QUALITY NOT QUANTITY IN SYNONYMY AVOIDANCE

Aja Altenhof[1,2] and Gareth Roberts[*2]

[*]Corresponding Author: gareth.roberts@ling.upenn.edu
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, Scotland
[2]Department of Linguistics, University of Pennsylvania, Philadelphia, USA

Words often compete for meaning and synonymy is extremely common. But *absolute* synonymy is exceedingly rare (Murphy, 2003). How does this situation arise? One possibility is that it results from a cognitive bias against synonymy, such as the mutual exclusivity bias (Markman & Wachtel, 1988; Hurford, 2003). Another possibility is that it is caused by differences in the perceived contextual distribution of candidate synonyms, which become amplified in learning (Hudson Kam & Newport, 2009).

To investigate this we conducted an experiment in which 362 participants were asked to learn three new slang words—presented in example sentences—and (as a cover task) to guess what the words meant. Two of the words (*snater* and *fincur*) were verbs, and participants were told they referred to the same action. The third word (*murp*) was a noun and functioned purely as a distractor. In the initial *Exposure phase*, 36 sentences were presented on screen, one by one, with 10 s between each sentence. Sentences were presented in a random order, with 12 per word. Verbs could appear in negative contexts (e.g., "Don't fincur in front of me!"), positive contexts (e.g., "My friends and family love my snatering skills."), or neutral contexts (e.g., "Dogs can't snater"). The noun *murp* appeared only in neutral contexts.[1]

After the Exposure phase came the *Generalization phase* in which participants were presented with 36 unseen sentences, each with a gap for participants to insert one of the three words. Of these sentences, 24 required verbs (12 positive and 12 negative) and the remaining 12 required nouns.

We manipulated the distribution of contexts in the Exposure phase. In the Neutral condition, all sentences were neutral. In the Random condition, both verbs were evenly distributed across positive and negative contexts. In the Consistent condition, one verb appeared only in positive sentences and the other appeared only in negative sentences. In the Overlapping condition, one verb occurred 75%

[1]Sentences were chosen following extensive piloting to ensure that the intended difference in valence existed.

of the time in positive sentences and 25% in negative sentences; the reverse was true for the other verb. For the 75%-positive condition, both verbs occurred 75% of the time in positive sentences and 25% in negative sentences; this distribution was reversed for the 75%-negative condition.

To identify how much participants differentiated the verbs in the Generalization phase we calculated, for each verb, a *word context score* by dividing its frequency in its higher-frequency context by the sum of its frequencies for both contexts. We then multiplied the context scores for the two verbs to get a *Differentiation score*. A Differentiation score of 0 meant that one verb had not been used at all. A score of 1 meant that the two verbs were 100% differentiated. 0.25 meant that both verbs were used equally often in both contexts (equivalent to random assignment). A score around 0.667 indicated "partial differentiation", where one verb was entirely specialized to one context and the other played a more default role and occurred in both (as with, e.g., English *thin : skinny*).

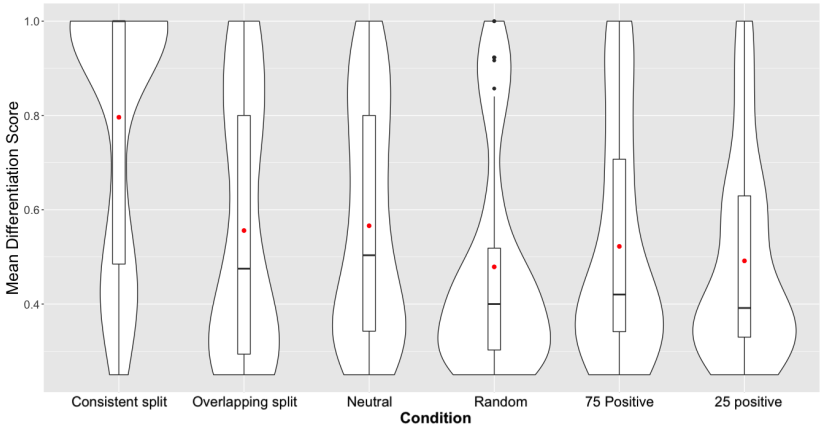

Figure 1. Differentiation score distributions for all conditions. Red dots indicate mean values.

Results are displayed in Figure 1. In the Consistent condition, full (or close-to-full) differentiation occurred. In all other conditions, regardless of Exposure-phase distribution, we found partial differentiation. A replication in which neutral sentences were included in the Generalization phase found the same result.

We consider these results to be consistent with an account based on a cognitive bias against synonymy (cf. Hurford, 2003). Participants differentiated potential synonyms at a greater than chance level and to an extent inconsistent with their distribution. This is not to say that distribution was irrelevant; there was significantly greater differentiation in the Consistent condition. However, it seems that the quality of differentiation in Exposure was more important than quantity.

References

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, & J. T. Kim (Eds.), *European conference on artificial life* (pp. 442–451). Dortmund.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms.* Cambridge: Cambridge University Press.