Information-theoretic efficiency and semantic variation: The case of color naming

Noga Zaslavsky[1,3] **(noga.zaslavsky@mail.huji.ac.il)** **Charles Kemp**[2] **(ckemp@cmu.edu)**

Terry Regier[3,4] **(terry.regier@berkeley.edu)** **Naftali Tishby**[1,5] **(tishby@cs.huji.ac.il)**

[1]Edmond and Lily Safra Centre for Brain Sciences, The Hebrew University, Jerusalem 9190401, Israel
[2]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213 USA
[3]Department of Linguistics and [4]Cognitive Science Program, University of California, Berkeley, CA 94720 USA
[5]The Benin School of Computer Science and Engineering, The Hebrew University, Jerusalem 9190401, Israel

Keywords: information theory; semantic typology; color naming; categories; language evolution.

A major question in semantic typology is how to account for semantic universals and variation in a principled and unified way. One approach to this question proposes that word meanings may reflect adaptation to pressure for efficient communication. Color is a semantic domain that has been approached in this spirit. Recent work has relied on the notion of the 'informativeness' of word meaning, has often cast that notion in terms borrowed from information theory, and has accounted for several aspects of color naming across languages on that basis (e.g. Jameson & D'Andrade, 1997; Regier et al., 2007; Baddeley & Attewell, 2009; Regier et al., 2015).

However, two fundamental issues have been left largely unaddressed. First, little is known about how a drive for efficiency may relate to present accounts of color term evolution. Second, while previous accounts of color naming invoked the general idea of efficiency (e.g. Regier et al., 2015), they did not ground their argument comprehensively in independently motivated formal concepts of efficiency from information theory. Thus, it is not yet clear to what extent variation in color naming can be explained in terms of information-theoretic efficiency, and whether doing so can explain aspects of color naming previously left unexplained.

In this work we examine this open question. We argue that languages efficiently compress ideas into words by optimizing the tradeoff between the complexity and accuracy of the lexicon according to the Information Bottleneck (IB) principle (Tishby et al., 1999), a formal principle with broad scope which is closely related (Harremoës & Tishby, 2007) to Shannon's rate distortion theory. We support this claim by showing that: (1) color naming systems across languages lie near the information-theoretic limit of efficiency; (2) small changes in a single tradeoff parameter account to a large extent for observed cross-language variation; (3) efficient IB color naming systems exhibit soft rather than hard category boundaries, and often leave large regions of color space inconsistently named, both of which phenomena are found empirically; and (4) these IB systems evolve through a sequence of structural phase transitions, in a process that captures key ideas from Berlin and Kay's theory (Berlin & Kay, 1969), as well as ideas associated with more continuous accounts (MacLaury, 1997; Levinson, 2000) of color category evolution. These results suggest that a drive for information-theoretic efficiency may shape color naming systems across languages. This principle is not specific to color, and so it may also apply to cross-language variation in other semantic domains.

Acknowledgments

This study was supported by the Gatsby Charitable Foundation (N.T.), IBM Ph.D. Fellowship Award (N.Z.), and DTRA award HDTRA11710042 (T.R.). Part of this work was done while N.Z. and N.T. were visiting the Simons Institute for the Theory of Computing.

References

Baddeley, R., & Attewell, D. (2009). The relationship between language and the environment: Information theory shows why we have only three lightness terms. *Psychological Science*, *20*(9), 1100-1107.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley and LA: UC Press.

Harremoës, P., & Tishby, N. (2007, June). The Information Bottleneck revisited or how to choose a good distortion measure. In *ISIT* (p. 566-571).

Jameson, K., & D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: An inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 295–319). Cambridge University Press.

Levinson, S. C. (2000). Yélî Dnye and the theory of basic color terms. *J. Ling. Anthropol.*, *10*(1), 3-55.

MacLaury, R. E. (1997). *Color and cognition in Mesoamerica: Constructing categories as vantages*. University of Texas Press.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *PNAS*, *104*(4), 1436-1441.

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 237–263). Hoboken, NJ: Wiley-Blackwell.

Tishby, N., Pereira, F. C., & Bialek, W. (1999). The Information Bottleneck method. In *Proceedings of the 37th annual Allerton conference on communication, control and computing.*