The shape of frequency distribution and novel construction learning

Grzegorz Krajewski[1], Anne-Kristin Siebenborn[2], and Elena Lieven[1,3]
[1]University of Manchester, [2]Ludwig-Maximilians-Universität München, [3]MPI EVA Leipzig
contact: krajewski@manchester.ac.uk

The aim of these studies was to test how the shape of its frequency distribution affects learning of a novel construction. There is some evidence in the literature suggesting that such learning is facilitated if one item appears in the construction much more frequently than others (skewed distribution; Casenhiser & Goldberg, 2005). However, studies showing this effect presented a novel construction with verbs that were novel too, which complicates the task and perhaps makes it ecologically less valid. Given that high token frequency reduces an item's contribution to the process of abstraction (Bybee, 1995), one might expect a more balanced distribution to be beneficial. Indeed, results of Matthews and Bannard (2010) suggest that generalisations of partially schematic constructions might be easier, if their distributions are characterized by greater entropy. Since the flatter a distribution the greater its entropy, it suggests a facilitatory effect of flat distributions.

We revisit the role of the skewness of a distribution, using familiar verbs to teach children a novel construction. To better accommodate for the continuous nature of the factor, we test three rather than two levels of skewness. To test whether the type of construction affects learning, we use a morphological construction (verbal prefix) in Study 1 and a syntactic construction (word order SOV) in Study 2. In both studies the meaning of the construction is "pretence". We use 6 verbs during the training. For each of them, children first see a film showing the corresponding action accompanied by a recorded sentence featuring the verb. A second film follows, featuring an actor pretending to perform the action and accompanied by a sentence with the verb used in the novel construction. The number of times the verbs are presented this way varies as a between-subject factor. In the flat condition each verb is presented twice (2-2-2-2-2-2), in the skewed condition one verb is unusually frequent (7-1-1-1-1-1), with the semi-skewed condition lying in-between (5-3-1-1-1-1). The overall number of presentations is always 12. Children in the control group don't receive any training. In the test phase, a pointing task, children hear 12 different familiar verbs, while watching 12 pairs ('real' and 'pretend') of simultaneously played films. For half of the items the verb is used in the novel construction ('pretend') and for half of the items it's not ('real'). Children have to point to a corresponding film.

117 English-speaking children (5;0-7;0, M=6;2, Mdn=6;4) took part in Study 1. The results show a significant and consistent effect of the skewness of the distribution: the flatter the distribution, the easier learning of a novel prefix is. In Study 2, 81 children were tested (5;4-6;5, M=5;11, Mdn=5;11). The results reveal a somewhat different picture: the skewed condition remains the most difficult one but the semi-skewed, rather than the flat one, turns out to be the easiest condition. It suggests that, while little variation (skewed distribution) makes learning of both morphology and word order difficult, too much variation (introduced by the flat distribution) is not so helpful either, when children have to spot and learn to use a new word order pattern. Perhaps, the difference between the two studies stems from the fact that a word order is more abstract than a morpheme and thus too much variation makes it more difficult to notice it as a recurring pattern.

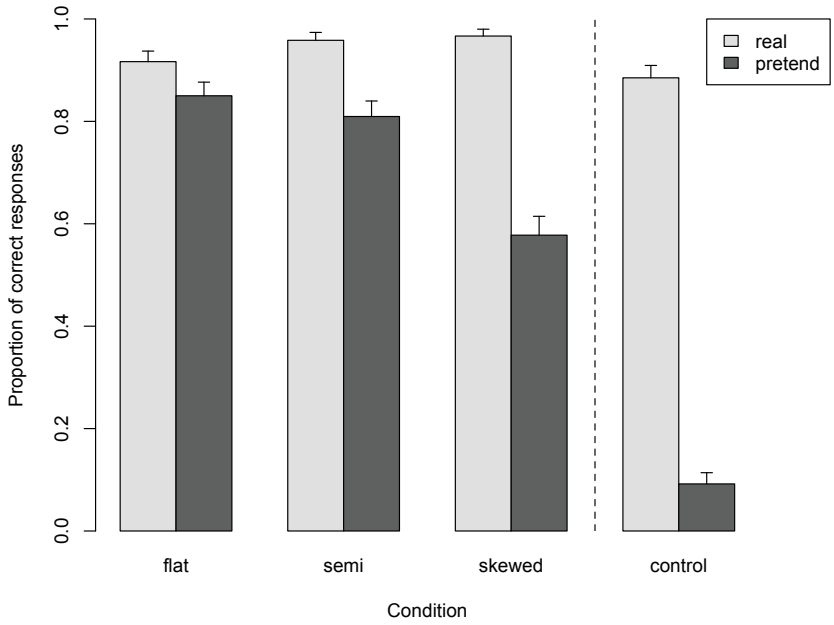

Figure 1. Results of Study 1 (novel prefix)

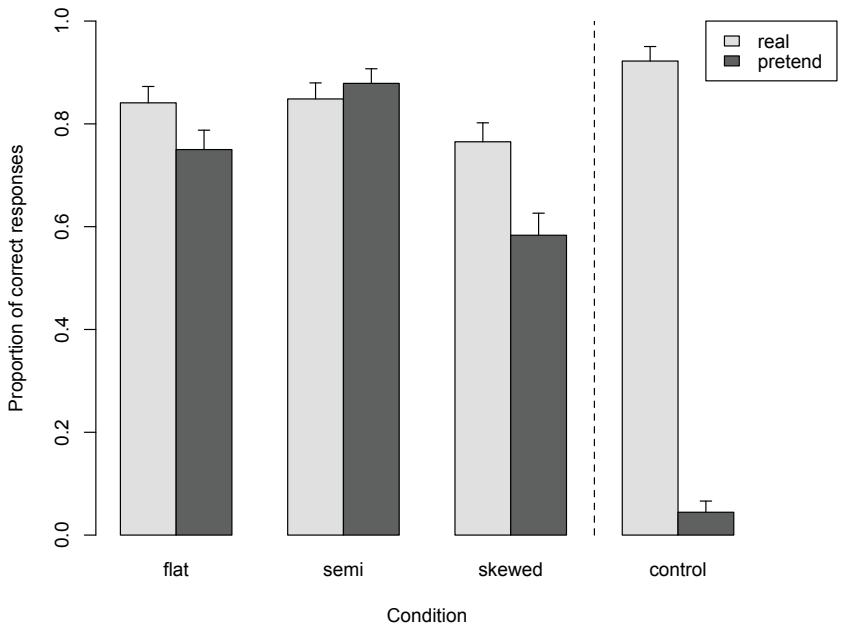

Figure 2. Results of Study 2 (novel word order)

References:

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*(5), 425–455.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8*(6), 500–508.

Matthews, D. & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science, 34*, 465-488.