Jeffrey Lidz and Sandra Waxman, Northwestern University

One-substitution Revisited: Experimental and Corpus Evidence

One of the classic illustrations of the poverty of the stimulus argument involves the acquisition of anaphoric "one". In short, the argument shows that learners must know in advance of experience that NPs have internal structure. The consequences of this logical argument for learners have been examined experimentally with 5-year-old children (Hamburger and Crain 1984). However, arguments for innate syntactic structure made on the basis of 5-year-olds are not fully convincing. Better evidence for the postulation of structure in the absence of relevant input would come from children at the earliest stages of grammatical development. In this paper, we show that 18-month-old children know that "one" is anaphoric to phrasal categories and thus that at the earliest stages of syntactic development, children know that NPs have internal structure. In addition, an analysis of the input to children shows that the evidence that would unambiguously lead children to a grammar with hierarchical structure is absent.

Consider the two hypotheses for the structure of NP given in (1).

(1) a. Flat structure hypothesis b. Nested structure hypothesis

We know, on the basis of anaphoric substitution, that for adults (1b) is the correct representation (Baker 1978). In (2), the element *one* refers anaphorically to the constituent [*red ball*].

(2) I'll play with this red ball and you can play with that one.

Since anaphoric elements substitute only for constituents and since it is only under the nested structure hypothesis that the string *red ball* is represented as a constituent, it follows that (1b) is the correct structure.

How children acquire this knowledge is more mysterious (Hornstein and Lightfoot 1981). Suppose that a learner is exposed to small discourses like (2) in which *one* is anaphoric to a previously mentioned discourse entity and that the learner has recognized that *one* is anaphoric. Suppose further that the learner has hypothesized a flat-structure grammar. Disconfirming evidence for this hypothesis would be extremely hard to come by because every situation that makes *one* = [*red ball*] true also makes *one* = [*ball*] true. The right kind of evidence might be a situation in which (3) is uttered and Max has a blue ball.

(3) Chris has a red ball but Max doesn't have one.

In such a situation, the learner who posited a flat structure in which *one* was anaphoric to *ball* would have to conclude that he had built the wrong grammar and thus be led to change the hypothesis. In order for learners to consistently build the correct grammar, such situations would have to be common enough for them to show up at levels distinguishable from noise in the linguistic environment. Since such situations are not likely to be so common, we conclude that the flat structure hypothesis could not be part of the hypothesis space of the learner.

The logic of the argument is based on the crucial assumption that the evidence that unambiguously supports the nested structure hypothesis does not occur often enough to impact learning. In addition, because it is an argument based on what adults know about their language, it is missing the important step of showing that at the earliest stages of syntactic acquisition, infants know that *one* is anaphoric to a phrasal category.

To examine the empirical status of the assumption that the evidence that *one* is anaphoric to phrasal categories is unavailable to learners, we examined the parental speech to the children Adam and Nina in the CHILDES database (MacWhinney 2000). Of the roughly 55,000 adult utterances in these corpora, there were 792 anaphoric uses of *one*. We examined these cases to determine whether the antecedent was unambiguously phrasal. If in a high proportion of cases, *one* referred unambiguously to a phrasal category, then it is possible that infants might be able to learn about the syntactic properties of *one* from the input. This was not the case. In 95% of the cases, the antecedent had insufficient internal structure to be informative. 4.3% of the cases were ambiguous; 0.5% contained an ungrammatical use of *one*, which would constitute noise on either hypothesis; and, only 0.2% occurred in a context that unambiguously indicated that *one* is anaphoric to a phrasal category. In sum, the vast majority of the anaphoric uses of *one* are syntactically uninformative; and, the data that the child would need in order to learn the syntax of *one* occurs at a rate that is indistinguishable from noise. Consequently, we can conclude that if children have command over the syntax of anaphoric *one*, this syntactic knowledge could not have come from the input.

To test infants' knowledge of the internal structure of NP, we conducted a preferential looking study with 18-month-olds. Each infant participated in 4 trials, each consisting of two phases. During the familiarization phase, an image of a single object (e.g., a yellow bottle) was presented three times accompanied by a voice naming the object with a Det-Adj-N sequence (e.g., "Look! A yellow bottle."). During the test phase, two objects appeared on opposite sides of the television monitor (e.g., a yellow bottle and a blue bottle). Both objects were from the same category as the familiarization object, but only one was the same color. Infants were assigned to one of two conditions differing only in the linguistic stimulus. In the control condition, subjects heard a neutral phrase ("Now look. What do you see now?"). In the anaphoric condition, subjects heard a phrase containing the anaphoric expression *one* ("Now look. Do you see another one?").

Our predictions were as follows. In the control condition, where the linguistic stimulus does not favor one image over the other, we expected infants to prefer the novel image (the blue bottle), as compared to the now-familiar image (the yellow bottle). In the anaphoric condition, infants' performance should reveal their representation of the NP. If infants represent the NP with a flat structure, and therefore interpret *one* as anaphoric to the category N°, then both images would be potential referents of the noun (bottle). In this case, the linguistic stimulus is uninformative with regard to the test images, and so infants should reveal the same pattern as in the Control condition. However, if infants represent the NP with a nested structure, and interpret *one* as anaphoric to the intermediate constituent (yellow bottle), then they should reveal a preference for the (only) image that is picked out by that constituent (the yellow bottle).

The predictions of the nested structure hypothesis were borne out. Subjects in the control condition devoted more attention to the novel object ($p<.01$). Subjects in the anaphoric condition devoted more attention to the familiar object ($p<.0008$). In addition, looking time to the familiar object was greater in the anaphoric condition than in the control condition ($p<.003$). These results show that, despite the unavailability of relevant input, 18-month-old infants interpret *one* as anaphoric to a phrasal category and thus that they represent the NP as having internal structure.