University of Massachusetts Occasional Papers in Linguistics

Volume 24 *UMOP 22: New Perspectives on Language Acquisition* | Article 13

1998

Optimality Theory, Child Language and Logical Form

Arild Hestvik
University of Bergen

Optimality Theory, Child Language and Logical Form

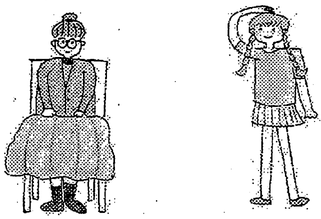
Arild Hestvik

University of Bergen

0. The Problem

As has been known since Chien and Wexler (1990), preschool children accept coreference where adults do not, in experimental contexts involving truth value judgment tasks. For example, when shown a picture like (1) and asked a question like "Is the girl patting her?", children typically respond "yes" roughly 50% of the time:

(1)



When asked similar questions involving pronouns with quantificational antecedents (and matched pictures), children typically respond "no" all of the time. This phenomenon is widely assumed to follow from a failure on behalf of children to compute a pragmatic rule governing the use of pronouns in context (cf. Chien and Wexler 1990, Grodzinsky and Reinhart 1993). However, these accounts are only compatible with experimental results from comprehension tasks. It is apparently a fact about children that their production do not exhibit this behaviour. In this paper, I propose a solution to this apparent contradiction between production and comprehension, using tools from Optimality Theory (Prince and Smolensky 1993).

New Perspectives on Language Acquisition

1. Reinhartian Binding is Characterized by Constraint Interaction

They key to solving this problem lies in uncovering the optimality theoretic nature of the standard account of this phenomenon. The classical explanation is framed within Reinhart's general approach to binding. I first review that theory, and then reformulate its insights, with some crucial modifications, into OT.

The central assumption of Reinhartian binding theory is that a referential dependency between antecedent and pronoun can be established in one of two ways: either via syntactic binding (i.e. coindexing and c-command), or via discourse reference; Reinhart labels only the latter "coreference". Syntactic binding is interpreted semantically as variable binding, and coreference means roughly "let pronoun P have the same reference as expression E in discourse D", i.e. the pronoun is not interpreted as a variable. Only syntactic binding is constrained by the binding theory in Reinhart's approach, whereas coreference is constrained by a pragmatic rule, formulated by Grodzinsky and Reinhart (1993) as "Rule I":

(2) Rule I: "NP A cannot corefer with NP B if replacing A with C, C a variable A-bound by B, yields an indistinguishable interpretation."

To illustrate, consider the string "The girl is patting her", and the intention that "her" should refer to "the girl". There are in principle two representational choices that can express this meaning: syntactic binding of the pronoun by the subject (as in 3a), or the establishment of discourse coreference between the two. Since discourse coreference is not mediated by syntactic mechanisms, I represent this without indices in the representation of the sentence (cf. 3b):

(3) a. The $girl_1$ is patting her_1
b. The girl is patting her

The syntactic binding relation in (3a) is blocked by Principle B. The coreference relation in (3b) is blocked by Rule I, since coreference here yields an interpretation that is indistinguishable from the one derived by variable binding.[1] Hence, the sentence cannot be interpreted reflexively under either representational choice. Consider however the well-known effect of a slightly different sentence in a wider context, and the two representations compatible with coreference:

(4) Nobody was patting the girl. Only the girl herself was patting her.
a. [Only the girl $herself]_1$ was patting her_1.
b. [Only the girl herself] was patting her.

[1] I.e., even though (3a) is blocked by another grammatical principle, namely Principle B. To avoid the confusion this might cause, note that the theory has the same effect on the string "John thinks that he is smart". Here, Rule I entails that an analysis under which "John" and "he" are taken as coreferent without binding violates Rule I, since the same interpretation can be derived from a syntactic binding representation, which in this case does not violate Principle B. Cf. also section 2 below.

In this context, there is a difference in entailments derived from the two representations. The semantic representation of the focus construction in (4a), roughly paraphrasable as "only the girl was an *x* such that *x* patted *x*", entails that no other person was an *x* such that *x* patted *x* (i.e. no other person had the property of self-patting), whereas (4b) entails no such thing, but rather that no other person had the property of patting the girl in question. Hence, the coreference representation yields an interpretation, in terms of entailments, which is distinguishable from the interpretation resulting from the bound variable representation, and Rule I therefore licenses (in this context) the use of a coreferent pronoun (i.e. the representation (4b) with coreference interpretation).

Note that Rule I is an "OT-like" rule. Its effect is that a constraint against coreference (i.e. Principle B) can be ignored for a string if certain other conditions are met. In a sense, Rule I competes with Principle B. In fact, Grodzinsky and Reinhart explicitly states that Rule I require that one must "...construct two representations, one for the binding option, and another for the alternative coreference reading", and that one must "compare the two representations" (p. 88). This invites an optimality theoretic reformulation, which, I argue, facilitates a better explanation for children's non-adultlike acceptance of sentences like (3a).

2. Optimizing Binding Theory

Let us recast Principles A and B and their interaction with Rule I as a system of competing and violable constraints. To review, an OT consists of:

(5)
a. a set of constraints $C_u=\{C_1, C_2, ..., C_n\}$
b. a ordering of C_u
c. a generator function GEN
d. an evaluation function EVAL

We can view the relevant constraints as Principle A, Principle B, and a constraint which encodes the gist of Rule I (although not the rule itself, as we will see below):

(6)
a. Constraint A: A reflexive must be bound in its binding domain.
b. Constraint B: A pronoun must be free in its binding domain.

These two are, as before, purely formal constraints on syntactic binding, and not on reference *per se*. The third constraint to be included is one that simply favours representations which encode referential dependencies at the syntactic level of representation, by means of coindexing and c-command. Informally, it says that if you are going to express coreference between two expression, then you must do it by entering the two elements into a syntactic binding relation. Call this constraint *Syntax-First*:

(7) Syntax-First: Referential dependencies are encoded by syntactic binding

This is also not a constraint on reference *per se*, but a constraint on the *coding* of reference. It is not identical to Rule I, but just a "stupid" rule which says that the system prefers referential relations that are syntactically encoded. This is different from the standard version of Rule I, which require special calculations to determine whether two sentences are semantically distinguishable. (I show below that this "competitive part" of Grodzinsky and Reinhart's Rule I is just a special case of how an optimality-theoretic system works in general, and therefore need not be stated as part of the rule.)

I furthermore assume that GEN takes a verb with its argument structure and an assignment of lexical heads to its arguments, and map this to a set of pairs of syntactic representations *and Logical Forms.*[2] EVAL will then evaluate subsets of the output of GEN for relative Harmony (cf. Prince and Smolensky 1993), where the subsets are constructed so that the pairs all have equivalent Logical Forms. I.e., EVAL is only interested in comparing syntactic structures that share the same LF. To illustrate this with the Syntax-First constraint in isolation, consider the input in (8):

(8) <thinks(x,Y), x=John, Y=<smart(y), y=he>>

GEN will generate outputs that are pairs of syntactic structures and Logical Forms, where the crucial property of the latter is the assignment of reference to the pronoun. Recall this this can be settled in one of two ways: either the pronoun is syntactically bound by the matrix subject, and the pronoun gets its value via variable binding, or the value of the pronoun is determined by a discourse relation. Let syntactic binding be represented as in Heim (1993), where an indexed R-expression undergoes QR and the indexed pronoun is replaced by a variable, and let coreference be represented by replacing the pronoun with a variable and an equation, DRT-style, between the variable and a discourse referent (cf. Berman and Hestvik to appear). The input in (8a) can thus be paired with either of the two logical forms in (9a,b), but will also generate e.g. (9c):

(9) a. think(j, smart(y)) & y=j ("discourse reference")
b. John [λx[think(x, smart(x))]] ("variable binding")
c. think(j, smart(y)) & y=b ("discourse reference")

Since (9a,b) have equivalent logical forms, EVAL will provide a ranking of those two outputs, but (9c) will not be ranked with respect to (9a,b). EVAL compares (9a,b) with respect to Syntax-First, and ranks (9b) higher than (9a), since (9b) has established the reference of the pronoun already at the syntactic level. This is illustrated in tableaux form below:

(T1)

	Syntax-First
a. <John thinks he is smart, think(j, smart(y)) & y=j>	*
b. <John_1 thinks he_1 is smart, λx[think(x, smart(x))](John)>	

[2] Legendre et. al (1995) assume that the input contains a skeletal specification of LF-structure. As far as I can see, it does not matter for the current analysis whether the relevant LF-properties are specified in the input, or whether GEN just takes words as input and generates combinations of structure and Lfs, and then let EVAL group those pairs into different comparison sets.

This means that the optimal and grammatical representation of the string "John thinks he is smart" with the meaning that John thinks that he, John, is smart, is the one with syntactic binding, and hence the only grammatical one.

Consider now the interaction of Syntax-First with the binding constraints. The input <like(x,y), x=John, y=him> generates, among other, the following tuples:

(10)
a. <[John [likes him]], like(j, b)>
b. <[John [likes him]], like(j, x) & x=j>
c. <[John$_1$ [likes him$_1$]], John λx[like(x, x)]>
d. <[John$_1$ [likes himself$_1$]], John λx[like(x, x)]>
e. ....

Notice that the output (10d) is not "faithful" to the input, since "him" has been replaced with "himself", and as such violate Faithfulness constraints. However, given the requirement of same logical form, it will nevertheless be included in the comparison set, and crucially so. (10a) will not be included in the comparison set, since it has a different LF. In order to evaluate these candidates, we must assume an ordering of the relevant constraints, which I take to be the following:

(11) Constraint A >> Constraint B >> Syntax-First.

This results in the following tableaux for the relevant outputs in (10) (i.e. those with equivalent logical forms):

(T2)

	A	B	SF
b. <John likes him, like(j, x) & x=j>			*
c. <John$_1$ likes him$_1$, John λx [like(x,x)]>		*	
d. <John$_1$ likes himself$_1$, John λx [like(x,x)]>			

As we see, (T2b) violates Syntax-First, but vacuously satisfies the other constraints, since the representation contains no syntactic binding. (T2c) violates constraint B, but satisfies Syntax-First (and vacuously, constraint A). (T2d) satisfies all the constraints: it satisfies constraint A, vacuously constraint B, and satisfies Syntax-First, and is therefore the optimal representation. In order words, the system determines that the syntactic representation in that tuple is optimal for the logical form in the tuple, and more optimal than, say, "John likes him" with the same meaning.

This yields the standard effect of binding theory in the OT-model. How does the effect of Rule I follow in this system? Note that there is no explicit rule here, only a "mindless" constraint that favours syntactically encoded coreference relations. Instead, the effect of Rule I follows as a consequence of the architecture of the theory. Recall that the OT-model compares outputs of GEN that are logically equivalent. They key to deriving Rule I turns on the properties of the logical forms to be compared. Consider again a typical Rule I case like (4), repeated below:

(4) Nobody was patting the girl. Only the girl herself was patting her.
a. [Only the girl herself]$_1$ was patting her$_1$. (her=the girl)
b. [Only the girl herself] was patting her. (her=the girl)

If we only look at the logical forms of sentences (4a) and (4b) in isolation, they will be equivalent. I.e., in isolation there is no truth-conditional difference between "only the girl was an x such that x patted x" and "only the girl was an x such that x patted y, and y=the girl". If, however, by the logical form of a sentence we mean its "context change potential" (cf. Heim 1988, Chierchia 1995), i.e. the ability of the sentence to alter the accumulated meaning representation of the whole discourse up to that point, then the two representations will differ in just this respect. I.e., as shown in the discussion above concerning entailments, the two representations contribute different entailments when combined with the entire context. Without formalizing meaning in such terms, it is clear enough what this means for the optimality-theoretic computations: it means that in the larger context, the meaning representations for the two sentences, couched in terms of their dynamic semantics, actually differ, and therefore will not be in the same comparison set. If (4a) and (4b) are not in the same comparison set, they will not compete. In particular, if (4b) does not compete with, say, (4a) or "Only the girl herself was patting herself", (4b) will be the optimal candidate for the relevant logical form and hence be grammatical. The details of this remains to be worked out, but it is clear how it would work. Therefore, there is no need to stipulate the comparison aspect of Rule I, since its effect will fall out as as a consequence of the optimality theoretic formulation of binding theory.

Given this model, we have to reformulate a solution to the problem of children's performance. In Grodzinsky and Reinharts theory, the explanation was that children failed to carry out the comparison of certain logical forms. In the current model, these logical forms are no longer compared by the system (they are not in the same candidate set), hence, the source of children's performance errors cannot lie therein. The next section develops the OT alternative.

3. Childrens vs. Adults' Use of OT-Grammar

Smolensky (1996) proposed a solution to a phonological production/perception dilemma. Although my proposal will differ from his, it is instructive to review his proposal, as I will adopt certain aspects of his analysis. Smolensky discussed the observation that children often display a mismatch between their phonetic production and their phonological perception. For example, they may produce "ta" as the phonetic realization of /kæt/, but still perceive [kæt] as the phonetic correspondent to /kæt/. The question arises: do they therefore have different grammars for production and perception?

Smolensky proposed that there is only one OT-grammar, with the same ranking of all constraints and the same function evaluating outputs for relative harmony, but that the grammar can be "used" in more than one way. In particular, production and perception can be modeled as two different uses of the grammar, where the difference lies in which structures compete for optimality. According to Smolensky, in production "what is fixed is an input; what competes are structures that share this particular underlying form. The overt expression of a given input is the [...] optimal structural description of that input." On the other hand, in perception "what is fixed is a surface form, and what competes are structures that share this given overt form. 'Comprehension' of a surface form is determined by the maximum-Harmony structural description containing that surface form." Smolensky continues,

"what differs between 'production' and 'comprehension' is only *which structures compete:* structures that share the same underlying form in the former case, structures that share the same surface form in the latter case."

Smolensky showed how this can account for production/perception mismatches of phonological words. Translated into syntax, "production" can similarly be viewed as a mapping from an input (a "deep structure") to an equivalence class of syntactic analyses (along with their LFs, which determines this equivalency). Perception on the other hand can be viewed as that function which maps from a surface string to a set of syntactic analyses of this string and their equivalent logical forms. The question for production is: given an input, what is the range of meanings for that input and what is the best expression for each of those meanings? For perception, the question is: which equivalence classes of syntactic representations (where equivalence is determined by "same" meaning) can a string of words be mapped to, and which structure is optimal among the members of each class? If the range of these functions differ, then different structures may be optimal for production and perception.

Let us apply this idea to the problem at hand. Consider again the tableaux of competing outputs from GEN in T2, repeated below:

(T3)

	A	B	SF
b. <John likes him, like(j, x) & x=j>			*
c. <John$_1$ likes him$_1$, John λx [like(x,x)]>		*	
d. <John$_1$ likes himself$_1$, John λx [like(x,x)]>			

The proposal is that the production/perception functions are identical for adults, but that they differ for children. Consider adults first. When an adult hears the string "John likes him", his grammar maps that string to (among others) a set of syntactic representations for coreference, such as those in (T3). These representations are then ranked, and the system determines that the optimal structure for this meaning is [John$_1$ likes himself$_1$]. This means that the structure [John$_1$ likes him$_1$] with the very same meaning is ungrammatical. This is the basis for performance in, say, a grammaticality judgment task, where the question is "can *John likes him* mean that John likes himself?" "Production" would be the task of producing a surface string for a certain meaning, say, the same meaning.[3] Since the comparison set is the same, the same structure and its corresponding string of words will be picked as the best expression of that meaning.

Consider next childrens "perception", which we view as a mapping from a surface string to one or more classes of syntactic structures, each with equivalent logical forms. My proposal is that the child grammar restricts comparison to only those structures *that share the same phonetic content*. This means that for perception purposes, they will not include (T3d) in the same comparison set as (T3b,c). If so, for the string "John likes him" with the intended meaning that 'him' is 'John', the system will only be comparing (T3b) and (T3c). (T3b) will be optimal, since it only violates the lower ranked Syntax-First, and hence be grammatical. The difference from adults is that adults include in their search space syntactic structures that contain different lexical items (within certain bounds), namely, a sentence containing a reflexive instead of a pronoun, and will then find that this is the optimal form with the given meaning. Children, on the other hand, search for optimality among structures that only have the actual surface form given. In a nutshell, this analysis says that the task of determining whether the string "John likes him" is grammatical with coreference requires comparing a range of alternative structures, and children "fail" in not including enough alternatives. Notice that this theory also predicts that children would accept coreference for "John likes himself", as this string would end up as optimal in a different comparison set.

In production, on the other hand, the child grammar goes from an input to a set of pairs of syntactic structures and equivalent logical forms. In this case, there is no issue of finding pairs of syntactic structures and their corresponding meanings for a given surface string, but rather to pick out the optimal syntactic representation for a given LF. I here assume that GEN operates automatically, and generates a representational choice with a reflexive in the relevant comparison set. The grammar will therefore automatically rank this as the optimal case. Hence, (T3d) with the reflexive then wins over the (T3b,c) with a pronoun. In an intuitive sense, production is driven by semantic form, whereas perception is driven by phonetic form.[4]

4. Matching Experimental Results

Does this theory match experimental findings? Consider first a pure grammaticality judgment task. Suppose a child is asked if the string "John likes him" can have the meaning that John likes himself (e.g. in a picture matching task, where a child is asked whether a sentence be matched to a picture). The task for the child is then to map that string to a set of competing candidates with the same phonetic contents, and then the answer is predicted to be "yes", because there is an optimal form among those which

[3] In fact, this would mean that the optimal form for an input with a pronoun and the "reflexive meaning" is an output with a reflexive pronoun.

[4] This way of modeling a production/perception difference differs from Smolensky's as follows: in Smolensky's analysis, the transition from child to adult grammar is achieved by reranking the constraints, but both adults and children keep using different parts of the grammar for production and comprehension. In the current analysis, the transition is effected by the children expanding their comparison set, so as to *neutralize* the difference between production and perception.

contains that surface structure. This means that in grammaticality judgment tasks, the child should display 100% non-adult like performance; it should always answer "yes" to the question. If an adult is asked the same question, he generates a larger set for comparison. In particular, he includes (T3d) as a candidate to compete with the structure containing a pronoun, and will find that this structure is a better expression for the same logical form, hence the other expressions are ungrammatical with that meaning. Therefore, an adult will answer "no" 100% of the time.

This appears to contradict the finding that children show 50% adultlike performance with such sentences. However, this is only apparently so. Notice that the experiments on which this finding is based are typically always truth-value judgment task. Consider then a truth-value judgment task, such as the one illustrated at the beginning of this article. Suppose a child is shown a picture like (1), and asked the question, "Is the girl patting her?", without any further indication about the intended reference of the pronoun. This was the case in the methodology employed by Hestvik and Philip (1997), where the picture is hidden from the experimenter asking the question, and the question is presented as a felicitous request for information (in the form of a yes/no-answer by the child), ostentatiously in order for the experimenter to determine "who is doing what to whom" in the unseen picture. The question for the child is then not only what the truth-value of the statement "the girl is patting her" is, but also what the truth-value of the statement is vis-a-vis a particular assignment of reference to the pronoun. If the discourse situation allows for two different and a priori possible antecedents for the pronoun, the answer depends on which meaning for the pronoun that the child assigns. (E.g., in the Hestvik and Philip experiment, the child had to guess which referent the experimenter had in mind with the question, before an answer could be determined.) If the child assigns *her=grandmother*, the answer will be "no". If the child assigns *her=girl*, the answer depends on the grammar of the child. If things are the way here argued, the answer will be "yes". If assigned and intended reference to *grandmother* and *girl* respectively is by chance throughout the experiment, the theory predicts 50% non-adultlike behaviour in this type of task. As such, the predictions are matched by experimental results.

The theory would be contradicted if a true grammaticality judgment task resulted in 50% non-adultlike behaviour rather than 100% non-adultlike behaviour. To the best of my knowledge, no study has undertaken to compare directly the results of these two experimental techniques with respect to children and Principle B.

5. Comparison with Grodzinsky and Reinhart

Building on Chien and Wexler's work, Grodzinsky and Reinhart (1993) argued that the computations involved in computing Rule I require resources that is lacking in the child. When the child attempts to compute Rule I, it therefore fails to carry out the computation. Consequently, they argue, the child guesses whether the pronoun may corefer or not. This invocation of Rule I and failure to compute occurs in every instance where a pronoun is encountered.

University of Massachusetts Occasional Papers in Linguistics, Vol. 24 [1998], Art. 13

There are several problems with account of children's behaviour with pronouns. First of all, it has never been established that children lack the computational resources to specifically compute such rules, or what those specific resources might be. In fact, if children's grammar is an OT grammar, similar processes of comparison between two (or more) forms are carried out all the time. Why should it only fail with this particular rule? Furthermore, Grodzinsky and Reinhart assume that Rule I is invoked by every instance of a string involving a pronoun, including strings where Rule I, if computed, could not in principle have licensed coreference (such as "Is the girl patting her?" out of the blue, without a potentially Rule I-licensing context). This in itself is implausible, and would be analogous to, say, checking whether subjacency has been violated in every sentence where no movement has taken place. It would have been more plausible if this effect had been hypothesized to occur only under conditions where Rule I indeed could have made a difference.

The final step of Grodzinsky and Reinhart's argument is also flawed. They assume that since the child cannot compute the result of Rule I, it consequently guesses what the outcome of the rule would have been, had it been successfully computed. This would correctly result in random performance with respect to acceptance or non-acceptance of coreference in sentences like (3). However, the problem is it does not follow from failure to compute the result of a rule that the child should *guess* what the outcome of the rule would have been. In other words, nothing about behaviour follows as a logical consequence of failure to compute. Just because I don't know whether it is legal to turn right on red in New Jersey doesn't mean that I will take a guess at every intersection. It would be just as plausible that not being able to compute the results of a rule would result in conservative behaviour, or the opposite: simply assume every time that a pronoun may corefer.

In the current model, there is no need to appeal to speculations about children's computational resources. Rather, it is a purely formal aspect of children's grammar that they choose different candidates to evaluate in a comprehension task than adults do. The real question is now: why do children not include the form with the reflexive when comparing which form optimally represents coreference, and what causes the transition to the adult system? I provide no answer to this question here, but merely note that this theory has shifted to question about what the difference is between children and adults, while at the same time avoiding some of the pitfalls of previous accounts.

There is one prediction which differentiates the current account and that of Grodzinsky and Reinhart. Avrutin and Wexler (1992) noted that the Reinhartian approach predicts that children should be insensitive to Rule I contexts, because according to this approach, they always attempt to compute Rule I, and always guess at the answer. The current approach, however, predicts that children *should* be able to differentiate between sentences like, say, (4), in an appropriate context, and sentences like (3) in out-of-the-blue contexts, because they are in different comparison sets. Similarly, children should be able, in an appropriately designed experiment, to determine that (4) is felicitous in discourses where a sentence like "only the girl is patting herself" is not, whereas Grodzinsky and Reinhart would predict no such sensitivity. I leave the testing of these predictions to future experimental investigations.

References

Avrutin, S. and K. Wexler. 1992. Development of Principle B in Russian: Coindexation at LF and Coreference. *Language Acquisition* 2:4.

Berman, S. and A. Hestvik. 1997. Split Antecedents, Noncoreference, and DRT. To appear in Bennis, H., P. Pica and J. Rooryck (eds.), *Atomism & Binding*, Foris, Dordrecht.

Chien, Y.-C. and K. Wexler. 1990. Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1, 225-295.

Chierchia, G. 1995. *Dynamics of Meaning*. University of Chicago Press.

Grodzinsky, Y. and T. Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24.1, 69-102.

Heim, I. 1988. *The Semantics of Definite and Indefinite Noun Phrases*. Garland Publishing.

Hestvik, A. and W. Philip. 1997. Reflexivity, Anti-subject Orientation and Language Acquisition. In *Proceedings of NELS 27*.

Legendre, G., C. Wilson, P. Smolensky, K. Homer and W. Raymond. 1995. Optimality and *Wh*-
Extraction. In Beckman, Urbanczyk and Walsh (eds.). *Papers in Optimality Theory*.
University of Massachusetts Occasional Papers, Department of Linguistics, Umass/Am

Prince, A. and P. Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. RuCCS-TR-2, Rutgers Center for Cognitive Science.

Smolensky, Paul. 1996. On the Comprehension/Production Dilemma in Child Language. *Linguistic Inquiry* 27, 720-731.

Department of Linguistics and Comparative Literature
University of Bergen
Sydnesplass 7
N-5007 Bergen
Norway

hestvik@uib.no