 rounds.

Populations were generated using the barabasi_albert_graph in the Python package, Networkx (Barabási & Albert, 1999). Density was manipulated using the m parameter (m=20 vs 50). In each round, each agent communicated to someone in their network about a randomly selected meaning. Success was measured by the distance of the comprehended meaning from the intended one. Agents selected a label to produce by searching their history for the closest and most successful label, and when none was available, created a new label by randomly combining 3 phonemes. Partners interpreted the label according to their past experience with the label (weighting tokens by their success). If a label or a meaning was not used for at least 500 interactions, the agent forgot it.

Results showed that larger communities divided the meaning space into more categories (Fig 1a). Furthermore, these categories were more balanced in size (Fig 1b). As a consequence, larger communities communicated more successfully (Fig 1c). The effects of density were smaller in magnitude and less consistent.

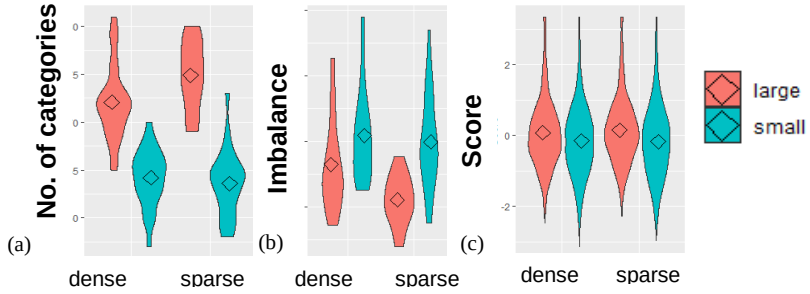

Figure 1. Properties of the categorization system as a function of community size and community density.

Further analyses revealed that it was larger communities' greater diffusion constraints that led to their superior behavior. The statistical analyses showed that: Labels were less likely to spread in larger communities, and even when they did, there was lower agreement on their meaning. This lesser agreement led categories to narrow their meanings, which, in turn, enabled their maintenance and the creation of more granular and balanced categories.

Lastly, matching data from the World Color Survey with population size from Ethnologue (controlling for language family) showed that larger populations have more color terms, providing initial real-life support for our findings.

This study shows that community structure influences the categorization system, such that larger communities create more granular and better structured categories that can support better communication. Moreover, the study shows how greater constraints – diffusion barrier – can ultimately improve a system. As greater granularity can add processing costs, future research should examine how larger communities might balance out the costs (e.g., more distinct labels).

References

Barabási, A. L., & Albert, R. (1999) Emergence of scaling in random networks. *Science 286*, 5439, 509-512.

Lupyan, G., and Dale, R. (2010) Language Structure Is Partly Determined by Social Structure. *Plos One* 5, e8559.

Freeberg, T. M. (2006) Social complexity can drive vocal complexity: group size influences vocal information in Carolina chickadees. *Psychological Science, 17, 7*, 557-561.

Hay, J. & Bauer, L. (2007) Phoneme inventory size and population size. Language 83, 388–400.

McComb, K. & Semple, S. (2005) Coevolution of vocal communication and sociality in primates. *Biology Letters 1*, 381 –385.

Raviv, L., Meyer, A.S. & Lev-Ari, S. (2019) Larger communities create more systematic languages. *Proceedings of The Royal Society B 286*, 201912.

Roberson, D., Davidoff, J., Davies, I. R., & Shapiro, L. R. (2005) Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology 50, 4*, 378-411.

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.