WORD LEARNERS REGULARIZE SYNONYMS AND HOMONYMS SIMILARLY

VANESSA FERDINAND

The Santa Fe Institute
Santa Fe, U.S.A
vanessa@santafe.edu

MATTHEW SPIKE

School of Philosophy, Psychology, and Language Sciences, University of Edinburgh
Edinburgh, U.K.
matthew.spike@ed.ac.uk

Unpredictable variation is rare in language. Explanations for this include a language-specific *regularization bias* (e.g. Reali & Griffiths, 2009), general constraints on memory (e.g. Hudson Kam & Newport, 2005), or both (e.g. Perfors, 2012; Ferdinand, Thompson, Kirby, & Smith, 2013). Experiments on lexical regularization typically study how words in free *synonymous* relationships become increasingly deterministic through use (e.g. some words drop from use). There is also experimental evidence that learners regularize *homonymous* relationships (Vouloumanos, 2010), but to date no experimental design has directly compared the relative regularization of synonyms versus homonyms. This is an important comparison to make because synonyms and homonyms have asymmetrical functional roles in communication (Hurford, 2003) and the jury is still out as to which of these two regularization biases are better for evolving effective communication systems. On one hand, Hurford proposes that there is less bias against homonyms because they are more common in language than synonyms and Piantadosi, Tily, and Gibson (2012) argue for the communicative function of ambiguous lexicons. On the other hand, Doherty (2004) demonstrates children's difficulty in learning homonyms and Spike, Stadler, Kirby, and Smith (2013) show that self-organizing novel lexicons require a bias against homonymy but not synonymy.

We extend the experimental paradigm of Ferdinand et al. (2013) to investigate the relative regularization of synonyms versus homonyms. 128 participants were trained on one of two artificial mini-languages with identically matched distributions of variation. In the synonyms condition, this variation was over word forms and in the homonyms condition it was over referents. Regularization is quantified by the drop in Shannon entropy of the words and referents that participants

produced when tested on their mini-language. Participants regularized 67% of the variation among homonyms ($t(63) = -12.8169, p < .001$) and 56% of the variation among synonyms ($t(63) = -10.5526, p < .001$). However, there was no significant difference between these conditions ($t(126) = 1.3518, p = 0.18$), suggesting that learners compress synonymous and homonymous variation similarly.

This experiment was repeated with non-linguistic stimuli, where participants learned the mappings between marbles and the different containers they were drawn from. Participants also regularized the non-linguistic stimuli, eliminating 42% of the variation among containers ($t(63) = -7.277, p < .001$) and 32% of the variation among marbles ($t(63) = -6.6908, p < .001$), again with no significant difference between conditions ($t(126) = 1.5049, p = 0.13$). This suggests a domain-general component to linguistic regularization. However, participants in the linguistic conditions regularized significantly more than those in the non-linguistic conditions ($F(252) = 11.259, p < .001$). We conclude that regularization results from general-purpose compression during learning, which can be ramped up for effective communication with linguistic stimuli, and operates similarly on synonyms and homonyms.

References

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*(1), 203–214.

Ferdinand, V., Thompson, B., Kirby, S., & Smith, K. (2013). Regularization behavior in a non-linguistic domain. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 436–441).

Hudson Kam, C., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195.

Hurford, J. (2003). Why synonymy is rare: Fitness is in the speaker. In *Proceedings of ECAL 7* (pp. 442–451). Berlin: Springer Verlag.

Perfors, A. (2012). When do memory limitations lead to regularization? An experimental and computational investigation. *Journal of Memory and Language*, 1–60.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Reali, F., & Griffiths, T. L. (2009, June). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–28.

Spike, M., Stadler, K., Kirby, S., & Smith, K. (2013). Learning, feedback and information in self-organizing communication systems. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3442–7).

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742.