

3-year-old's referring expressions balance familiarity with informativity

Colin Bannard¹, Marla Rosner¹ and Danielle Matthews²

1. University of Texas at Austin

2. University of Sheffield

As has long been understood (Zipf, 1949), the relationship between the frequency of linguistic forms and their utility to language users is complex. While frequent forms are quicker to be learned and easier to deploy than rare forms, rarer linguistic forms usually carry more information. For example if we refer to someone as a *young man*, we are, in most contexts, giving less information about them (discriminating them less from other men) than if we use a less frequent term (e.g. *mean man*). Recent work (Bannard and Matthews, 2008) has shown children are better at producing sequences of words that they hear frequently than matched rarer phrases. While it is valuable for the child to be competent with such familiar expressions, children also need to produce informative utterances. We here explore how children negotiate these competing motivations early in language development.

We paired 8 nouns with 2 different adjectives each (giving 16 unique phrases, all attested in CHILDES). The corpus frequency of each phrase and of the component words was varied so that one of each phrase for each noun was a collocation (higher frequency than would be expected given the frequency of the component words, as measured using pointwise mutual information, Mackay, 2003; e.g. *old woman*, *pretty dress*) and the other was not (e.g. *kind woman*, *little dress*). Twenty-three 3-year-old children have taken part so far. In a first stage they would sit with E1 and look through pictures depicting each phrase and E1 would describe them. A second experimenter would then come in. They would look through the pictures again and the child was asked to tell E2 what s/he could see. For 11 of the children E2 was attending to the picture described, while for the other 12 she was unable to see it. We were interested in how the children's choice of referring term would be affected by the frequency and information content of E1's phrases, as well as in how this might interact with E2's attentional state.

All except one of the children's descriptions featured a noun, and in 80% of cases this was the same noun as E1 used. However children showed great variability in their use/non-use of adjectives. We analyzed the data using multilevel logistic regression models. As can be seen in figure 1, the probability that a child would use an adjective-noun phrase rather than a bare noun to describe each picture increased as a function of the information content of the adjective the experimenter had used (Odds ratio=1.3, $p=0.05$). E2's attentional state had no effect. This indicates that children were more likely to refer to an attribute referenced by E1 when that attribute was referred to with a high information adjective. However, when the children did produce an adjective they were more likely to produce exactly the adjective produced by E1 (rather than e.g. a paraphrase) when that adjective was part of a collocation. These two effects suggest that children's early referring expressions balance a preference for familiarity with a drive to be informative.

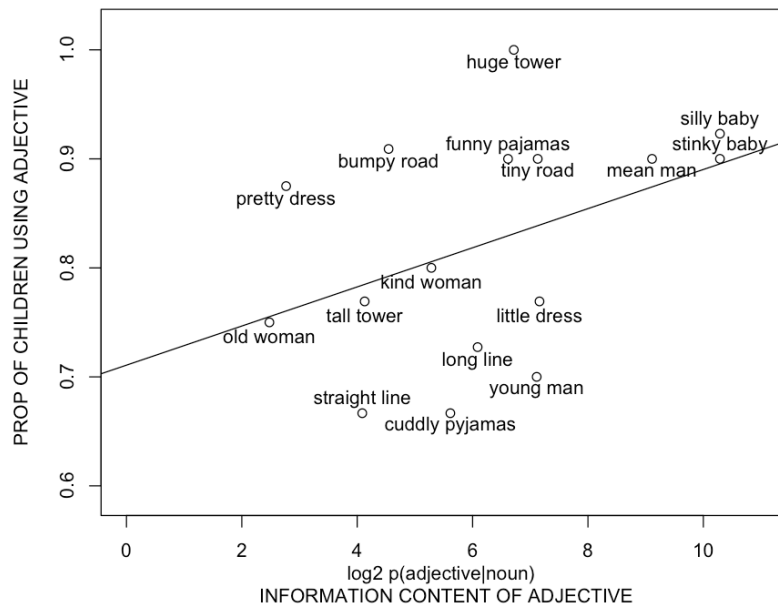


Figure 1.

References

- Bannard, C. and Matthews, D.E. (2008). Stored Word Sequences in Language Learning, *Psychological Science, 19*, 241-248.
- Zipf, G.K (1949). *Human behavior and the principle of least effort*. Addison-Wesley.
- MacKay, D.J.C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.