SMITH COLLEGE

Smith ScholarWorks

Philosophy: Faculty Publications

Philosophy

2021

# A Touchy Subject: Optimality and Coreference

Jill de Villiers
*Smith College*, jdevilli@smith.edu

Jacqueline Cahillane
*University of Massachusetts Amherst*

Emily Altreuter
*Smith College*

Follow this and additional works at: https://scholarworks.smith.edu/phi_facpubs

 Part of the Linguistics Commons, Philosophy Commons, and the Psychology Commons

## Recommended Citation

de Villiers, Jill; Cahillane, Jacqueline; and Altreuter, Emily, "A Touchy Subject: Optimality and Coreference" (2021). Philosophy: Faculty Publications, Smith College, Northampton, MA.
https://scholarworks.smith.edu/phi_facpubs/49

This Article has been accepted for inclusion in Philosophy: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

# A Touchy Subject: Optimality and Coreference

**Jill G. de Villiers,**
**Smith College**

**Jacqueline Cahillane,**
**University of Massachusetts, Amherst**

**Emily Altreuter,**
**Smith College**

**Abstract**

Four studies are reported that compare production and comprehension of structures involving Principle A and B with 68 English speaking children. The stimuli included simple and complex sentences combined with simple and quantified NPs, each with reflexives and pronouns. A novel technique using a laptop proved successful for eliciting stimulus descriptions as well as truth value judgment. The results test a recent Optimality account of binding by Hendriks and Spenader (2004), but it is argued that more constraints are needed. Although the data can be fit well by the constraints, questions remain about whether it is theoretically satisfactory.

**Introduction**

A long-standing problem in child language research has been to account for the failure of children to reject the coreference of a pronoun with an antecedent noun in a sentence like 1):

1) Papa Bear hit him.

Many different studies have shown that children as late as age six treat the pronoun as possibly coreferent with Papa Bear, in apparent violation of Principle B of the binding principles. The landmark study was that of Chien and Wexler (1990) who showed in an extensive range of tests that children *did* obey Principle B in the case of a quantified NP:

2) Every bear hit him.

suggesting that the failure lay elsewhere than knowledge of grammatical principles. They argued that variable binding and coreference of a pronoun followed two different principles, with binding being absolute, but establishing coreference being dependent on context (Reinhart, 1983). Since quantified NPs cannot be antecedents for coreference, that interpretation of the pronoun is ruled out in 2), but 1) remains open to a coreferent interpretation of the pronoun for children. Thornton and Wexler (1999) analyzed data to show that in the case of 1), young children perform at chance, patterning their answers according to a binomial distribution.

Chien and Wexler (1990) and Grodzinsky and Reinhart (1993) explained the child's failure in different terms. Chien and Wexler argued that children lacked knowledge of the proper contextual considerations to interpret the pronoun in 1), a factor they called Principle P. When this matured, then they could perform as adults. Reinhart (1983) argued for an inference, Rule I, that was based on a decision procedure between two competing representations, more recently termed "reference set computation" (Reinhart, in press). The idea of Rule I is that in deriving a representation, another representation with a bound variable should be constructed and compared in meaning to the first in context. If they are not equivalent, then a coreference interpretation is permitted. Grodzinsky and Reinhart attribute the failure of children in these tasks to a computational difficulty, namely, they do not have the processing resources to compute Rule I. Thornton and Wexler dismiss the processing limitation account, but retain the idea that there is a pragmatic principle captured in Rule I that children do not know. The difficulty that Reinhart (in press) points out lies in the learnability puzzle that then emerges. For Reinhart, processing is the only thing that matures, but all the linguistic knowledge is innate. Thorton and Wexler must postulate that the child acquires knowledge of the right pragmatic conditions for coreference from experience.

What are the implications for production of sentences such as 1)? Surprisingly little attention has been paid to this problem. In reviews on the subject only a brief mention is made that children produce pronouns and reflexives readily and accurately (from the perspective of binding) at a young age. Only one study (Bloom, Barss, Nichol & Conway,1994) examined corpora of child speech in detail, but found rather few *third person* cases on which to base an analysis. Nevertheless, there did not appear to be Principle B errors in production. Theoretically, the above theories might differ on the matter of production. Reinhart could argue that no reference set computation is involved in production, because the child knows the principles innately and never compares alternatives in deriving a production, just in processing. Thornton and Wexler have a more challenging task, since the child surely needs the information from contexts to

acquire the missing pragmatic knowledge that would seem as necessary in production as in comprehension.

In the study that follows, the original plan was simply to explore elicited production of the same forms in production as in comprehension, to expand the data available for theorizing in this domain. However, a new theoretical idea emerged as we were collecting the data, in a paper by Hendriks and Spenader (2004). This paper provides a new account of the difference in children's processing of sentences like 1), using an Optimality framework. But in addition, it explicitly addresses the (presupposed) disparity between production and comprehension of such sentences and provides an explanation for it in terms of Optimality.

Hendriks and Spenader also acknowledge a sharp distinction between the phenomena under Principle A and Principle B, but in a different way than the theories above. They adopt Burzio's (1998) "soft constraints" approach that uses only Principle A with other constraints to explain the distribution of pronouns, not Principle B at all. Generalizing beyond a principle that Burzio calls "Referential Economy", they postulate that there is a constraint such that referentially "heavy" items are to be reduced whenever possible. So, pronouns are preferred to nouns and reflexives are less costly than pronouns. Hendriks and Spenader set these in an Optimality tableau, in which Principle A outranks Referential Economy, and the latter is only a constraint on forms, not on interpretations. Tableau 3) represents production of a sentence in a context in which Papa bear is touching himself:

(3) Hendriks and Spenader: Reflexive Production

(☞ indicates a winning candidate, * a violation, !* the most serious violation)

| | /Coreferential meaning/ | PRINCIPLE A | REFERENTIAL ECONOMY |
|---|---|---|---|
| a. ☞ | Reflexive | | |
| b. | Pronoun | | *! |

In this case, the pronoun is ruled out for reasons of referential economy, so the reflexive wins. If Papa Bear is touching someone else, then tableau 4) pertains:

(4) Hendriks and Spenader: Pronoun Production

| | / Disjoint meaning/ | PRINCIPLE A | REFERENTIAL ECONOMY |
|---|---|---|---|
| a. | Reflexive | *! | |
| b. ☞ | Pronoun | | * |

The reflexive is ruled out by Principle A, and the pronoun is ruled out by referential economy, but since the latter is a weaker constraint, the pronoun wins. In Comprehension, the input is a linguistic form, and the choice is of *interpretations* as in 5).

(5) Hendriks and Spenader: Reflexive Comprehension

| / Reflexive / | | PRINCIPLE A | REFERENTIAL ECONOMY |
|---|---|---|---|
| a. | Disjoint | *! | |
| b. ☞ | Coreferential | | |

Now reflexives are easily understood, because Principle A rules out a disjoint interpretation. But pronouns have in their account in 6) no constraints limiting their interpretation:

(6) Hendriks and Spenader: Pronoun Comprehension

| / Pronoun / | | PRINCIPLE A | REFERENTIAL ECONOMY |
|---|---|---|---|
| a. | Disjoint | | |
| b. | Coreferential | | |

The prediction is that children should choose at 50/50 under such a condition, and they do according to some studies (Thornton & Wexler, 1999).

But how, on Hendriks and Spenader's account, do adults ever get the right interpretation of the pronoun in 1)? They argue that adults (and older children) are able to carry out a second round of optimality analyses, in which they find the optimal *combination* of form and meaning. In the first pass, because of Principle A, the reflexive and coreferential meaning is one such optimal pair. But that choice then rules out any new pair involving either reflexives or coreferential meaning, leaving pronoun and disjoint meaning as the default second optimal pair. Hendriks and Spenader argue that the child is incapable of this bidirectional optimization. This could be either because of processing limitations, or because the optimization requires considering what is best for listeners as well as speakers, and this is a pragmatic skill that is still developing at age five or six. They write:

> "A child must, when hearing a pronoun, reason about what other non-expressed forms are associated with the potential interpretations of pronouns, realize that a coreferential meaning is better expressed with a reflexive, and then by a process of elimination realize that because this potential meaning is already better expressed with a reflexive, the pronoun should be interpreted as disjunct."

Hendriks and Spenader provide an interesting account that simultaneously handles the issue of the child's difference from adults and the disparity in comprehension and production. They acknowledge that other constraints may be needed, such as a feature-agreement requirement that matches the features on nouns with their anaphors, but usually experimental studies have controlled these aspects for the sake of ambiguity so the effects of the constraint can not be seen (butt see Vasic, Avrutin & Zuckerman, 2004). In the case of a quantified NP, Hendriks and Spenader would invoke a constraint

of “Referential Antecedent”, essentially ruling out coreference between a pronoun and a quantified NP because the quantified NP is not referential (Reinhart, 1983). However, they do not provide in their paper any tableaux for this, nor do they suggest where the constraint might rank.

The opportunity arises in the current study to test these theoretical predictions against an empirical data set collected to test and compare the performance of children in production and comprehension of the same stimuli. Not by design, since the study was begun before Hendriks and Spenader wrote their paper, the experiment has some features that give us the opportunity to stretch their analysis in two new directions, namely, to how children treat quantified NPs and to binding in two clause sentences. As we will show, this is quite fortuitous. However, we remain entirely agnostic about the bidirectional optimization hypothesis, having no data to speak to what might allow children success in the end.

**Design**

Four different conditions were designed to test children’s production and comprehension in parallel. These were:

A. Classic Chien & Wexler (1990):
   Here is Baby Bear and Papa Bear. (Order varied)
   Baby Bear is washing him/himself.
B. Two-clause (Solan,1987; Jacubowitz. 1987)
   Papa Bear says Baby Bear is washing him/himself.
C. Quantified Chien & Wexler (1990)
   Here is Big Bird and all the bears. (Order varied)
   Every bear is washing him/himself.
D. Quantified two-clause
   Big Bird says every bear is washing him/himself.

An example picture for each scenario type is presented in Figure 1. The characters were all familiar ones to children of this age, and the three verbs used were *touch, wash* and *point to*.

Subjects

Subjects were 68 children, average age 6;2 years, range 4;6 to 7;2. All were normally developing and had no speech or hearing deficits, and English was their first language. There were 37 girls and 31 boys.

Each child received only one of conditions A-D in comprehension, then again in production, always in that order because we felt we needed to warm them up to attempt the speaking task. Usually the two tasks were done in a single session, sometimes with other tasks interspersed, and rarely, in two sessions a few days apart.

The numbers of subjects in each condition were as follows:

| Condition | Number of subjects | Mean age |
|---|---|---|

| | | |
|---|---|---|
| A | N=19 | 6;3 |
| B | N=18 | 6;3 |
| C | N=15 | 5;11 |
| D | N=16 | 5;7 |

**Procedure**

Comprehension

To make the stimuli and presentation as constant as possible, we used a laptop computer to present them together with recorded narration in Powerpoint. Each child was brought individually from their classroom to a quiet space where the experimenter let them sit at the computer. A video-camera was positioned behind the child and recorded all responses for later checking. The child was told the computer would present a picture, then a sentence would be heard about it. If the child thought the sentence was the right one to use about the picture, the child should say "yes" or push a green smiley face in the right top section of the screen. If the child thought the sentence was not the right one about the picture, then the answer would be "no", or pressing the red frowning face at the bottom left of the screen. No child had any problem following these instructions, and usually spoke and pressed a face simultaneously. To keep attention and interest at a high level, only a small number of stimuli were presented, (12 to 15), with both pronouns and reflexives in each condition, in semi-random orders to avoid creating set responses.[1]

Production

It was in Production that the use of the laptop proved most beneficial. Sometimes children are shy about full productions of picture description, but we used the following subterfuge:

> "Now you have seen the pictures, we think it would be much better with a kid's voice, don't you? Do you think you can do the right sentences for us? We'll record your voice on the computer, so just speak nice and loudly and this little mike will pick it up. Just to remind you, here are some good examples of what we want you to say".

The child then heard eight new forms, all correct, with four reflexives and four pronouns. Following that, we opened a new Powerpoint, reminded the child of the form to use, by saying e.g. "See, you'll want to say something like "every bear…." and then tell me what they say", and began recording narration. The child proceeded at his or her own pace, and

[1] One important caveat is in order about the data. Because when we began the study, we believed that the case of 1) was the most theoretically significant, namely the *false* pronoun case, in conditions A and B we included only false pronoun examples, and to balance it by required answer, we used true reflexive cases. In conditions C and D, which were run subsequently, we added true pronouns and false reflexives for a balanced design. However, in the present analyses we restrict our attention to the cases in common across the comprehension of all conditions, namely false pronouns, and true reflexives. Fortunately, truth makes little difference to the high level of performance on reflexives (Chien & Wexler, 1990). Neither is it the case that the makeup of the set created different results, because we were able to compare the performance on false pronouns in this study to the performance of an earlier test group of 36 children who had been tested with the same stimuli within a design that included all types. Their performance on the false pronoun case is discussed in the results for comparison.

occasionally needed reminding e.g. “First say “here is…”, or “start with “Big Bird says…”. In the production Powerpoint there were ten new scenarios made by creating combinations of old stimuli, that is, the same characters and the same verb types: *wash*, *touch* and *point to*.

As a side comment, two aspects of this procedure persuade us that it is worth adopting more widely. First, the child’s verbal responses are recorded with the stimuli they are describing and with great fidelity compared to the usual video microphone. Second, the children at this age were very excited by the computer and most of them made great efforts to speak clearly and succinctly “for the computer”, ignoring the experimenter. This meant that there was much less need for the usual wrangling with a child to use language as specific as you need, especially in the pragmatically odd testing circumstance when a picture is right in front of both of you.

**Results**

The results fall into two categories: very general ANOVAs are presented first to deal with the data in broad strokes, but then the particular predictions of Optimality tableaux are presented together with the particular statistical tests of them.

Before the other analyses were run, an ANOVA tested age as a covariate on the children’s comprehension of pronouns, with clause (one versus two) and quantifier (none versus *every*) as the between-subject variables. Age was added as a covariate in this ANOVA because the mean age varied slightly across the conditions. Since there was no effect of age in this ANOVA, it was not used in the remainder of the analyses.

A repeated measure ANOVA was conducted with the percentage of correct responses as the dependent measure and the within-subject variables as task (production versus comprehension) and proform (pronoun versus reflexive), and the between subject variables being quantifier (none or *every*) and clause number (one versus two).

The results show first that there is a massive effect of task, namely production is significantly better than comprehension in all four experiments ($F(1,54) = 113.57$, $p < .000$, $eta^2 = .678$. But task also hugely interacts with the form used ($F(1,54) = 144.8$, $p < .000$, $eta^2 = .728$) because it is for pronouns that the task difference is most dramatic. Pronoun *comprehension* is better with quantified NPs, resulting in a three way interaction between task, proform and quantifier ($F(1,54) = 23.16$, $p < .000$, $eta^2 = .300$). Neither clause number nor quantifier is significant as a main effect on all performance. The complex results are presented in Table 1, but the details of what is happening are inevitably obscured in such an omnibus test. Mean percentages across conditions are presented in Table 2.

Before continuing, it is necessary to consider whether the scoring of the production results might have inflated these differences, and whether they should be adjusted to correspond more to the comprehension data. There are two problems we faced, which later turn out to be important factors in revising the Optimality tableaux:

a) In conditions A and C, namely the one clause sentences with an introducing clause, children in comprehension frequently asked “who’s ‘him’”?, as if they couldn’t access the proper antecedent. In production, they often avoided producing pronouns and used proper names. We did not count those responses, but it meant we lost several subjects from the ANOVAs as a result

of missing data. So on a second pass, we recoded these as production "errors", to see if the difference between the production and comprehension was an artifact of our generosity in coding.

b) In conditions C and D, namely the quantifier cases, children were reluctant to use the form "himself" bound to "every N", instead often saying "himselves", "themselves", "theirselves", "theirself", or "themselves". In other words, they believed it needed a plural feature. We coded these as correct productions in the first analyses, but as mistakes in the second analyses, again to see if the task difference would disappear.

Table 3 shows the analysis redone, and Table 4 the new means. The result of counting the plural reflexives as errors means that the *overall* task advantage of production is now washed out, and reflexive production looks poor in comparison to comprehension. However, a big difference remains for *pronoun* production over comprehension, varying widely across conditions.

Clearly, it is time for some more refined analyses, so let us turn to Part 2 of the results, in which some new Optimality tableaux are introduced and their particular predictions tested against the data.

**Part 2.**

In this section, we present some adaptations of the Hendriks and Spenader Optimality tableaux, driven by the general empirical observations made in carrying out these tests with children. First, the general assumptions are laid out, then each case is considered for each condition and each task.

It must be pointed out that what follows may not be a correct account for all extant and possible studies of binding. One of the promises of this approach is that it might be modifiable to predict performance when individual tasks and sentences change e.g., when the antecedent of a pronoun is made more salient, or when person or number features are made to match or mismatch. Since the model is successful for these four studies, it may hold promise as a framework for the analysis of other studies but we would expect the results to change as the variables do.

Assume the following constraints from Hendriks and Spenader:

1. Principle A (for reflexives).
2. Referential Economy: Nouns< pronouns<reflexives
3. Referential Antecedent (i.e. not a quantified NP for a pronoun)

To these we add two more constraints:

4. Referential Salience: the antecedent of a pronoun must be *salient*, i.e. readily available in the discourse. Notice that in the conditions A and C, children seemed bewildered by who "him" was, as if they had lost track of the referents in the discourse. For a pronoun, in the absence of salience, Referential Economy is cancelled and a noun is used for a pronoun. So Referential Salience outranks Referential Economy. It may also outrank Referential Antecedent. This idea resembles other proposals linking Optimality

constraints to pragmatics and discourse, such as Beaver (2004). The most immediate resemblance is to a proposal in Bouma (2003) of a constraint called SENTENTIAL PROXIMITY, i.e. “Don’t have a sentence boundary between the pronoun and the antecedent.”

5. Feature Matching. The anaphor or pronoun must match its features (number, person) to the antecedent. This resembles proposals in Bouma (2003 for matching agreement features on antecedents. This not usually a significant issue in binding studies, but in the quantifier case, children might believe a quantified NP such as “every bear” requires a plural reflexive. This seems to be only a weak constraint in the present study due to children’s misassumptions about “every N”, but it is possible that more definite plurals or different persons would change the ranking of this constraint (see also footnote in Hendriks and Spenader (2004), where they assume it is highly ranked but usually irrelevant in the studies).

We postulate the following rank ordering of the constraints:

> Principle A> Referential Salience> Referential Antecedent> Referential Economy>Feature matching.

Now tableaux for each particular condition are considered for production, then comprehension, each for reflexives and then for pronouns. In the titles, scenarios are represented in < >, sentences in quotes. We present the tableaux in a sequence corresponding to the order of our conditions: A,B,C.D., first the non-quantified cases, then the quantified.

To facilitate discussion, we number all results in {parentheses} and asterisk { *} those that do not agree with the predictions.

A.1.1.) <Big Bird touched himself> Production: 1 clause, no quantifiers, reflexive

| | /Co-reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. | Reflexive | | | | | | 96% |
| b. | Pronoun | | | | *! | | 3% |
| c. | Proper N | | | | *! | | 1% |

Since referential economy counts against the pronoun, there is an easy reflexive answer, which is strongly confirmed by the results. Only one noun was used[2] {1}

A.1.2) <Big Bird touched him> Production: 1 clause, no quantifiers, pronoun.

[2] Answers that fell into none of these categories such as “his head” were not considered in the percentages. They represent about 5% of the total. The form “hisself” was very common and considered as a variant of “himself”.

| /Disjoint reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Reflexive | *! | | | | | 0% |
| b. ☜ Pronoun ? | | * | | * | | 38% |
| c. ☜ ?Proper N | | | | * | | 62% |

The pronoun wins over the reflexive, because the reflexive is a violation of Principle A, but the choice is not a happy one because it violates both Referential Economy and the Referential Salience constraints. The child has to say, "Here are Big Bird and Grover. Big Bird is hitting….". The pronoun performance is predicted to be impaired by the lack of salience of the discourse antecedent, in which case Referential Economy is overridden and a noun is produced instead. So the proper noun should be a strong competitor. It is not clear how to rank pronouns and nouns, since the violation of referential economy is presumably not the same in the two cases: Nouns are worse than pronouns.

In fact, pronouns are only produced at 38%, and nouns at 62%, with 0% reflexives, as predicted {2}. Pronoun production is markedly lower than the production of reflexives (96%) in A.1.1, which have no constraints against them. ($t(18) = -4.989$, $p<000$) {3}.

A. 2.1. "Here is Big Bird and Grover". Big Bird touched himself" Comprehension: 1 clause, no quantifiers: reflexive

| /Reflexive/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Disjoint reference | *! | | | | | 4% |
| b. ☜ Co-reference | | | | | | 96% |

The coreference choice is predicted to win, and does so decisively {4}. Here we must note an important point about the methodology. Because the task is truth value judgment, and the child is merely saying "yes" or "no", we have to assume when the child disagrees with the sentence, that the opposite meaning (e.g. disjunction) *would* be chosen instead. We will return to this problem in the discussion.

A. 2.2. "Here is Big Bird and Grover". Big Bird touched him" Comprehension: 1 clause, no quantifiers: pronoun

| /Pronoun/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Disjoint reference | | *! | | | | 6.6% |
| b. ☜ Co-reference | | | | | | 93.4% |

Coreference wins, on this account, because of the lack of salience of the antecedent for the pronoun in the preceding clause. In other words, the child should get the pronoun

*systematically wrong,* which they do (6.6% correct judgments) {5}. By a One-sample t-test this percentage is significantly lower than chance (50%), so Referential Salience is a genuine factor. (t(18)= -16.73, p<.000)[3] {6}.

**The two clause cases:**

The effect of the two clause cases is to make much more salient the antecedent outside the clause: the speaker is highlighted and is considered the topic. So Referential Salience should be no longer a factor for pronouns. But does it slightly impair reflexives?

B.1.1. <Big Bird said Grover touched himself> Production: 2 clause, no quantifiers: reflexive

| /Co-reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Reflexive | | | | | | 85.4% |
| b. Pronoun | | | | *! | | 14.6% |

The reflexive wins, as predicted. The result should not be different from A.1.1. but it just misses being significantly different (F(1.35)=4.01, p=.051), a marginal fact for which we have no account unless referential salience has a weak negative effect on the reflexive{*7}

B.1.2. <Big Bird said Grover touched him> Production: 2 clause, no quantifiers: pronoun

| /Disjoint reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Reflexive | *! | | | | | 2.8% |
| b. Pronoun | | | | * | | 97.2 |

[3] This percentage is surprisingly low compared to other studies. Was it because we only had *false* pronouns in condition A? It is initially hard to see how this might work to increase the likelihood of saying "yes" to the false case, because the false pronouns provided the only opportunity to say "no". Several children turned to us and said, "I thought you said there'd be some wrong ones – these are all right!" In our previous unpublished study (alas, with no production) the 36 subjects aged 4 to 6 performed at a slightly higher level (31%), but this value was still significantly below chance of 50% (t (31)=-2.94, p<.006). It is a real possibility that in the presence of true pronoun cases, which the child generally answers correctly, the salience of the preceding discourse is brought into visibility for the *false* trials. That is, the saliency constraint is weakened. This would predict the difference we see between the studies (7% versus 31%).

The pronoun wins. Now the salience of the topic antecedent in the main clause helps. Comparing the result to the 1 clause case where salience was a problem confirms a massive difference in pronoun production $F(1,35)=38.822$, $p<.000$ {8}. Notice: nouns are never produced, as they would violate referential economy even more than pronouns do {9}.

B.2.1.. "Big Bird said Grover touched himself" Comprehension: 2 clause, no quantifiers: reflexive

| /Reflexive/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Disjoint reference | *! | | | | | 5.6% |
| b. ☜ Co-reference | | | | | | 94.4% |

Coreference is predicted to win, and does so decisively {10}. It is not significantly different from the one clause case (A.2.1), as predicted {11}.

B.2.2. "Big Bird said Grover touched him" Comprehension: 2 clause, no quantifiers: pronoun

| /Pronoun/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Disjoint reference | | | | | | 32.9% |
| b. Co-reference | | | | | | 67.1% |

Now that the disjoint referent is made salient, the result should be a 50/50 choice as no constraints are in operation. First, compare pronoun comprehension in this case to the case with 1 clause (A.2.1.), from which it should be significantly better, and it is ($F(1.35)=10.318$, $p<.003$) {12}. However, compared to chance (50%) it is marginally different by a One Sample t test ( $t(17)$ -2.15, $p<.05$ (33% not 50%)) {*13}. This might reflect an effect of distance between the pronoun and the topic, causing coreference to be preferred, since nothing *forbids* it. That is, it's a preference, but one that is not accounted for by the constraints we have proposed.

**Quantified NPs**

Now we consider the quantified NP variants of A and B. We predict that they introduce two new factors: Referential Antecedence, constraining coreference of pronouns with quantified NPs, and Feature Matching, influencing reflexives to the extent that the child believes "every N" requires a plural anaphor.

C.1.1. <Every bear touched himself> Production: 1 clause, quantifier, reflexive

| | /Co-reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. | Reflexive | | | | | * | 26.7% |
| b. | Pronoun | | | *! | * | | 14.6% |
| c. ☜ | Plural reflexive | | | | | | 58.7% |

The pronoun is ruled out because of the Referential Antecedent constraint and Referential Economy. The reflexive wins {14}, but Feature Matching between "every" and "himself" is a constraint on form for the child, who believes "every N" is plural.

If the form "himself" is considered the appropriate outcome, then performance is much impaired relative to tableau A.1.1. ($F(1,32)=66.120$, $p<.000$){15}.

C.1.2. <Every bear touched him> Production: 1 clause, quantifier, pronoun

| | /Disjoint reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. | Reflexive | *! | | | | * | 2% |
| b. ☜? | Pronoun | | * | | * | | 53.3% |
| c. ☜ ? | Proper N | | | | * | | 44.7% |

Here the reflexive is ruled out, but pronouns should be impaired for the same reasons of the lack of salience of the antecedent. Nouns are predicted to become a strong competitor despite the economy constraint{16}, but the economy constraint may not be the same for pronouns and nouns . At least, no significant difference is predicted between this outcome and that of A.1.2., and this is confirmed ($F(1,32)=1.33$, n.s.) {17}.

C.2.1. "Here are Big Bird and all the bears. Every bear touched himself." Comprehension: 1 clause, quantifier, reflexive.

| | /Reflexive/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. | Disjoint reference | *! | | | | | 20% |
| b. ☜ | Co-reference | | | | | * | 80% |

Coreference wins, which is confirmed by the high success rate {18}. However, the answer may be cast into doubt for children because of the lack of Feature Matching (in their grammars). So the percentage should be lower than in A.2.1. This is confirmed ($F(1,32)=6.90$, $p<.013$) {19}.

C.2.2. "Here are Big Bird and all the bears. Every bear touched him." Comprehension: 1 clause, quantifier, pronoun.

| | /Pronoun/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. ☜ ☐ | Disjoint reference | | * | | | | 55.6% |

| | | | | | | |
|---|---|---|---|---|---|---|
| b. ☒ Co-reference ☐ | | | * | | * | 44.4% |

This tableau allows a test of which ranking to give to Referential Antecedent versus Referential Salience, both predicted to be strong constraints on pronoun comprehension. Because of the Referential Antecedent constraint that disallows coreference between a pronoun and a quantified NP, disjoint reference should win. But if Referential Salience is a higher ranked constraint for pronouns, the result should be coreference. As can be seen, the result is something of a draw. The difference in the choices is not statistically significant by a matched pairs t-test {20}.

Pronoun comprehension is predicted to be much better compared to the "no quantifier" case, i.e. A.1.1.. This is true: F (1,32)=26.819, p<.000 {21}. It should also be *less* than perfect because of lack of salience of the antecedent. A One-Sample t-test proves that the disjoint reference choice is significantly below 100%: t (14)=-7.338, p<.001 or even 90% if that criterion is too stringent (t=-6.0, p<.001) {22}.

D.1.1.< Big Bird says every bear touched himself> Production: 2 clause, quantifier: reflexive.

| | /Co-reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. | Reflexive | | | | | * | 29.2% |
| b. | Pronoun | | | *! | * | | 38.6% |
| c. ☒ | Plural reflexive | | | | | | 32.3% |

The pronoun should be excluded by the referential antecedent constraint, and certainly reflexives (plural or singular) are preferred. The result should be matched to the one clause quantifier case (C. 1.1). There is no significant difference as predicted {23}. Plural reflexives are frequent as an alternative that preserves the Feature Matching the children prefer {24}. However, the 38.6% use of the pronoun suggests the Referential Antecedent constraint is not much of a constraint {*25}.

D.1.2. <Big Bird says every bear touched him> Production: 2 clause, quantifier: pronoun

| | /Disjoint reference/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|---|
| a. | Reflexive | *! | | | | * | 2.8% |
| b. ☒ | Pronoun | | | | * | | 92.8% |
| c. | Proper N | | | | * | | 4% |

The pronoun should win against both the reflexive and the noun, since the noun represents a worse violation of referential economy. The choice of pronoun should be more definite than in the one clause case with the quantifier (C.1. 2.) due to the alleviation of the Referential salience constraint. The difference is statistically highly significant (F (1, 28)=15.72, p<.001){26}.

D.2.1. “Big Bird says every bear touched himself” Comprehension, 2 clause, quantifier: reflexive

| /Reflexive/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. Disjoint reference | *! | | | | | 37.5% |
| b. ☜ Co-reference | | | | | * | 62.5% |

Coreference wins, but it should be impaired because of the Feature Matching problem. Performance should be worse on the reflexive in the quantifier case because of the Feature Matching, but NOT different from the one clause quantifier case (C.2.1). The presence of the quantifier makes a huge difference, $F(1,64)=15.34$, $p<.000$) {27}, but there is no significant effect of the two clauses {28}. So here again Feature Matching makes a difference.

D.2.2. “Big Bird says every bear touched him”. Comprehension, 2 clause, quantifier: pronoun

| /Pronoun/ | PRINCIPLE A | REFERENTIAL SALIENCE | REFERENTIAL ANTECEDENT | REFERENTIAL ECONOMY | FEATURE MATCHING | RESULT |
|---|---|---|---|---|---|---|
| a. ☜ Disjoint reference | | | | | | 59.4% |
| b. Co-reference | | | *! | | * | 40.6% |

Disjoint reference wins, as there are no constraints. In this case, children should do best on pronoun comprehension. It is predicted to be better than the case (C.1.2) of two clauses with no quantifier because of the Referential Antecedent constraint ( $F(1.64)=10.49$, $p<.002$){29}. It is also predicted to be better than (B1.2.) one clause, quantifier because of the lifting of the Referential Salience constraint ($F(1.64)=25.569$, $p<.000$) {30}. Nevertheless, there is the significant 40% error rate, suggesting again that the Referential Antecedent constraint may not be very strong for “every N”. In fact, the difference in choice between disjoint and coreference is not statistically significant by a matched pairs t-test {*31}.

Finally, A regression analysis was run with pronoun comprehension as the dependent variable and quantifier and number of clauses as the two predictive variables. The results show that there are two independent significant predictors ($R^2=.367$, Beta =6.33 for number of clauses ($p<.001$); Beta=6.355 for quantifier, $p<.000$), that is, there is no interaction between the two factors.

**Discussion**

First and foremost, it is important to notice that there is a massive confirmation of the fact that production is ahead of comprehension in this domain. This fact cries out for explanation, given that it is not the usual case in language acquisition. Hendriks and Spenader presupposed a production/comprehension asymmetry in their Optimality

analysis that is now confirmed. However in the present study it proved necessary to elaborate the proposal by Hendriks and Spenader, generating sixteen different tableaux to test against the empirical data in the four conditions, two tasks, and two proforms. The new tableaux generated thirty-one predictions, of which 28 were confirmed to a high level of statistical significance. For convenience these are listed again in Table 5. The predictions not only included which element should "win" by the Optimality ranking, but also the *level* of performance compared to other scenarios. Furthermore it predicted correctly cases of *no difference* between performances, and also the *match* between actual and expected performance e.g. 50% or 100%. It is always statistically risky to conduct a large number of individual tests, but the observed p level was so low in most cases as to render this problem irrelevant.

Three exceptions occur, predictions {7}, {13} and {31}. The case of {7} occurred in testing whether the reflexive was understood at the same high level in one clause versus two clause sentences with no quantifier. Although the Optimality tableaux predicted equivalent 100% performance for both conditions, the two clause sentences - "Big Bird said Grover touched himself" - slightly reduced performance (96% to 85%, $p<.05$). The only suggestion we have is that the two-clause sentence made the matrix subject a salient alternative, but as it overrides Principle A it is not a common error. It is harder to argue that the two clause sentences created additional processing load that reduced performance, given that this condition so significantly *improved* pronoun comprehension.

The case of prediction {13} arose in comprehending "Big Bird said Grover touched him", where choice of interpretations was predicted to be 50/50, but was instead slightly ($p<.05$) in favor of the coreference alternative. One suggestion we have here is that some other weak factor might come into play, such as a bias towards the closest accessible salient referent, given that there is no block for this according to the scheme, because Principle B is not part of the analysis. The other more significant possibility here is that the difference is because of a "yes" bias, leading the children to say "yes" rather than "no", common in a truth-value judgment task.

The case of prediction {31} calls into question the strength of the Referential Antecedent constraint. Children disobey it a full 40% of the time, and this reduces their performance on D.2.2. "Big Bird said every bear touched him" to a 50/50 choice. We have already seen that some portion of this may be a "yes" bias. But in scenario C.2.2. "Here are Big Bird and all the bears. Every bear touched him." the children also overrode Referential Antecedent 44% of the time. Should the constraint of Referential Antecedent be removed from the tableaux? It must be admitted that it does little to improve the predictiveness of the other constraints, so {14} for example, could be explained purely on the basis of Referential Economy considerations.

Why might the Referential Antecedent constraint be so weak in this study? It is tempting to consider the other finding, namely that children regard "every" as having plural features, at least in so far as its anaphor goes. There is a wealth of literature suggesting that children in this age range have trouble with the quantifier "every", and may not yet treat it in a fully adult way. In fact, instances of "every +N" are very rare in CHILDES: a recent study of 18 children studied longitudinally found only 10 examples of forms like "every + N (as opposed to potentially frozen forms like "everybody") (Merchant, 2005). A recent elicited production study of "every" showed that even 5 year

olds have considerable difficulty (Altreuter & de Villiers, 2006). There is evidence that children treat "every N" as a simple plural (Roepe, Struass & Pearson, 2006), and as an event quantifier (Philip, 1995), and do not respect constraints on its movement (Coles-White, de Villiers & Roeper, 2004). But what about the classic finding that in Chien and Wexler (1990) and several other studies, children's performance on "Every bear touched him" is significantly better than "Grover touched him"? The possibility arises that it is children's constraint on Feature Matching that tips the balance in favor of a disjoint reading in this condition. This would suggest that "The boys touched him" should be easy, and "The boys said the bears touched them" should be hard. If Hendriks and Spenader are right that there is such a constraint as Referential Antecedent, then perhaps it will be operative only later in development, when children fix *every* as a DP quantifier. Or, perhaps a different methodology might render "every N" more like a proper quantified NP. In our study, the bears were pictured as a set of definite referents, also referred to in the discourse as "all the bears". If one told children that they were to be asked what they know about animals, and then asked e.g. "Does every cat wash him?" (parallel to "does every narcissist admire him?"), we suspect most children would say, "Who?" or "Who's him?", showing that they do respect the Referential Antecedent constraint. But *we*'d be speaking ungrammatically, which we try to avoid in experiments for ethical reasons.

In contrast to the constraint of Referential Antecedent, the proposed constraint of Referential Salience proved powerfully predictive of the children's performance with pronouns, both in production and comprehension. It was introduced to explain children's expressed bewilderment about suitable discourse referent for "him", showing that they had not kept track of the possible referents even across such short spans of sentences. Unfortunately, this constraint is still somewhat vaguely defined for child language, though it sounds promising to use the similar constraints proposed for adult anaphora resolution in Bouma (2003) and Beaver (2004). Grimshaw and Rosen (1989) formulated a similar explanation for children's failure to obey Principle B, but their account was met with a torrent of counter proposals and counter-evidence. Nevertheless, their proposal had a similar concern for children's ability to keep track of referents in the experimental discourse. Notice that the experimental sentences are very reduced texts, in which no real topics are maintained, and children may need more cohesion to establish and maintain a discourse referent. Two obvious follow-ups suggest themselves and may help to define the notion of Referential Salience with more precision. One is to make the antecedent referent salient by ONLY mentioning him, e.g. "Here is Grover. Big Bird is touching him". The prediction on the basis of Referential Salience would be that such a discourse focus would improve performance on pronoun comprehension. A second is to have a larger narrative, in which the topic is some salient character who has various adventures. It is also likely that children will then understand the pronoun as referring extra-sententially to this character. However, there is still nothing that *forbids* a coreferential interpretation of the pronoun, under the Burzio approach without Principle B. One should also test the idea that having true pronoun sentences intermixed with the false can change the salience of the disjoint referent.

Discussing the adequacy of the three constraints, Referential Antecedent, Referential Salience, and Feature Matching, brings up serious questions about the Optimality approach. Its great attraction is that the linguistic and extra-linguistic factors

that experiments introduce can be schematized and weighted to predict children's performance. This is illustrated by our success in predicting performance across these tasks. But precautions are necessary to stop the proliferation of particular and narrow constraints. Validation could be achieved by asking hard questions: Are these constraints general enough to transfer across studies? Across structures? Across languages? For example, it is well established that the delay of Principle B does not occur in languages with clitic pronouns (Baauw, 2003). Can the Optimality account be adjusted in some non ad-hoc way to predict this difference? What is the status of linguistically principled constraints such as Principle A, and those that seem more like attentional difficulties at the interface with pragmatics, such as Referential Salience? Or do the latter reflect genuine linguistic constraints that are as yet poorly formulated? Finally, is Optimality a theory of performance, or competence? Can it substitute for a competence account?

A word is necessary about methodology. The truth-value judgment task has been heralded as the most refined and respectable methodology (Crain and Thornton; Thornton and Wexler), but we have our doubts. It seems to produce inevitable "yes-bias", and when one is trying to predict performance *levels* this bias can get in the way of precision. It might be very interesting to try a methodology somewhat neglected in Principle B work, namely picture-choice. Especially when one is looking at preferences for one interpretation over the other, it seems more direct to see which picture a child would prefer. Consider the result in {19}, where the child says "no" to the correct sentence "every bear is touching himself" 20% of the time. But is the child violating Principle A and choosing a disjoint reading, or is he rejecting the sentence because the form "himself" should be "themselves", as in the child's production grammar? Presumably a picture choice task would allow the child to show us more about the basis for his rejections. However, picture choice also has some problems with complex forms in which a child has difficulty simultaneously remembering the stimulus sentence and encoding and comparing two or more pictured scenes. Nevertheless, it would be worth comparing the two comprehension methodologies.

Let us return to the production/comprehension asymmetry. There are at least two other cases that show a similar pattern, and there may be more. Children's ability to use the information carried by 3rd person /s/ reveals a similar delay of comprehension over production (Johnson & de Villiers, in press), and children's comprehension of referential opacity is delayed relative to their own production of nominals in opaque environments (de Villiers, 2004). What these three cases have in common is not immediately apparent: one concerns a syntactic principle, the next a remnant agreement rule of English morphosyntax, and the third, a semantic competence that involves keeping track of what others know a thing to be called. There is a tantalizing correspondence in the age at which comprehension becomes more adult-like, namely about age six, but there is as yet no evidence that the phenomena are empirically linked. The only general commonality, and it is one that has no real place in current theory, is that in each case something spreads from the *subject* to other forms in the sentence. In the case of 3rd /s/, number and person features dictate the form of the verb. In binding, similar number and person features spread from the subject to the reflexive, appropriately limited by Principle A. Of course pronouns are *not* in agreement with the subject, but with some other antecedent form. In referential opacity, de Villiers (2005) conjectures that "point of view" features

are dictated by the subject onto the nominals in an intentional complement under its verb. So, for instance,

Oedipus thought he married Jocasta.

cannot be rephrased as:

Oedipus thought he married his mother.

and retain its truth value, because from Oedipus' point of view, the person is not "his mother". Why should this result in an asymmetry of production and comprehension? Possibly, because in production the subject is formulated in the child's production system together with its features (of all kinds: number, person, point of view) and then these force other elements in the sentence to agree. However, in comprehension, the child must recover those subject properties from context and then determine if the agreement features are satisfactory. Agreement might be considered an automatic part of the production system, but it may be lower in importance in comprehension, relative to truth, for example, or relevance. Nevertheless, one begins to see the outline of a common story.

Next, consider the transition to adult performance, which we have not yet confronted. While the Optimality analysis goes a considerable way towards accounting for the preferences children have in the stage prior to adult grammar, we have made no progress at all in understanding the factors that shift the child to obedience to adult constraints. First, we suspect that the "lean" Optimality analysis with which Hendriks and Spenader began may represent a final developmental stage prior to adult-like performance. That is, the new constraints we introduced, of Referential Salience, and of Feature Matching, both of which disrupt the straightforward account that they offer, may resolve slowly as the child gets better at discourse and works out the status of the quantifier "every". Then the remaining principles would be the same as in Hendriks and Spenader. Notice that at this refined stage, production would be perfect, though comprehension would still be non-adult.

How then get to adult performance? Hendriks and Spenader's "bidirectional optimization" bears a certain similarity to Reinhart's "Rule I", in the sense that both require the child to do a computation across alternative representations, and draw a kind of implicature involving what would be most informative. But mastery of the idea that other people have minds whose contents may differ from one's own comes in at least two years before pronouns are understood in these environments (Wimmer & Perner, 1983; Perner, 1988; de Villiers, de Villiers, 2000). There are other tantalizing phenomena in language of a late-acquired sort that might correlate, such as the work on referential opacity (Russell, 1988; de Villiers & Fiteva, 1996; Apperly & Robinson, 1998; Kawamar & Olson, 2000; de Villiers, 2001). Especially in the light of the recent theoretical discussions of coreference involving the child's mastery of "guises" (Heim, 1998; Thorton & Wexler, 1999) it seems of interest to explore the connection of binding with referential opacity. There is also broader work on children's awareness of speaker's intent and their communicative adequacy (Robinson & Whittaker, 1987; Apperley & Robinson, in press) pertinent to the ideas behind the failure of bidirectional optimization, and of course work on Gricean implicatures (Chierchia, 2001), but no work has explored whether these phenomena relate empirically to failures in coreference. The interconnections among these phenomena deserve further exploration.

However, we return to the roughly sketched idea involving agreement of features with some antecedent, cases that are easy in production, but harder in comprehension.

Suppose that the breakthrough in each case comes about as the child is able to compare two representations: what the sentence was, and what the child himself would have said. The difference between this and bidirectional optimality is that there is no computation here of an implicature about listener's needs, just the child's own production. Compared to Rule I, it seems less awkward than the generation of alternatives to see if they could mean something different. All that is needed is a reformulation of our assumptions about comprehension and production: that in certain well-defined cases, production is what you need before you can comprehend.

Since this proposal sounds conceptually incoherent, it needs close examination. How could a child learn to produce something that he cannot first understand? The paradox emerges only because the comprehension stimuli in experiments are designed so carefully to remove other clues. In real life, the child must hear hundreds of unambiguous cases that enable the appropriate classification of pronouns and reflexives, assuming this is guided by innate principles such as Principle A. So the child might hear "Don't hit him!" or "I'm looking at myself." or " Wash yourself, not me!" , all of which in rich referential contexts would help fix the status of the forms. So linguistic features of the subject can be tied to agreement by the child's comprehension of less demanding scenarios than those that are represented in the experimental settings. By this means, productive competence is finally established. In the experiments, Optimality considerations come into play, affecting performance especially in comprehension. We are proposing that these are only overridden when the child can recruit a representation of what he would say in such a situation and compare it with what was said.

**References**


Apperly, I., & Robinson, E.J. (1998). Children's mental representation of referential relations. *Cognition*, 67, 287-309.

Apperly, I., & Robinson, E.J. (In press). Children's difficulties handling dual identity. *Journal of Experimental Child Psychology*.

Beaver (2004). The Optimization of Discourse Anaphora *Linguistics and Philosophy* 27(1), pp. 3-56)

Bloom, P., A. Barss, J. Nicol and L. Conway. Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language* 70(1):53-71, 1994.

Bouma, G. (2003) Doing Dutch pronouns automatically in Optimality Theory. Proceedings of the EACL 2003 Workshop on The Computational Treatment of Anaphora. Budapest.

Burzio, L. Anaphora and Soft Constraints. In: P. Barbosa et al. eds, *Is the best good enough? Optimality and competition in syntax*, pages 93-116. MIT Press, Cambridge, MA, 1998.

Chien, Y.-C. and K. Wexler. Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* (13):225-295, 1990.

Chierchia, Gennaro. 2001. Scalar implicatures, polarity phenomena, and the syntax/ pragmatics interface. Ms., University of Milan.

Crain, S & Thornton, R. (1998) *Investigations in Universal grammar: a guide to experiments on the acquisition of syntax and semantics.* Cambridge, MA:MIT press

de Villiers, J. & Fitneva, S. (1996). On the role of language in the development of propositional attitudes. Presented at the International congress for the study of child language. Istanbul.

de Villiers, J.G. & de Villiers, P.A. (2000) Linguistic determinism and false belief. In P. Mitchell & K. Riggs (eds) Children's Reasoning and the Mind. Hove,U.K.: Psychology Press.

Grodzinsky, Y. and T. Reinhart. The innateness of binding and the development of coreference. *Linguistic Inquiry*, 24:69-101, 1993.

Jakubowitz, Celia. 1984. On markedness and binding principles. *Proceedings of the North East Linguistic Society,*14,154-82.Kamawar, D.

Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In Astington, J.W., Harris, P.L. & Olson, D.R., (Eds) *Developing theories of mind*. Cambridge: Cambridge University Press.

Reinhart, T. Coreference and bound anaphora: A restatement of the anaphora questions.*Linguistics and Philosophy*, 6:47-88, 1983.

Reinhart, T. (in press) Processing or pragmatics?- explaining the coreference delay In T. Gibson and N. Pearlmutter (eds) *The processing and acquisition of reference*. MIT press.

Robinson, E.J. & Whittaker, S.J. (1987). Children's conceptions of relations between messages, meanings and reality. *British Journal of Development psychology*, 5, 81-90.

Russell, J. (1987). 'Can we say… ?' Children's understanding of intensionality. *Cognition*, 25, 289-308.

Solan, E. (1978) *Anaphora in child language*. Unpublished PhD dissertation, University of Massachusetts, Amherst.

Thornton, R & Wexler, K. (1999) *Principle B, VP ellipsis and interpretation in child grammars.* Cambridge, MA: MIT press.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.

# Table 1

**ANOVA**
**Tests of Within-Subjects Contrasts**

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| TASK | 42430.510 | 1 | 42430.510 | 107.122 | .000 | .669 |
| TASK * CLAUSENO | 4601.362 | 1 | 4601.362 | 11.617 | .001 | .180 |
| TASK * QUANT | 1642.585 | 1 | 1642.585 | 4.147 | .047 | .073 |
| TASK * CLAUSENO * QUANT | 61.784 | 1 | 61.784 | .156 | .694 | .003 |
| Error(TASK) | 20992.949 | 53 | 396.093 | | | |
| PROFORM | 15536.079 | 1 | 15536.079 | 17.792 | .000 | .251 |
| PROFORM * CLAUSENO | 5114.617 | 1 | 5114.617 | 5.857 | .019 | .100 |
| PROFORM * QUANT | 10493.262 | 1 | 10493.262 | 12.017 | .001 | .185 |
| PROFORM * CLAUSENO * QUANT | 27.040 | 1 | 27.040 | .031 | .861 | .001 |
| Error(PROFORM) | 46281.203 | 53 | 873.230 | | | |
| TASK * PROFORM | 59044.064 | 1 | 59044.064 | 138.006 | .000 | .723 |
| TASK * PROFORM * CLAUSENO | 380.489 | 1 | 380.489 | .889 | .350 | .017 |
| TASK * PROFORM * QUANT | 9466.573 | 1 | 9466.573 | 22.127 | .000 | .295 |
| TASK * PROFORM * CLAUSENO * QUANT | 311.899 | 1 | 311.899 | .729 | .397 | .014 |
| Error(TASK*PROFORM) | 22675.402 | 53 | 427.838 | | | |

# Tests of Between-Subjects Effects

Transformed Variable: Average

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| Intercept | 1224112.812 | 1 | 1224112.812 | 2160.251 | .000 | .976 |
| CLAUSENO | 150.249 | 1 | 150.249 | .265 | .609 | .005 |
| QUANT | 185.035 | 1 | 185.035 | .327 | .570 | .006 |
| CLAUSENO * QUANT | 780.841 | 1 | 780.841 | 1.378 | .246 | .025 |
| Error | 30032.611 | 53 | 566.653 | | | |

Table 2

Means

| CLAUSENO | QUANT | | Produce "him",ignore nouns | Produce any reflexive | Comprehend pronoun | Comprehend reflexive |
|---|---|---|---|---|---|---|
| 1.00 | None | Mean | 100.0000 | 92.5922 | 6.9444 | 95.8333 |
| | A | N | 10 | 18 | 18 | 18 |
| | | Std. Deviation | .00000 | 17.36088 | 11.52221 | 9.58706 |
| | Every | Mean | 98.0769 | 85.3887 | 44.4438 | 80.0004 |
| | C | N | 13 | 15 | 15 | 15 |
| | | Std. Deviation | 6.93375 | 28.45017 | 29.32242 | 24.55968 |
| 2.00 | None | Mean | 97.2222 | 76.4811 | 32.8706 | 94.4444 |
| | B | N | 18 | 18 | 18 | 18 |
| | | Std. Deviation | 8.08452 | 31.02778 | 33.75637 | 13.70797 |
| | Every | Mean | 93.2292 | 67.7083 | 59.3748 | 72.9173 |
| | D | N | 16 | 16 | 16 | 16 |
| | | Std. Deviation | 12.25453 | 39.19021 | 25.79726 | 27.80539 |

Table 3

ANOVA with adjusted production scores
Within-subjects tests

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| TASK | 169.429 | 1 | 169.429 | 0.288 | 0.594 | 0.005 |
| TASK * CLAUSENO | 3478.549 | 1 | 3478.549 | 5.906 | 0.018 | 0.088 |
| TASK * QUANT | 15257.878 | 1 | 15257.878 | 25.906 | .000 | 0.298 |
| TASK * CLAUSENO * QUANT | 82.998 | 1 | 82.998 | 0.141 | 0.709 | 0.002 |
| Error(TASK) | 35926.772 | 61 | 588.963 | | | |
| PROFORM | 24942.391 | 1 | 24942.391 | 29.432 | .000 | 0.325 |
| PROFORM * CLAUSENO | 24765.923 | 1 | 24765.923 | 29.224 | .000 | 0.324 |
| PROFORM * QUANT | 49085.195 | 1 | 49085.195 | 57.921 | .000 | 0.487 |
| PROFORM * CLAUSENO * QUANT | 3367.24 | 1 | 3367.24 | 3.973 | 0.051 | 0.061 |
| Error(PROFORM) | 51694.18 | 61 | 847.446 | | | |
| TASK * PROFORM | 66011.228 | 1 | 66011.228 | 115.245 | .000 | 0.654 |
| TASK * PROFORM * CLAUSENO | 5592.205 | 1 | 5592.205 | 9.763 | 0.003 | 0.138 |
| TASK * PROFORM * QUANT | 1184.227 | 1 | 1184.227 | 2.067 | 0.156 | 0.033 |
| TASK * PROFORM * CLAUSENO * QUANT | 1121.037 | 1 | 1121.037 | 1.957 | 0.167 | 0.031 |
| Error(TASK*PROFORM) | 34940.297 | 61 | 572.792 | | | |

Between Subjects tests:

Transformed Variable: Average

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| Intercept | 991645.218 | 1 | 991645.218 | 1656.789 | .000 | .964 |
| CLAUSENO | 16395.501 | 1 | 16395.501 | 27.393 | .000 | .310 |
| QUANT | 4355.089 | 1 | 4355.089 | 7.276 | .009 | .107 |
| CLAUSENO * QUANT | 390.790 | 1 | 390.790 | .653 | .422 | .011 |
| Error | 36510.591 | 61 | 598.534 | | | |

Table 4
Means on adjusted production scores

| CLAUSENO | QUANT | | Produce "himself" | Produce "him" | Comprehend pronoun | Comprehend reflexive |
|---|---|---|---|---|---|---|
| One clause | None A | Mean | 32.3529 | 92.5922 | 6.9444 | 95.8333 |
| | | N | 17 | 18 | 18 | 18 |
| | | Std. Deviation | 37.25410 | 17.36088 | 11.52221 | 9.58706 |
| | Every C | Mean | 53.3333 | 26.6667 | 44.4438 | 80.0004 |
| | | N | 15 | 15 | 15 | 15 |
| | | Std. Deviation | 36.43324 | 30.07728 | 29.32242 | 24.55968 |
| Two clause | None B | Mean | 97.2222 | 76.4811 | 32.8706 | 94.4444 |
| | | N | 18 | 18 | 18 | 18 |
| | | Std. Deviation | 8.08452 | 31.02778 | 33.75637 | 13.70797 |
| | Every D | Mean | 92.7778 | 29.1671 | 59.3748 | 72.9173 |
| | | N | 15 | 16 | 16 | 16 |
| | | Std. Deviation | 12.54621 | 37.26745 | 25.79726 | 27.80539 |

Table 5

| Prediction | Stimulus situation | Prediction | Confirmation status |
|---|---|---|---|
| 1 | <Big Bird touched himself> | Reflexive strongly preferred | √ |
| 2 | <Big Bird touched him> | Reflexives disallowed, Nouns preferred | √ |
| 3 | <Big Bird touched him> | Pronoun production lower than reflexive in {1} | √ |
| 4 | "Big Bird touched himself" | Coreference strongly preferred | √ |
| 5 | "Big Bird touched him" | Coreference strongly preferred | √ |
| 6 | "Big Bird touched him" | Percentage of disjoint lower than chance | √ |
| 7 | "Big Bird said Grover touched himself" | Reflexive preferred to same degree as {1} | X (marginal) |
| 8 | <Big Bird said Grover touched himself> | Pronoun production markedly better than {3} | √ |
| 9 | <Big Bird said Grover touched himself> | Nouns should never be used | √ |
| 10 | "Big Bird said Grover touched himself" | Coreference strongly preferred | √ |
| 11 | "Big Bird said Grover touched himself" | Coreference no different than {4} | √ |
| 12 | "Big Bird said Grover touched him" | Pronoun comprehension much better than in 1 clause case {6} | √ |
| 13 | "Big Bird said Grover touched him" | Pronoun comprehension should be at chance | X (marginal) |
| 14 | <Every bear touched himself> | Reflexive wins | √ |
| 15 | <Every bear touched himself> | Reflexive performance impaired compared to {1} | √ |
| 16 | <Every bear touched him> | Nouns and pronouns win | √ |
| 17 | <Every bear touched him> | Pronoun performance same as {2} | √ |
| 18 | "Every bear touched himself" | Coreference wins | √ |

| | | | |
|---|---|---|---|
| 19 | “Every bear touched himself” | Coreference performance weak compared to {4} | √ |
| 20 | “Every bear touched him” | Draw between two strong constraints | √ |
| 21 | “Every bear touched him” | Pronoun comprehension better than {5} | √ |
| 22 | “Every bear touched him” | Disjoint reference below 100% | √ |
| 23 | <Big Bird said every bear touched himself> | Reflexive performance matched to {14} | √ |
| 24 | <Big Bird said every bear touched himself> | Plural reflexives should be preferred | √ |
| 25 | <Big Bird says every bear touched himself> | Pronouns should not be used b/c of quantified NP | X |
| 26 | <Big Bird says every bear touched him> | Pronouns win more decisively than in one clause case {16} | √ |
| 27 | “Big Bird says every bear touched himself” | Reflexive comprehension worse than with no quantifier | √ |
| 28 | “Big Bird says every bear touched himself” | Coreference at same level as one clause {18} | √ |
| 29 | “Big Bird says every bear touched him”. | Disjoint reference more than 2 clause, no quantifier {12} | √ |
| 30 | “Big Bird says every bear touched him”. | Disjoint more than one clause, quantifier {21} | √ |
| 31 | “Big Bird says every bear touched him”. | Coreference should not be allowed | X |