Chapter 26
Backward Anaphora

Consider the following sentence: *He thinks the Troll is the best jumper*. Every adult speaker of English knows that the pronoun *he* in this sentence does not refer to the Troll; its referent must be someone else. Linguists explain this using one of the principles of the binding theory, that component of Universal Grammar concerned with anaphoric relations among NPs. The principle that determines that a name (e.g., *the Troll*) cannot be the referent of a pronoun in certain structural configurations is Principle C. Since Principle C is a negative statement about illicit anaphoric relations, it is a constraint on meaning: (sentence, *meaning*). The sentence itself, *He thinks the Troll is the best jumper*, is perfectly acceptable; it simply cannot receive a particular interpretation. Because Principle C is a constraint, it is a viable candidate to be viewed as an innately specified aspect of linguistic knowledge. As a result, it is anticipated to be a linguistic universal, not mastered by children on the basis of their linguistic experience. Granting these assumptions, we are led to ask whether Principle C shows up early in the course of language development. The truth value judgment task has proven to be a useful technique to address this question. In this chapter, we review the findings of previous experimental investigations of the acquisition of Principle C. First, some theoretical background.

26.1 Theoretical Background

Principle C governs the anaphoric relations between pronouns and referring expressions (e.g., names and definite descriptions). We observed that in some constructions, a referring expression (r-expression) cannot corefer with a pronoun, as in (1).

(1) He thinks the Troll is the best jumper.

Here, the pronoun *he* cannot refer to the Troll; it must refer to someone else—Cookie Monster, perhaps. However, if the order of the pronoun and the name is switched, as in (2), then coreference between them is permitted. In (2), the pronoun can be interpreted as referring to the Troll. There is also another reading in which it refers to someone else—again, perhaps Cookie Monster.

(2) The Troll thinks he is the best jumper.

Principle C explains why (2) tolerates an interpretation that (1) does not. Without considering further evidence, instead of Principle C, we might propose the following hypothesis to explain the facts discussed so far:

(3) If a pronoun precedes a referential NP (an r-expression), then they cannot refer to the same thing.
In symbols, we can represent this as the following, linear prohibition:

\[(4) \quad \ast \text{pro}_1 \ldots \text{NP}_1\]

In this representation, the fact that the two NPs (\(\text{pro}\) and \(\text{NP}\)) have the same subscript indicates that they have the same reference. As noted in chapter 20, structure-independent hypotheses like this, which treat sentences as strings of words, are not characteristic of the theory of Universal Grammar. It is therefore not surprising to find evidence that this hypothesis is incorrect. The following example will suffice:

\[(5) \quad \text{While he}_i \text{ was reading the paper, the Troll}_i \text{ ate a bagel.}\]

In this sentence, the pronoun \(\text{he}\) can refer to the Troll. Example (5) therefore shows that in some cases a pronoun can precede a name and still be coreferential with it.

The difference between (1), on the one hand, and (3) and (5), on the other, is structural. The structural relation that determines coreference is called \(c\)-command, and the constraint that invokes \(c\)-command is called Principle C. For our purposes, the following definition of Principle C will suffice:

\[(6) \quad \text{Principle C}\]

An r-expression \(R\) cannot be coreferential with a pronoun \(P\) that \(c\)-commands it.\(^1\)

As (6) makes clear, our concern is with r-expressions that are \(c\)-commanded by NPs that are pronouns. At a coarse level of description, a pronoun \(c\)-commands an r-expression if the pronoun is positioned higher than the r-expression in the phrase structure analysis (tree diagram) of the sentence. A more precise definition is that a pronoun \(P\) \(c\)-commands an r-expression \(R\) in a phrase marker if and only if there is a path beginning at \(P\), proceeding upward to the first branching node above \(P\) and then extending downward to \(R\). Thus, \(P\) \(c\)-commands \(R\) in (7) but not in (8).

\[(7)\]

\[\text{CP} \]

\[\text{IP} \]

\[\text{P} \quad \text{VP} \]

\[\text{he}_i \quad \text{thinks} \quad \text{CP} \]

\[\text{IP} \]

\[\text{R} \quad \text{the Troll}_i \quad \text{VP} \]

\[\text{is the best jumper}\]
Since Principle C is a constraint within the theory of Universal Grammar, it is expected to exhibit all of the hallmarks of innate specification. It should be universal; it should emerge in the absence of linguistic experience; all children should abide by it, regardless of differences in primary linguistic data and despite its apparent (In the indexing notation used here, which replaces the underlining notation used in previous chapters, two NPs that have different referents are given different indices \((i,j,k, \ldots)\), and two NPs that have the same referent are given the same index. The \(\text{Troll}_{i/j} \) and \(\text{he}_i\) in (7) means that the Troll and he cannot refer to the same individual \((*i,i)\) but can refer to different individuals \((i,j)\).

When a pronoun (e.g., he in (7)) c-commands an r-expression (e.g., the Troll), coreference is not permitted; on the noncoreferential interpretation, which is permitted, these expressions have different indices \((i,j)\), the result of applying Principle C. By contrast, Principle C does not apply to the structure in (8) because the pronoun does not c-command the r-expression; therefore, coreference is possible. Principle C is a negative statement; it states when coreference is not permitted. As another example, notice that in (5), While he was reading the paper, the Troll ate a bagel, the pronoun does not c-command the r-expression, because it is deeply embedded inside a subordinate clause, as shown in (9). Therefore, Principle C does not apply, and coreference is permitted between the pronoun and the r-expression.

Since Principle C is a constraint within the theory of Universal Grammar, it is expected to exhibit all of the hallmarks of innate specification. It should be universal; it should emerge in the absence of linguistic experience; all children should abide by it, regardless of differences in primary linguistic data and despite its apparent
complexity (pretheoretically speaking); it may appear early in the course of language development. The experiments reviewed in the rest of the chapter were concerned with the hallmark of early emergence. As we will show, several early studies of Principle C indicated that it did not bear this hallmark. Fortunately, a subsequent study using the truth value judgment task did confirm the predictions of the theory of Universal Grammar, namely, that children’s grammars do include Principle C.

26.2 Previous Research

There is a large body of research on children’s knowledge of coreference relations between pronouns and r-expressions. The findings of the initial studies of children’s knowledge of Principle C were somewhat disconcerting. Researchers were led to conclude that, at least initially, children apply a purely linear (structure-independent) hypothesis in interpreting pronouns and r-expressions. Summarizing one such experiment, by Lust (1981), Solan (1983) concludes that children’s initial hypothesis about anaphora is to prohibit coreference between a pronoun and an NP whenever the pronoun appears first. This conclusion was based on an act-out study by Tavakolian (1978) in which children were instructed to act out the meanings of sentences using toy figures and other props placed in the experimental workspace. Tavakolian found that two-thirds of the 3- to 5-year-old subjects responded in a manner that, according to Solan, indicated that children’s grammars include a linear prohibition against backward coreference (cf. the directionality factor of Lust, Eisele, and Mazuka 1992).

In the figure manipulation experiment by Tavakolian, the majority of the time children selected an animal that was not mentioned in the sentence, but was present in the workspace, as the referent of the pronoun in sentences like (10a) and (10b).

(10) a. For him to kiss the lion would make the duck happy.
   b. That he kissed the lion made the duck happy.

Of the 24 subjects, 14 consistently acted out these sentences in this fashion. Solan concludes from this that “children use direction rather than structural principles in restricting anaphora, … never allowing backward anaphora” (pp. 83–84).

This conclusion is unwarranted, for several reasons. First of all, since one-third of the subjects’ responses indicated acceptance of backward anaphora, the statement that children never allow backward anaphora is too strong. But suppose that every child had chosen an unnamed referent on every opportunity. Even this would not be evidence of a prohibition against backward anaphora. The possibility exists that there is a strong preference for the deictic interpretation of the pronoun (according to which the pronoun refers to someone who is not mentioned in the sentence) over the backward anaphora interpretation (according to which the pronoun and an r-expression within the sentence are anaphorically linked).²

It seems more likely that the apparent lack of availability of backward anaphora was simply the result of children’s preference to assign a discourse referent to a pronoun as quickly as possible, without waiting for linguistic information about the intended referent. This kind of preference, which conserves memory resources, should also show up in adult sentence processing, according to the Modularity Matching Model. The pressure to interpret constituents, as rapidly as possible will be encountered by children (and adults) on every trial; therefore, we would expect to see a
strong bias against backward anaphora in an act-out task. By contrast, as we will show, the truth value judgment task can be used to boost the availability of the backward anaphora interpretation. This is one of its virtues, as compared to other tasks. It is important to note that if children actually have a linear prohibition against backward anaphora, then we cannot test their knowledge of Principle C using sentences like (1). To provide evidence of children's knowledge of Principle C, we must show two things: (a) that children allow backward anaphora wherever adults allow it and (b) that children disallow backward anaphora when it is excluded by Principle C.

26.3 A Positive Finding

The question of children's knowledge of the backward anaphora interpretation of sentences was pursued in an experiment by Crain and McKee (1985). In this experiment, children encountered sentences like (11) in circumstances appropriate to the backward anaphora (or intersentential) interpretation and the deictic (or extrasentential) interpretation.

(11) While he was dancing, the Ninja Turtle ate pizza.

The experimental procedure was the truth value judgment task. Ambiguous sentences such as (11) or its equivalent were presented on two separate occasions, in two contexts. Following both situations, the puppet, Kermit the Frog, uttered the sentence in (11). In one context for (11), the Ninja Turtle was dancing and eating pizza. This was the situation appropriate for the backward anaphora reading. Because the backward anaphora interpretation was dispreferred in previous experiments, children's knowledge of this interpretation was tested in contexts in which it was true. In the other context presented to children, someone else was dancing while the Ninja Turtle was eating pizza. The context also made (11) true. Presented in this context, the sentence was a control to check whether children would accept the deictic interpretation of the pronoun more often than the backward anaphora interpretation. Since both interpretations were presented in contexts that made the test sentences true, children's preference for one or the other interpretation could readily be assessed.

The results were that the 62 children tested (mean age 4;2) accepted the backward anaphora reading 73% of the time, in appropriate contexts. The extrasentential reading was accepted only slightly more often, 81% of the time. Only 1 of the 62 children interviewed in this way consistently rejected backward anaphora. On the basis of these findings, we can conclude that children's grammars allow backward anaphora, just as adults' grammars do. By boosting the availability of the backward anaphora reading, the truth value judgment task allowed Crain and McKee to demonstrate the existence of a reading that did not surface readily in the act-out task where children were forced to choose between interpretations.

In this study, knowledge of Principle C was tested using both one-clause sentences like "He washed Luke Skywalker" and two-clause sentences like "He ate the hamburger when the Smurf was inside the fence." It was found that children rejected coreference between the r-expression and the pronoun 90% of the time for the one-clause sentences and 84% of the time for the more complex two-clause sentences. This finding underscores our contention, stated in chapter 3, that neither children nor adults find it more difficult to comprehend sentences with two clauses than sentences with a single clause.
The findings support the view that children do not rely solely on their linguistic experience in making judgments about the appropriate mappings of sentences with their meanings. Presumably, there is nothing in children's experience to tell them that certain sentence/meaning pairs are not allowed; therefore, they have no way to learn the structural constraint prohibiting coreference, Principle C. Nevertheless, they appear to know it at an early age. By the logic underlying the poverty-of-the-stimulus argument, we are led to conclude that children's knowledge of Principle C is innately specified.