

**LANGUAGES SUPPORT EFFICIENT COMMUNICATION
ABOUT THE ENVIRONMENT: WORDS FOR SNOW
REVISITED**

TERRY REGIER¹, ALEXANDRA CARSTENSEN²

*Departments of Linguistics¹ and Psychology², University of California, Berkeley,
Berkeley, CA, USA
terry.regier@berkeley.edu, abc@berkeley.edu*

CHARLES KEMP

*Department of Psychology, Carnegie Mellon University
Pittsburgh, PA, USA
ckemp@cmu.edu*

We bring together two influential ideas about language: the vexed and much exaggerated claim that Eskimo languages have multiple words for subtypes of snow because of the nature of their environment (Boas, 1911; Whorf, 1956; Martin, 1986; Pullum, 1991), and the independent theory that language is shaped by the functional need for efficient communication (e.g. Piantadosi et al., 2011; Fedzechkina et al., 2012; Kemp & Regier, 2012; Kirby et al., 2015). We show that a natural variant of the Eskimo snow claim is supported across a broad range of languages, in line with this independent theory.

We take a communicative system to be efficient to the extent that it is informative and precise, yet requires minimal effort to use. From this perspective, a system of fine-grained categories is both more informative than a single broad category, and more complex, requiring more effort to store and process. The added complexity of a fine-grained system may be worth the investment if the gain in informativeness is compounded by frequent use of the fine-grained categories. This reasoning predicts that semantically fine-grained categories will tend to appear in frequently referenced parts of semantic space (Greenberg, 2005; Kemp & Regier, 2012), consistent with Boas' original Eskimo snow claim.

The same line of reasoning applied in the opposite direction predicts that semantically broad linguistic forms that can refer to either snow or ice will tend to be found in warmer climates, as in Aztec (Nahuatl) (Whorf, 1956: 216), because of reduced need to refer to snow or ice in such climates.

We conducted a survey of library reference works covering 50 genetically diverse languages, following as closely as possible the survey by Bybee et al. (1994), and a separate survey of 166 languages based on the online Intercontinental Dictionary Series. We asked whether languages spoken in warm climates are more likely than those spoken in cold climates to have a semantically broad form that can refer to either snow or ice, rather than separate forms for snow and ice. Mixed-effects logistic regression analyses revealed that this was the case, in each survey, and that this predicted association between temperature and use of a single semantically broad form is not attributable to relatedness of the languages concerned. For example, we found that a single semantically broad form for snow and ice is found in, among many others: Hausa (Chadic, African Savannah), Hawaiian (Austronesian, Oceania), Panare (Cariban, NE South America), Telugu (Dravidian, Indic), and Yaqui (Uto-Aztecan, Mesoamerica); all genetic and areal affiliations listed here are drawn from the Autotyp database (Nichols et al., 2013). Finally, a cross-language analysis of messages posted to Twitter confirmed that speakers of languages in warm climates do refer to snow and ice less often than speakers of languages in cold climates. Our results confirm that variation in semantic categories across languages may be traceable in part to local communicative needs. They suggest moreover that despite its awkward history, the topic of “words for snow” may play a useful role as an accessible instance of the principle that language supports efficient communication.

Acknowledgments

We thank Bernard Comrie for helping us to access the IDS data, Brendan O'Connor for compiling and helping us to access the Twitter data, Jongmin Baek and Linda Lee for their help with the library survey, and Susanne Gahl, Sharon Inkelas, Paul Kay, Michael Klein, Steven Piantadosi, Geoffrey Pullum, Sean Roberts, and Yang Xu for their comments; their being listed here should not be taken to imply their endorsement of our argument. Any errors are our own.

References

- Boas, F. (1911). Introduction. In: *Handbook of American Indian Languages*, Vol. 1, Bulletin 40 (pp. 1-83). Washington, DC: Government Print Office.
- Bybee J. et al. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Fedzechkina, M. et al. (2012). Language learners restructure their input to facilitate efficient communication. *PNAS*, 109, 17897-17902.
- Greenberg, J.H. (2005). Universals of kinship terminology. In: *Language universals: With special reference to feature hierarchies* (pp. 72-87). Berlin: Mouton de Gruyter.
- Kemp, C. & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336, 1049-1054.
- Kirby, S. et al. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87-102.
- Martin, L. (1986). "Eskimo words for snow": A case study in the genesis and decay of an anthropological example. *American Anthropologist*, 88, 418-423.
- Nichols, J. et al. (2013). The AUTOTYP genealogy and geography database: 2013 release. <http://www.autotyp.uzh.ch/>
- Piantadosi, S.T. et al. (2011). Word lengths are optimized for efficient communication. *PNAS*, 108, 3526-3529.
- 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). Chicago: University of Chicago Press.
- Whorf, B.L. (1956). Science and linguistics. In: Carroll, J.B. (Ed.), *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* (pp. 207-219). Cambridge, MA: MIT Press.