

Productivity of a Polish Child's Inflectional Morphology: Corpus-based Studies

Grzegorz Krajewski (University of Manchester, UK & MPI for Evolutionary Anthropology, Leipzig, Germany)

The aim of the two studies was to evaluate how productive a 2-year-old Polish child's use of noun-inflections is. Schema-based approach predicts that early use of inflections is lexically highly restricted, whereas rule-governed inflections should be used unrestrictedly from the onset. Study 1 compared lexical productivity of inflections in child and child-directed speech, thus avoiding arbitrary criteria of productivity. In Study 2, it was predicted that early use of schema-based inflections should also be restricted contextually, reflecting gradually emerging semantics of schemas, whereas rules should encode when given inflections can be used from the onset.

The studies used a densely sampled corpus of a girl recorded for 5 hours per week for 6 weeks, starting at 2;0.3. During this period she produced all noun-inflections occurring in the mother's speech; only 36 erroneous forms were sampled (0.64% of all tokens).

In Study 1, number of inflections per noun-type was counted in the child's and the mother's sample. To control differences in sample sizes, number of tokens in the mother's sample was randomly reduced to match that of the child. It was repeated 1000 times, each time comparing the distributions using Wilcoxon signed-rank exact test. To control differences in the lexicon, only noun-types shared by both speakers were included (672.95 on average). The mean z -score was 3.00, the mean p -value was below 0.006 ($SD=0.01$). The results suggest that the mother was likely to use the same noun in a greater number of inflections, which couldn't be explained by the child having fewer inflections available or producing fewer tokens.

Study 2 was concerned with how informative noun-inflections were about their context. It employed the Contextual Distinctiveness index, a relative-entropy measure indicating the difference between the distribution of context-items co-occurring with a target-item and their overall distribution. For each noun-form used by both speakers, its CD was calculated in both samples. The set of context-items consisted of all word-forms shared by the speakers, controlling differences in the lexicon. Because of the child's low MLU, the co-occurrence window was +/-1. Again, random reduction was repeated 1000 times, each time comparing the average difference in CD using paired-samples t -test. The mean t -score was 4.32 (mean $df=760.85$), and the mean p -value was below 0.0001 ($SD<0.0003$). The results show that context of a noun-form was significantly easier to predict in the child's speech, which can't be explained by a smaller lexicon, suggesting that the child's use of inflections is contextually restricted.

Both studies, using rigorous methods, reveal that a 2-year-old child's use of noun inflections, despite appearing adult-like at first, is highly restricted, both lexically and contextually, much in line with the schema-based approach to morphology.