When production precedes comprehension:

An optimization approach to the acquisition of pronouns

Petra Hendriks and Jennifer Spenader

University of Groningen

Running head: When production precedes comprehension

Abstract

Data from child language comprehension shows that children make errors in interpreting pronouns as late as age 6;6, yet correctly comprehend reflexives from the age of 3;0. On the other hand, data from child language production shows that children correctly produce both pronouns and reflexives from the age of 2 or 3. Current explanations of this asymmetric delay in comprehension have either rejected the comprehension data outright or have argued that the problems are pragmatic or caused by processing limitations. In contrast, our account, formulated in the framework of Optimality Theory, handles the comprehension data as well as the production data by arguing that children acquire the ability to take into account the alternatives available to their conversational partner relatively late. It is this type of bidirectional optimization, we argue, that is necessary for correctly interpreting pronouns.

1. Children's grasp of binding principles

1.1. Children's comprehension of reflexives and pronouns

There is a well-known asymmetry in children's pattern of acquisition of the binding principles A and B. Children correctly interpret reflexives like adults from the age of 3;0 but they continue to perform poorly on the interpretation of pronouns even up to the age of 6;6 (Jakubowicz (1984); Koster and Koster (1986); Chien and Wexler (1990); McDaniel, Smith Cairns and Hsu (1990); McDaniel and Maxfield (1992); McKee (1992); see also Grimshaw and Rosen (1990) and Kaufman (1992) for a review). For example, presented in a context with two male referents, say Bert and Ernie, sentences like (1) are correctly understood from a young age (95% of the time according to some studies). However, children misinterpret the *him* in (2) as coreferring with the subject about half the time, which seems to be the result of chance performance.

(1) Bert washed himself.

(2) Bert washed him.

For this Pronoun Interpretation Problem (also referred to as the Delay of Principle B Effect), a good explanation has yet to be given.

1.2. Children's production of reflexives and pronouns

Most experiments investigating the acquisition of the binding principles focus on comprehension. However, children's production data complicates the picture. The production research suggests that children do not have problems in producing reflexives or pronouns correctly.

De Villiers, Cahillane and Altreuter (to appear) studied the production as well as the comprehension of reflexives and pronouns in 68 English speaking children between the ages of 4;6 and 7;2 (average age 6;2 years). Their study looked at two sentence types: two-sentence sequences like (3) and sentences with an embedded subordinate clause like (4).[1]

(3) Here is Baby Bear and Papa Bear.
Baby Bear is washing him/himself.

(4) Papa Bear says Baby Bear is washing him/himself.

After being tested for comprehension with a truth-value judgment task, children were shown the pictures with the same type of content and asked to describe them. Production was significantly better than comprehension for both forms and both sentences types.

Children showed minimal difficulties in correctly producing reflexives. In the embedded condition, when the target was a reflexive, they produced a pronoun only 3% of the time, and in the two-sentence condition, 14.6% of the time. This difference for sentence types was not significant.

Further, children showed superior performance in producing pronouns correctly. They almost never produced a reflexive when a pronoun was the target (never for the two-sentence sequence and only 2.8% of the time with an embedded sentence). However, children did tend to use proper names instead of pronouns in the two-sentence condition.

For the pronoun-target sentences they produced a pronoun 38% of the time and a proper name 62% of the time. A proper noun is arguably as natural as a pronoun in the context tested. The choice between them seems to depend on subtle distinctions in information structure, such as whether a relation of contrast can be established between this item and some other item. Because we do not have any information on how often adults would choose each form in the task we cannot evaluate how close children were to adults norms. However, the relevant results are that children correctly produce reflexives, and that when they do use a pronoun, they use it correctly.

These results are further supported by Bloom, Barss, Nicol and Conway's (1994) study of naturalistic data which looked at the spontaneous production of the English pronoun *me* and the reflexive *myself* in data from the CHILDES database (MacWhinney and Snow (1985; 1990)). They focused on reflexives and pronouns occurring as the object of a verb, as in *I hit myself* or *Give it to me,* since these yield the clearest test for mastery of Principles A and B. The study was limited to first person forms because these are the only forms unambiguous in a transcript. Bloom et al. were able to identify 2,834 such contexts for *me* and 75 for *myself*. Even in the youngest age groups investigated (ranging from 2;3 or 2;4 to 3;10), the children consistently used the pronoun *me* to express a disjoint meaning (99.8% correct), while they used the reflexive *myself* to express a coreferential interpretation (93.5% correct).

1.3. Strategies for reconciling experimental results and Binding Theory

How can it be explained that children are able to correctly produce forms which they are not yet able to correctly understand? Usually, comprehension of a given form precedes

production of this form (Bates, Dale and Thal (1995); Benedict (1979); Clark (1993); Fraser, Bellugi and Brown (1963); Goldin-Meadow, Seligman and Gelman (1976); Layton and Stick (1979)). Thus how do we reconcile children's poor performance on comprehension tasks with their near-perfect production data?

One possible approach is to simply reject the comprehension data. Convinced of the solidity of their production data but reluctant to accept the idea of a comprehension delay, Bloom et al. (1994) do just that. They suggest that the tasks used in the comprehension experiments do not adequately test children's grammatical competence, and argue that the production data supports a conclusion that there is no actual delay.

Grimshaw and Rosen (1990) follow the same strategy, but without appealing to the production data. Instead they argue that children do not always obey Principle B in an experimental setting when asked to interpret sentences like (2). A problem with such an account is that it is unable to explain why children's comprehension of reflexives in the same experiment is almost adult-like, given that Principle A and Principle B are generally assumed to be interrelated. To circumvent this problem, Grimshaw and Rosen disconnect Principle A from Principle B, claiming that "Knowledge of Principle A is logically independent of Principle B" (Grimshaw and Rosen (1990, 197)). This position, however, is diametrically opposed to most other accounts of reflexives and pronouns, which assume a close connection between the principles guiding the behavior of these elements or even assume a strict complementarity in the distribution of reflexives and pronouns.

Another possible explanation is to posit a dissociation between a comprehension grammar and a production grammar. If these grammars develop at different rates, this might explain why children's comprehension of certain forms lags behind their production of these forms. However, Bloom et al. reject this explanation almost

immediately. The hypothesis that there exists a fundamental dissociation between comprehension and production encounters many conceptual and empirical problems. In principle, it seems to be the case that whatever a speaker can understand, she is able to produce, and vice versa. The child language data discussed in this paper seems to represent an exception to this general pattern. Positing a complete dissociation between a comprehension grammar and a production grammar, while explaining the exceptional cases, fails to account for the general pattern.

A third possible strategy is to revise the binding principles, in particular to revise Principle B. This strategy is taken by Reinhart (1983), Chien and Wexler (1990) and Grodzinsky and Reinhart (1993). These authors first make a distinction between coindexation and coreference. Coreference interpretations are governed by pragmatic principles, such as a version of what they term Principle P (Chien and Wexler (1990)) or Rule I (Grodzinsky and Reinhart (1993)). One of the main arguments for this approach is that children seem to correctly interpret pronouns in the scope of quantified noun phrases.[2] This approach is discussed in more detail in section 5.3. Whereas Reinhart (1983) and Chien and Wexler (1990) attribute children's problems to a delay in acquiring the contextual considerations underlying the pragmatic principles, Grodzinsky and Reinhart (1993) and Reinhart (2004; to appear) argue that it is the computational complexity of constructing an alternative derivation while holding the previous one in working memory, and comparing the two derivations, which explains children's difficulties. Both of these explanations could conceivably claim that Principle P and Rule I are only involved in comprehension and that correct production does show that children know the binding principles.

Finally, one could choose to accept the results of both the comprehension and production data and try to determine how the original binding theory could account for them. This is the strategy that we will adopt, accepting the existence of a pronoun comprehension delay. That is, children are able to produce pronouns and reflexives correctly at age 2 or 3, but have problems interpreting pronouns until the age of 6;6. The aim of this paper is to explain how such a situation can arise in which a child knows how to produce a given form, but nevertheless selects an incorrect interpretation when presented with this form. Our explanation is consistent with the majority of the experimental results of children's production and comprehension of reflexives and pronouns. Additionally, our analysis is based on a linguistic model that distinguishes the speaker perspective from the hearer perspective while at the same time recognizing that these roles are obviously related and make use of the same knowledge. We choose to use a version of bidirectional Optimality Theory because it has these desirable characteristics.

The paper is organized as follows. In section 2, we introduce Optimality Theory, and the constraints that are relevant to the binding principles. Section 3 then presents our analysis of pronouns and reflexives. As we will show, the patterns found in child language can best be explained as the result of unidirectional optimization from form to meaning and from meaning to form. In contrast, the pattern found in adult language is consistent with the results of bidirectional optimization where optimization is performed on form-meaning pairs. This difference in optimization strategies is crucial, and leads to our conclusion that children make errors in pronoun interpretation because they are unable to optimize bidirectionally. In section 4, we formulate several explicit predictions

that our analysis makes and briefly discuss related phenomena. Finally, section 5, looks at some issues related to our analysis.

2. Anaphora and soft constraints

In Optimality Theory (henceforth OT, see Prince and Smolensky (2004)), a candidate set of possible outputs is generated from a given input. These possible outputs are evaluated on the basis of constraints. Constraints in OT are potentially conflicting, soft (i.e. violable) and ordered in a hierarchy according to strength. If two constraints are in conflict, it is more important to satisfy the stronger constraint than it is to satisfy the weaker constraint. The candidate that performs best in this competition is the optimal candidate. This is the output for the given input. All other candidates must be rejected. Because the constraints are potentially conflicting, it is possible that the optimal candidate also violates one or more of the constraints. Therefore, constraints in OT must be violable: a constraint violation is not always fatal. It only renders a candidate suboptimal if its competitors do not violate this constraint and behave similarly with respect to stronger constraints. For the present purposes, an important property of OT is that it can model both language production and language comprehension. In language production, the input is a meaning and the output a form. Conversely, in language comprehension the input is a form and the output a meaning.

Returning to reflexives and pronouns, it has been pointed out that their production and comprehension within a given language are highly dependent on the other anaphoric expressions in the language (e.g., Burzio, 1998). Because the set of anaphoric devices available to languages can differ greatly, an account of reflexives and pronouns in terms

of morphological class is problematic. Burzio (1998) therefore proposes to describe their behavior in terms of implicational hierarchies, which can be straightforwardly translated into soft constraints. Burzio's soft-constraint alternative to the principles A, B and C of Binding Theory is based on the following two constraints:

(5) PRINCIPLE A: A reflexive must be bound locally[3]

(6) REFERENTIAL ECONOMY: a >> b >> c

a. bound NP = reflexive

b. bound NP = pronoun

c. bound NP = R-expression

Although Burzio does not do so himself, we refer to the first constraint as PRINCIPLE A, since its effect is similar to that of Principle A of Binding Theory. PRINCIPLE A relates a particular form (a reflexive) to a particular meaning (a locally bound interpretation) and as such can be viewed as a faithfulness constraint evaluating the mapping from input to output (i.e., from form to meaning or from meaning to form). It will have a similar effect in production and comprehension. The constraint should be interpreted like material implication, i.e., every reflexive, whether in the input or output, must be associated with a locally bound interpretation in the corresponding output or input.[4]

The second constraint, which Burzio terms REFERENTIAL ECONOMY, actually is a markedness hierarchy (cf. Prince and Smolensky (2004); see Aissen (1999, 2003) for a recent investigation of markedness hierarchies in OT syntax) consisting of three markedness constraints which are ranked with respect to each other. REFERENTIAL

ECONOMY reflects the view that expressions with less referential content are preferred over expressions with more referential content. Because Burzio considers "reflexives to have no inherent referential content, pronouns to have some, and R-expressions[5] to have full referential content" (Burzio (1998, 93)), the effect of this constraint sub-hierarchy is that reflexives are preferred to pronouns as bound NPs, and pronouns are preferred to R-expressions as bound NPs.

Several other researchers have also argued for the existence of an economy constraint on the specification of bound elements (Richards (1997); Wilson (2001)). Motivation for such a constraint is provided by cases of competition among anaphors. Consider the following data from Icelandic (data from Maling (1984, 212; 1986, 284), cited in Wilson (2001)):

(7) Haraldur$_i$ skipaði mér að raka *hann$_i$/sig$_i$.
Harold ordered me to shave(infinitive) him/anaphor
Harold ordered me to shave him.

(8) Jón$_i$ veit að María elskar hann$_i$/*sig$_i$
Jon knows that Maria loves(indicative) him/anaphor
Jon knows that Maria loves him.

These examples illustrate the partial complementarity of the third-person pronoun *hann* 'he' and the SE anaphor *sig* (not present in English). When a binding relation is sufficiently local, as in (7), the bound element must be realized as a reflexive, not as a pronoun. But when the binding relation is non-local, as in (8), the reflexive is excluded

and the pronoun must be used. Such distributions can be accounted for by assuming that reflexives are preferred to pronouns by a referential economy constraint. A sentence containing a pronoun loses the competition to a sentence containing a reflexive, except when the binding relation in the latter sentence is excluded by some other constraint, for example PRINCIPLE A. Hence, a constraint like REFERENTIAL ECONOMY is needed in any theory that wants to relate the grammaticality of a bound pronoun to the unavailability of an reflexive in this case.

We adopt Burzio's constraints PRINCIPLE A and REFERENTIAL ECONOMY for our analysis, but because we are concerned with the distribution of reflexives and pronouns as well as their interpretation, we revise them to distinguish the effects they have on the form of linguistic expressions from the effects they have on their interpretation. In particular, we adapt the constraint sub-hierarchy REFERENTIAL ECONOMY in such a way that it applies to the form of an expression only:

(9) REFERENTIAL ECONOMY: Avoid R-expressions >> Avoid pronouns >> Avoid reflexives

According to this formulation, certain forms are preferred to other forms, irrespective of their interpretation. Because reflexives are preferred to pronouns, every occurrence of a pronoun yields a more serious violation of REFERENTIAL ECONOMY than any occurrence of a reflexive. Since REFERENTIAL ECONOMY is a constraint pertaining to forms only, in the output-oriented framework of OT this constraint will have an effect on production only. In the remainder of this paper, we abbreviate the constraint sub-hierarchy of REFERENTIAL ECONOMY as just one constraint, and evaluate every occurrence of a

pronoun in the output as a violation of this constraint, and every occurrence of a reflexive in the output as satisfying this constraint. Since in OT it is not important whether or not a candidate violates a constraint, but rather whether it satisfies the total set of constraints better than its competitors, using REFERENTIAL ECONOMY in this abbreviated form yields the same results as using the full sub-hierarchy in our discussion of pronouns and reflexives.

If REFERENTIAL ECONOMY were the only constraint applying to the forms in a language, then the only noun phrases occurring in the language would be reflexives. However, the selection of a form is also constrained by faithfulness constraint PRINCIPLE A. We hypothesize that PRINCIPLE A is stronger than REFERENTIAL ECONOMY. This accounts for the generalization that a reflexive is used only if the speaker intends to express a coreferential meaning. In all other cases, a pronoun or R-expression must be used. Thus, the interaction between these two constraints explains Burzio's observation that pronouns (in English but also cross-linguistically) seem to fill the space from which reflexives are excluded, an observation which is extremely difficult to explain by an analysis based on inviolable principles.

3. From child language to adult competence: unidirectional and bidirectional optimization

In this section, we will show that the interaction between PRINCIPLE A and REFERENTIAL ECONOMY explains the child language data discussed in section 1 as well as the correct adult pattern. The key difference between our explanation of the child language data and our explanation of the adult pattern is the type of optimization. We will take Principle A

as a primitive, and together with the interaction of REFERENTIAL ECONOMY derive Principle B effects from it.[6] To simplify the exposition we will limit our discussion to examples where the reflexive or pronoun is in the local domain with a referential subject, like those used in most experiments on child language acquisition (but see section 4 for a discussion of pronouns outside this local domain, and section 5.3 for a discussion of pronouns with non-referential subjects). Using only these constraints, children's production data can be described by unidirectional optimization from meaning to form, and children's comprehension data by unidirectional optimization from form to meaning.

The same constraints under the same ranking predict the adult pattern of production and comprehension when bidirectional optimization is used. Thus we argue that children begin with unidirectional optimization. In order to arrive at the correct adult interpretation for pronouns, children must start to optimize bidirectionally. In other terms, children must start to take into account not only their own alternative interpretations in comprehension, but also the alternatives for production that were available to their conversational partner.

3.1. Unidirectional optimization as a model of child language

In section 2, we introduced the violable constraints PRINCIPLE A and REFERENTIAL ECONOMY and informally discussed their interaction in production. In this section, we formalize this interaction within the framework of OT and, in addition, present an account of the interaction between these two constraints in comprehension.

In OT, possible forms and possible meanings are evaluated with respect to an ordered set of constraints. Constraint evaluation is usually illustrated in OT by means of a tableau:

(10) Tableau for producing a coreferential meaning

Input: coreferential meaning	PRINCIPLE A	REFERENTIAL ECONOMY
☞ reflexive form		
pronominal form		*!

In OT, candidate outputs are generated on the basis of a given input. The input is given in the top left-hand corner of the tableau. For tableau (10) we can see that the input is a coreferential meaning. Candidate outputs are listed in the first column below the input. Here the speaker has two relevant potential forms to choose from, a reflexive form and a pronominal form.[7] Constraints in a tableau are ordered from left to right in the first row, in order of descending strength. The linear order of the two constraints shows that PRINCIPLE A is stronger than REFERENTIAL ECONOMY.

In tableau (10), the pronominal form violates REFERENTIAL ECONOMY. This is marked by the asterisk in the corresponding cell. Because the reflexive form does not violate any of the constraints, the violation of REFERENTIAL ECONOMY by the pronominal form is a fatal violation. Because of this violation the pronominal form is a suboptimal form. There is a better candidate output, namely the reflexive form, which does not violate either of the constraints. Fatal violations are indicated by an exclamation mark.

Because the reflexive form satisfies the constraints best, this form is the optimal output. Optimal outputs are indicated in OT by the pointing hand. Thus the tableau in (10) predicts that a reflexive is preferred for a coreferential meaning.[8]

Note that the input and the candidate outputs in this tableau and following tableaux are abbreviations. The input in (10) actually is a full semantic representation such as {see(Bert,Bert), tense=past} (cf. Grimshaw (1997)). This semantic representation includes the information that the two arguments of the verb are coreferential. So when we refer to meanings in the remainder of this paper, this includes indexing information. In (10), the relevant candidate outputs are the full syntactic representations [$_{IP}$ Bert [$_{VP}$ saw himself]] and [$_{IP}$ Bert [$_{VP}$ saw him]]. However, for the sake of clarity we focus on the form and interpretation of the anaphoric element, and restrict ourselves to the selection of the anaphoric expression (reflexive or pronoun) and its interpretation (coreferential with the subject or disjoint from the subject in a local domain).

In (10), optimization proceeds from meaning to form, i.e., from a speaker's perspective. This direction of optimization also allows us to determine the optimal form for a disjoint meaning (see tableau (11)). In tableau (11) we see what the grammar predicts for producing a disjoint meaning. We can see that the reflexive form violates PRINCIPLE A because the speaker's intention is to express a disjoint meaning. If the reflexive form is chosen it will be associated with a disjoint, non-locally bound, interpretation (see footnote 4).

(11) Tableau for producing a disjoint meaning

Input: disjoint meaning	PRINCIPLE A	REFERENTIAL ECONOMY

reflexive form	*!	
☞ pronominal form		*

Both the reflexive form and the pronominal form violate one of the constraints. However, because PRINCIPLE A is ranked higher than REFERENTIAL ECONOMY, a violation of PRINCIPLE A is more serious than a violation of REFERENTIAL ECONOMY. As a result, the pronominal form is the optimal form.

If optimization proceeds from form to meaning, i.e., if a hearer's perspective is taken (cf. Hendriks and de Hoop (2001)), the input is a syntactic representation without coindexation information. The candidate outputs are assumed to be semantic representations which include indexing information pertaining to the semantic relation between the two arguments (i.e., whether they are coreferential or disjoint).

The tableaux in (12) and (13) give the results of interpretation. Because REFERENTIAL ECONOMY is a markedness constraint on forms only, it is satisfied vacuously here. In interpretation the form is already given as the input. Because markedness constraints apply to candidate outputs only and do not refer to the input, this constraint is not relevant in distinguishing among candidates. Thus based on the effects of the faithfulness constraint PRINCIPLE A, it is predicted that the optimal interpretation of a reflexive is a coreferential interpretation.

(12) Tableau for interpreting a reflexive form

Input: reflexive form	PRINCIPLE A	REFERENTIAL ECONOMY
☞ coreferential meaning		

disjoint meaning	*!	

Because PRINCIPLE A only has an effect when a reflexive is present (i.e. when found in the input or as a candidate output), it is satisfied vacuously when the input form is a pronoun. REFERENTIAL ECONOMY does not distinguish among the candidates either because the form is already given as the input. The result of optimization is thus that both interpretations are equally preferred.

(13) Tableau for interpreting a pronominal form

Input: pronominal form	PRINCIPLE A	REFERENTIAL ECONOMY
☞ coreferential meaning		
☞ disjoint meaning		

If both interpretations are equally preferred, Optimality Theory predicts that each interpretation will be chosen equally as often.

As we showed in this subsection, our OT model of unidirectional optimization predicts that children who have acquired the adult ranking of the two relevant constraints prefer a coreferential meaning to be expressed by a reflexive (tableau 10) and a disjoint meaning to be expressed by a pronoun (tableau 11). These predictions are borne out by de Villiers et al. (2005) and Bloom et al.'s (1994) studies summarized in section 1.2, which showed that children correctly produce reflexives and pronouns. Unidirectional optimization also predicts that these children interpret a reflexive as expressing a

coreferential meaning (tableau 12). This prediction is consistent with the results of the comprehension experiments mentioned in section 1.1 that found that children correctly understand reflexives from a young age. A final prediction of our OT model of unidirectional optimization is that, for a pronoun, children who have acquired the adult constraint ranking will select a coreferential meaning and a disjoint meaning equally as often because both meanings are optimal for that form (tableau 13). This prediction interestingly enough parallels the observation made in comprehension experiments that children seem to perform at chance levels when interpreting pronouns, i.e. choosing each interpretation about 50% of the time. The 50% results that have been found in experiments are not only group results. Statistical analysis of the performance of individual children shows that many children also perform at chance level individually (see Reinhart (to appear) for discussion).

3.2. Adult language: Considering the speaker's alternatives

With unidirectional optimization our analysis models the child language data. However it does not predict an adult-like pattern. Using the same two constraints under the same (adult) constraint ranking, bidirectional optimization (Blutner (2000)) achieves the adult pattern.

Briefly, bidirectional optimization considers production and comprehension simultaneously by optimizing over form-meaning pairs. Bidirectional optimization is based on the principle that during a first round of optimization the best form will be associated with the best meaning according to the standard OT algorithm for constraint evaluation. This results in one or more superoptimal form-meaning pairs. Further rounds

of optimization compare the remaining form-meaning pairs. Among this new set, further superoptimal pairs can be identified. Crucially, forms and meanings that are in use in an already identified superoptimal form-meaning pair cannot be part of another superoptimal form-meaning pair.

Formally, bidirectional Optimality Theory is defined by Jäger (2002, 435) as in (14). OT has its roots in neural network theory. In neural network theory, how well an output pattern of a neural network conforms to the constraints that are implicit in the neural network can be given a numerical value: the harmony of the network. In the definition in (14), therefore, "more harmonic" means roughly "better", i.e., satisfying the constraints better than other candidates:

(14) Bidirectional Optimality (Jäger's version):[9]

A form-meaning pair <f,m> is superoptimal iff:

a. there is no superoptimal pair <f',m> such that <f',m> is more harmonic than <f,m>.

b. there is no superoptimal pair <f,m'> such that <f,m'> is more harmonic than <f,m>.

Consider a case where in a first round of optimization a form-meaning pair <f,m> is identified as superoptimal on the basis of its behavior with respect to a set of ranked constraints. Assume that there are three remaining form-meaning pairs: <f',m>, <f,m'>, and <f',m'>, and that both <f',m> and <f,m'> outperform <f',m'> with respect to the constraints. Surprisingly, perhaps, <f',m'> can be superoptimal even though the distinct pairs <f',m> and <f,m'> satisfy the constraints better. This is because <f',m> and <f,m'>

are no longer possible form-meaning pairs: there is already a superoptimal pair, namely <f,m>, that incorporates one of their component forms or component meanings. In this way a seemingly suboptimal pair like <f',m'> can win over other pairs when it is taken into account how competitors relate to already identified superoptimal pairs.

For our data, given the two meanings and the two forms there are four logically possible form-meaning pairs. These pairs are listed in the first column of the bidirectional optimization tableau in (15). The input to bidirectional optimization can be either a form or a meaning, depending on whether bidirectional optimization is used for production or for comprehension. The output is formed by one or more superoptimal pairs consisting of forms and their corresponding meanings. The form-meaning pair <reflexive, coreferential> can already be identified as a superoptimal pair in the first round of optimization. This is marked in the tableau in (15) with the symbol ☝. It is superoptimal because it incorporates the best form and the best meaning and satisfies the constraints under consideration best. No other form-meaning pair satisfies both constraints.

(15) Bidirectional optimization tableau for the production and interpretation of reflexives and pronouns

	PRINCIPLE A	REFERENTIAL ECONOMY
☝ <reflexive, coreferential>		
<reflexive, disjoint>	*	
<pronoun, coreferential>		*
✌ <pronoun, disjoint>		*

Bidirectional OT then allows a further round of optimization, considering the remaining three candidate pairs, while keeping the first superoptimal pair in mind. Notice that the second candidate associates the reflexive form with a disjoint meaning. Because the already identified superoptimal pair <reflexive, coreferential> associates this reflexive form with a coreferential meaning, this second candidate pair falls out of the competition. The third candidate pair falls out of the competition for a similar reason. This pair associates a coreferential meaning with a pronominal form. Because we have already identified an optimal form to be associated with a coreferential meaning, i.e. a reflexive form, this third candidate pair also drops out of the competition. Thus any candidate pairs that incorporate a form or a meaning that is part of the already identified superoptimal pair fall out of the competition in further rounds of optimization. The remaining pair <pronoun, disjoint> will then be identified as the second superoptimal pair, marked by ✌ in our tableau.

Because the pairs <reflexive, coreferential> and <pronoun, disjoint> are superoptimal in bidirectional OT (see tableau 15), reflexives are predicted to carry a coreferential meaning and vice versa, and pronouns are predicted to carry a disjoint meaning and vice versa. No other form-meaning pairs are superoptimal. Therefore, a reflexive which is in the local domain with a referential subject can only be interpreted as coreferential with this subject, and a pronoun can only be interpreted as disjoint in this environment. Also, a coreferential meaning can only be expressed by a reflexive, and a disjoint meaning can only be expressed by a pronoun. Thus a bidirectional OT analysis predicts normal adult production and comprehension of pronouns and reflexives.

3.3. Acquiring adult competence

We propose that children begin with unidirectional optimization, and only later acquire the ability to optimize bidirectionally. A child must, when hearing a pronoun, consider the other non-expressed forms the speaker could have used, compare the interpretation associated with the pronoun and realize that a coreferential meaning is better expressed with a reflexive. Then, by a process of elimination, the child must realize the pronoun should be interpreted as disjoint. Optimizing bidirectionally inherently involves taking into account alternatives not present in the current situation, which may be a skill acquired very late, thus explaining the lag in acquisition.

The analysis proposed in this paper is compatible with ideas in Grodzinsky and Reinhart (1993), who argue that if a coreferential interpretation for a pronoun is not distinguishable from a bound variable interpretation (i.e., an interpretation that would be obtained by using a reflexive), the coreferential interpretation is blocked. This blocking is controlled by the pragmatic Rule I (Grodzinsky and Reinhart (1993, 79)):

(16) Rule I: Intrasentential Coreference

NP A cannot corefer with NP B if replacing A with C, where C is a variable A-bound by B, yields an indistinguishable interpretation.

Rule I in effect makes coreference only possible in cases where a bound variable reading would have a different interpretation. Through Rule I speakers can reason that if the bound variable interpretation and the coreferential interpretation are the same, then the speaker cannot have intended the bound variable interpretation. Otherwise, the speaker would have chosen a reflexive form.

Our account, while very similar to Grodzinsky and Reinhart's (1993), differs in several crucial ways. Rule I is a rule that specifically pertains to the possible division of form-meaning pairs for coreference relationships. Our analysis obtains the same effects via the more general process of bidirectional optimization, a strategy that has been independently argued for in several other aspects of language (e.g. see Blutner (2000); de Hoop, Haverkort and van den Noord (2004); Zeevat (2000)).[10] From a learnability perspective, how speakers learn to optimize bidirectionally can more easily be accounted for as one step in the developmental process. Note also that our analysis of children's acquisition of pronominal interpretations parallels de Hoop and Krämer's (to appear) analysis of children's acquisition of the interpretation of indefinites in Dutch. According to de Hoop and Krämer, unmarked forms and unmarked meanings are easy for children, whereas marked forms and marked meanings are difficult. The same generalization appears to hold for pronouns and reflexives. Under our analysis, the reflexive form is the optimal and hence unmarked form. The unmarked meaning corresponding to this unmarked form is the coreferential reading. Associating the unmarked form with the unmarked meaning is easy for children. Associating the marked form (the pronoun) with the marked meaning (the disjoint reading), however, is difficult for children, since it involves bidirectional optimization. De Hoop and Krämer argue that children may crucially lack the ability to optimize bidirectionally even as late as the age of 7;0.

Another difference between our account and Grodzinsky and Reinhart's is that, rather than revising Principle B and postulating a pragmatic rule accounting for the interpretation of pronouns (Grodzinsky and Reinhart's Rule I), we are able to derive Principle B effects and the effects of Rule I from Principle A alone, through bidirectional optimization.

Additionally, our analysis more clearly distinguishes the task of a speaker from the task of a hearer. As a result our analysis is also able to model different results for production and comprehension. Grodzinsky and Reinhart's account, on the other hand, is limited to comprehension. One could, of course, argue that, since Rule I is relevant for comprehension only, no problems are predicted to arise in production, in agreement with the results in de Villiers et al. (2005) and Bloom et al. (1994). However, it is equally well conceivable that under Grodzinsky and Reinhart's account processing problems are expected to arise in production as well: if children make errors in comprehension due to the fact that they lack the processing resources to apply Rule I, shouldn't they make similar errors in production? Grodzinsky and Reinhart's analysis fails to make predictions for production.

Finally, like Grodzinsky and Reinhart's processing account of Rule I but in contrast to Thornton and Wexler's (1999) pragmatic account of Rule I, our account straightforwardly predicts the trends found in experiments that children interpret pronouns as disjoint or coreferential at chance level. Under our analysis, however, this trend is not the result of a guessing strategy, which Grodzinsky and Reinhart (1993) and Reinhart (to appear) appeal to (note that other strategies are conceivable as well), but is consistent with the predictions of our OT grammar.

On all these counts, the analysis presented here is a simpler explanation that appeals to more general principles than Rule I, and is able to more completely account for the majority of the experimental results on children's production and comprehension of binding principles. Moreover, in contrast to pragmatic explanations such as Reinhart (1983), Chien and Wexler (1990), and Thornton and Wexler (1999), our explanation integrates the pragmatic computations and the syntactic computations in one system of

optimization, in effect yielding a *grammatical* explanation of the Pronoun Interpretation Problem. This has the advantage that the interaction between the syntactic computations and the pragmatic computations is formalized. As a result, our account yields explicit predictions with respect to comprehension as well as production.

While our explanation, which attibutes children's problems in pronoun interpretation to their inability to take into account the speaker's alternatives, is a grammatical explanation, bidirectional optimization obviously is computationally more complex than unidirectional optimization. In particular, bidirectional optimization is a recursive procedure, whereas unidirectional optimization is not. In this respect, our explanation bears some resemblance to processing explanations such as Grodzinsky and Reinhart (1993), Avrutin (1999) and Reinhart (to appear). In the next section, we will discuss some differences between our account and these processing accounts.

A remaining issue to be explained is how children develop from applying unidirectional strategies to applying bidirectional strategies. Because bidirectional optimization requires awareness of their conversational partner's choices, Theory of Mind (Perner, Leekam and Wimmer (1987)) seems to be a prerequisite for making the transition. Indeed, we don't see the adult pattern until after the age at which it is assumed that children have acquired a Theory of Mind (around age 5). Because the adult pattern in pronoun interpretation seems to arise well after the age of 5, there must be some additional difficulties involved in learning or using bidirectional optimization. But note that bidirectional optimization not only requires awareness of the fact that other conversational participants have different knowledge from oneself, but in addition requires that the perspective of the other conversational participant is taken and the

optimization process is performed in the opposite direction as well. This additional step may be responsible for children's difficulties with bidirectional optimization.

4. Predictions of our analysis

Our analysis of the Pronoun Interpretation Problem makes a number of interesting predictions which do not follow from any of the alternative explanations.

Our first prediction has to do with the choice of referential expressions in non-local domains. Notice that under the proposed account a pronoun only loses the competition to a reflexive if there is a suitable antecedent within the local domain. The preceding sections focused on pronouns and reflexives within such a local domain. But often no local antecedent is available, for example if the form to be selected is a subject in canonical position. In this situation, no c-commanding potential antecedent is available within the local domain. As a result, a reflexive cannot be bound locally and PRINCIPLE A is violated. Because PRINCIPLE A is stronger than REFERENTIAL ECONOMY, in the absence of a local antecedent pronouns and referential expressions are preferred to reflexives. Because pronouns have less referential content than referential expressions and hence yield a less serious violation of REFERENTIAL ECONOMY, a pronoun is the optimal output in this situation. So whereas reflexives can be optimal in a local domain, pronouns are always preferred to reflexives outside of this local domain.

But obviously speakers also sometimes use referential expressions. So how can referential expressions ever be optimal under the proposed account? Although many other factors may influence the selection of a referential expression instead of a pronoun (e.g., demands on information structure), one major factor is the avoidance of ambiguity.

Sometimes using a pronoun results in an ambiguity which would not arise if a referential expression were used. In this situation, adult speakers realize that the intended meaning would not be recoverable for the hearer if they would use a pronoun. Hence, they use a referential expression. This recoverability condition on the production of ellipsis and anaphora automatically follows from bidirectional optimization (Blutner, de Hoop and Hendriks (to appear); Buchwald, Schwartz, Seidl and Smolensky (2002); Kuhn (2001); Vogel (2004)). If a given meaning yields as its optimal form a reduced (i.e., elliptical or anaphoric) form, but this reduced form does not return the initial meaning back again, then the reduced form is blocked for the given meaning in a bidirectional optimization model. As a result, a less reduced form is chosen.

If this explanation for the condition on recoverability is correct, and if the child is not yet able to optimize bidirectionally, we predict that outside the local domain the child will use pronouns more often than adults, even in cases where a pronoun is ambiguous for hearers and hence the intended meaning is not recoverable. This is exactly the pattern found by Karmiloff-Smith (1985). In her large-scale experimental study of children's use of cohesive devices in discourse production, involving 150 native English speaking children and 90 native French speaking children from 4 to 9 years, Karmiloff-Smith presented children with booklets containing sets of six pictures. They were asked to tell a story as the experimenter turned the pages. Children in the youngest age group, from 4;0 to 5;11 years, were found to produce pronouns much more often than older children, even in situations where the pronoun could be assigned a different meaning than the intended meaning by a hearer (Karmiloff-Smith (1985), 71). An illustrative example is the following:

(17) The little boy's walking along. He's in the sunshine and he's got a hat on. The man's giving him a balloon ... a green balloon. He asks for some money so he gives him some money and then he gives him the balloon. And then he goes home to show it to his mummy. [...].

At this stage, children typically produced strings of subject pronouns, referring at times to the main protagonist (i.e., the discourse topic) and at other times to the subsidiary protagonist (a non-topic). Only in the older age groups in the study (i.e., from the age of 6 on), were children able to block this non-adult use of pronouns in production. They then started to use definite noun phrases such as *the balloon man* for non-topics, which is to be expected under our account, given that adults prefer topics as the referent of pronouns.[11]

The observed patterns as well as the ages of the children in Karmiloff-Smith's experiment correspond to the predictions made by the proposed account. If children are not yet able to optimize bidirectionally until the age of 6 or 7, they will experience ambiguity in comprehension, and select potentially ambiguous forms in production. Both non-adult patterns are the result of children's inability to take into account the conversational partner's alternatives. Crucially, neither the pragmatic accounts nor Reinhart's (to appear) processing account predict these particular difficulties in production.[12] Reinhart's processing account is based on the process of reference-set computation, which is a computation performed by the parser rather than the grammar. Reference-set computation involves constructing, for a given derivation, a reference set consisting of pairs of derivation and interpretation, and determining whether the given derivation is appropriate, or whether the pair of derivation and interpretation could be obtained more economically. If the latter is true, the given derivation is blocked. Reinhart

(2004, 135-136) emphasizes that reference-set computation, which she argues elsewhere (Reinhart (to appear)) to be necessary for the adult-like interpretation of pronouns, is not required for production. Because she assumes children to have the relevant linguistic knowledge, she predicts no problems to arise in children's production of pronouns. This prediction is correct for the local domain (see the previous sections), but is incorrect for the non-local domain.

A second prediction of our analysis pertains to adult's comprehension of reflexives and pronouns within the local domain. Here again, our account yields different predictions than Reinhart's account. According to Reinhart, the demanding process of reference-set computation is only required for the comprehension of marked forms such as pronouns, but not for the comprehension of unmarked forms such as reflexives or for the production of marked or unmarked forms. According to our account, on the other hand, reflexives should not be easier to process than pronouns because in both cases adult's comprehension of these items involves bidirectional optimization.

It would be possible to experimentally test between Reinhart's explanation and our alternative. An experiment that compares the processing load induced by pronouns and reflexives under exactly the same circumstances could do this. Ambiguity in the form of two potential antecedents for a pronoun yields additional processing costs (Sekerina, Stromswold and Hestvik (2004)). Variable binding by a quantifier, on the other hand, seems to be easier for pronouns than determining their reference through a referential antecedent (Piñango et al. (2001)). Furthermore, sentence-internal antecedents might be easier to access than sentence-external antecedents. The crucial experiment would therefore have to be one in which pronouns with exactly one sentence-internal referential

antecedent are compared to reflexives with exactly one sentence-internal referential antecedent.

One such experiment was performed by Badecker and Straub (2002). They investigated reading times for sentences such as the following:

(18) a. John thought that Bill owed him another opportunity to solve the problem.

b. John thought that Bill owed himself another opportunity to solve the problem.

According to both the standard version of Binding Theory and our account, the pronoun in (18a) can have as its antecedent the subject of the matrix clause, *John*, but not the subject of the embedded clause, *Bill*. In contrast, the reflexive in (18b) must have as its antecedent the subject of the embedded clause, *Bill*, and cannot have *John* as its antecedent. In a self-paced reading task, these sentences were contrasted with sentences in which the noun phrase that could not be the antecedent according to Binding Theory was replaced by a proper name with an incompatible gender feature. So in the pronoun condition (18a) *Bill* was replaced by *Beth*, and in the reflexive condition (18b) *John* was replaced by *Jane*. If Binding Theory would filter out unacceptable antecedents, this change is expected not to have any effect on the reading times of the sentences, since only Binding Theory-compatible antecedents would have to be considered. However, in the pronoun condition as well as in the reflexive condition reading times were shorter in the sentences in which only one preceding NP matched the pronoun or reflexive in gender compared to reading times for the sentences where both NPs were the same gender. This result suggests that identifying the antecedent of a pronoun or reflexive

involves considering all potential antecedents, and not only referents which are compatible with the binding principles (cf. Kennison (2003), but see Nicol and Swinney (1989) for a different view). So in (18a) as well as (18b), both *John* and *Bill* are considered as possible antecedents for the pronoun and reflexive, respectively. This clearly reflects the competition among candidates which is characteristic of OT. In addition, the experiment showed pronouns and reflexives to give rise to similar reading times and hence to induce comparable processing load in adults. This is in accordance with the proposed account, but contradicts the predictions made by Reinhart's explanation based on reference-set computation. However, Badecker and Straub's (2002) study is just one study comparing pronouns and reflexives. Evidently, more work needs to be done with other on-line experimental methods such as fMRI, to confirm this result.

A final prediction of our account relates to linguistic phenomena beyond pronouns and reflexives. Bidirectional OT was originally introduced and motivated by the need to deal with the interpretation of the lexical phenomena of blocking and partial blocking (Blutner (2000)), and has since been successfully used to analyze many other phenomena, including presupposition accommodation and anaphora resolution. Other phenomena that have been argued to involve alternative forms and hence might be better treated by taking a bidirectional perspective are conversational implicatures (Blutner (2000); van Rooy (2004)), word order freezing (Beaver and Lee (2004); Lee (2001a,b); Vogel (2004)), the interpretation of indefinite subjects and objects (de Hoop and Krämer (to appear)), and the production and interpretation of marked sentential stress (Aloni, Butler and Hindsill (to appear); Hendriks, Hendrickx, Looije and Pals (2005)). We therefore predict that also in these other cases where considering alternative forms or meanings seems crucial there can be a similar gap between children's production and

their comprehension extending over several years, with acquisition delays being possible not only in comprehension but also production. The acquisition of sentential stress seems to be a case in point. Cutler and Swinney (1987, 145) discuss the "apparent anomaly in that young children's productive skills appear to outstrip their receptive skills", but provide a rather ad hoc explanation for this performance paradox which is not extendable to the Pronoun Interpretation Problem.[13] We believe a more general explanation along the lines presented here could account for this data (cf. Hendriks et al. (2005)).

5. Related Issues

In this section we consider the extendibility of our analysis to other languages than English. Also, we discuss a number of apparently problematic cases for our analysis. However, as we show, our analysis can easily be extended to handle these cases.

5.1. Cross-linguistic differences

In section 3 we presented an OT account of adult production and interpretation of reflexives and pronouns in English, and provided an explanation for the pattern that is observed with respect to the acquisition of these lexical items. At this point, an obvious question to ask is whether the same acquisitional pattern should be expected to arise in languages other than English. The answer is no. As was already mentioned earlier, the distribution and interpretation of anaphoric expressions from the same morphological class may differ from one language to the other. OT explains cross-linguistic variation through a different ranking of the same set of constraints. Other constraint rankings may

result in other optimal forms and other optimal meanings. As a consequence, when looking at the developmental transition from unidirectional to bidirectional optimization under the same constraint ranking (as we did in this paper), the acquisitional pattern of anaphoric expressions within a given language can only be determined by looking at the constraint hierarchy for that language, not by looking at the acquisitional patterns of comparable expressions in other languages, since the ranking of the constraints may differ in these languages. As a result, although we believe that our account of anaphoricity has universal explanatory power, the resulting acquisitional pattern may differ among languages. Indeed, it has been argued that the Pronoun Interpretation Problem is much more limited in a language like Spanish (Baauw and Cuetos (2003)).

That the variation in the distribution and interpretation of anaphoric expressions can be explained through constraint reranking is nicely illustrated by Fischer's (2004) cross-linguistic OT account of binding. Fisher explains the behavior of anaphoric expressions cross-linguistically using two universal sub-hierarchies of constraints, the first of which is the inverse of our constraint REFERENTIAL ECONOMY, and the second one a hierarchy of constraints referring to binding domains of different size. While all languages observe the relative ranking of the constraints within each sub-hierarchy, languages can differ in the way these two universal sub-hierarchies are interleaved. Depending on the way these two sub-hierarchies are interleaved and on the set of anaphoric expressions available in the language, the distribution and interpretation of an anaphoric expression can vary as a result of a different ordering of the same set of constraints. Thus Fischer is able to explain the different behavior of, among others, English reflexives and pronouns, German and Dutch SE anaphors, Italian clitics and Icelandic long distance anaphors.

5.2. Breakdown of complementarity

In general, reflexives and pronouns are in complementary distribution. This observation is reflected in the original formulation of Principles A and B (Chomsky (1981)), according to which the environments in which a reflexive must be bound are identical to the environments in which a pronoun must be free. In a number of contexts, however, this complementary distribution breaks down:

(19) The man hid a book behind himself.

(20) The man hid a book behind him.

These sentences are equally acceptable, and both *himself* and *him* are interpreted with *the man* as its antecedent. Cases like (19) and (20) yield a challenge for the standard binding theory as well as for our OT account of pronominal distribution and interpretation. Because a pronoun is only possible if a reflexive is ruled out, we predict complementary distribution in all cases, in child language as well as adult language.

Reinhart and Reuland (1993) propose a reformulation of the standard binding theory, according to which the binding principles constrain the dependencies between coarguments of the same predicate only. Dependencies such as in (19) and (20), where the reflexive/pronoun and its antecedent are not coarguments of the same predicate, fall outside the scope of these binding principles. The use of reflexives and pronouns in these contexts is governed by other principles, which do not force complementarity. In Fischer's (2004) OT account of binding, mentioned in the previous section, the same

effect is obtained by the interaction among two constraint hierarchies. By distinguishing among binding domains of differing sizes, it is possible to differentiate between binding of a reflexive or pronoun within the entire sentence (the subject domain, in Fischer's terminology) and binding of a reflexive or pronoun within a locative prepositional phrase (the theta domain, in Fischer's terminology). Depending on the way the two constraint hierarchies are interleaved, the binding domain of a pronoun in a given language may partially overlap with the binding domain of a reflexive in the same language, thus giving rise to a breakdown in complementarity.[14]

5.3. The coreference-coindexation distinction

There is additional experimental evidence relating to children's interpretation of pronouns and reflexives in the scope of quantified noun phrases and in VP ellipsis. This evidence has often been used to motivate revising Principle B and making a distinction between coreference and coindexation, as well as to motivate the need for a pragmatic principle like Rule I or Principle P. Therefore, how our account handles this additional data is relevant to its plausibility.

Recall from section 1.3 that in order to explain children's apparent lag in Principle B, Chien and Wexler (1990) and Grodzinsky and Reinhart (1993) revise Principle B so that it only governs bound variable uses of pronouns, making a distinction between a syntactic process of coindexing and a pragmatic process of coreference.

(21) Bert washed him.

For examples like (21), revised Principle B stipulates that *Bert* and *him* must have different indices. However, nothing prevents each of these indices from being pragmatically resolved to the same discourse referent. Thus it is impossible to distinguish errors in indicing from errors in the resolution of indices to discourse referents, where the latter is argued to be a pragmatic skill that is acquired late.

This revised Principle B can be tested by factoring out coreference by experimental design. Two common ways to do this are to test children's interpretation of sentences with universally quantified NP subjects such as *each* and *every*, and of sentences with VP ellipsis.

(22) Every bear washed him.

When used to describe a situation where every bear washes himself, example (22) should be judged as unacceptable by children who know Principle B because Principle B rules out binding in a local domain. Additionally, because universally quantified noun phrases like *every bear* are non-referential, they do not introduce discourse referents that indices could be resolved to, so pragmatic principles do not come into play. If children perform well on examples like (22) then it has been claimed that they have shown that they have mastered or have access to a principle like revised Principle B.[15] Chien and Wexler (1990) report that children correctly interpret pronouns in the scope of quantified noun phrase subjects from an early age.[16,]

Without further constraints, our own account would predict that children using a unidirectional optimization strategy would show the same type of errors with universally quantified subjects as with other NP subjects, that is, that they would be equally likely to

resolve a local pronoun to the universally quantified NP subject, as with a non-local antecedent, misinterpreting the quantified NP as referential.

However, as Hendriks and De Hoop (2001) have argued, many different types of information interact in the interpretation of quantified NPs, so it is not unlikely that additional constraints play a role in the interpretation of pronouns in the scope of quantified NPs as well. When Chien and Wexler model children's comprehension of pronouns with revised Principle B and the coindexation-coreference distinction, they make two additional assumptions: that children are aware of the non-referential property of universally quantified NPs and that children are aware that pronouns can only have referential antecedents. By making these same assumptions, and incorporating the latter one as a constraint in our analysis, we obtain the same predictions as Chien and Wexler. For example, incorporating a constraint such as the following into the analysis would make the correct predictions:

(23) REFERENTIAL ANTECEDENT: A pronoun must have a referential antecedent

This is shown in tableau (24), where the input is a pronominal form occurring in the context of a quantified NP subject (QNP) such as *every* N or *each* N. Assuming a constraint banning QNPs as antecedents for pronominal forms leads to a disjoint interpretation as the optimal meaning.

(24) Tableau for interpreting a pronominal form if the subject is a quantified NP

Input: pronominal form,	REFERENTIAL	PRINCIPLE	REFERENTIAL

	subject = QNP	ANTECEDENT	A	ECONOMY
	coreferential meaning	*!		
☞	disjoint meaning			

Once children have acquired this constraint they should be able to correctly resolve pronouns in the scope of universally quantified NPs even with a unidirectional optimization strategy.[17] The constraint will not affect the correct predictions for unidirectional optimization in production, or unidirectional optimization in the comprehension of reflexives. Bidirectional optimization will also continue to predict adult forms.

Results on experiments designed to test children's mastery of revised Principle B and the coreference-coindexation distinction have unfortunately been inconsistent. Work done by Kaufman (1988), Jakubowicz (1991) and Koster (1993) found that children made the same errors in interpreting pronouns in the scope of quantified NP subjects as they do for pronouns in the scope of referential NP subjects. Additionally, in a recent article Elbourne (2005) has called into question the methodology and conclusions of the earlier experiments done by Chien and Wexler and other related studies, concluding that there are other more likely explanations for the results obtained.[18] Our original set of two constraints would be sufficient to deal with these results.

It remains to be seen what future experiments show about children's abilities to correctly interpret bound variables. But regardless of the results, either our current proposal seems to handle these examples, or the incorporation of generally accepted assumptions in the model correctly extends the account.

6. Conclusions

In this paper, we presented an explanation for the Pronoun Interpretation Problem. Our analysis accounts for the data without assuming a more complex version of the binding principles or their parts, and also without rejecting the robust findings of comprehension experiments. We also avoid having to posit a complete dissociation between the system for comprehension and the system for production. We predict that lags in acquisition occur in cases where comprehension involves taking into account the speaker's alternatives, and that it is this bidirectional optimization, and not the principles themselves, that are acquired late (from the age of 6-7). That is, if a speaker uses a pronoun, the child must learn to draw the conclusion that the coreferential interpretation is not possible because, if the speaker would have wanted to express a coreferential interpretation, the speaker would have used a reflexive. Hence, the pronoun must receive a disjoint interpretation. Only if the child has learned to optimize bidirectionally, will she consistently assign a disjoint interpretation to a pronoun. While our explanation is similar in spirit to the processing account given in Grodzinsky and Reinhart (1993) and Reinhart (to appear), it explains the data from the properties of the grammar rather than from the properties of the parser. Moreover, it is supported by experimental evidence.

Our analysis also has additional advantages and implications. It accounts for the production data as well as the comprehension data. By using Optimality Theory, the interaction between syntax and pragmatics can be formalized within the grammar. Additionally, using Optimality Theory also makes more explicit what claims affect production, and what claims affect comprehension. The bulk of the experimental

evidence presents an example where production does not directly follow the development of comprehension. From our account it follows that theory and experimental design must carefully distinguish between claims and tests of production, and those of comprehension.

Acknowledgments

This investigation was supported in part by grants from the Netherlands Organisation for Scientific Research, NWO (grants no. 051-02-070 and 015-001-103 for Petra Hendriks, and grant no. 355-70-005 for Jennifer Spenader). Earlier versions of this paper were presented at the Tabu day, Groningen (2004); the ESSLLI'04 workshop on Semantic Approaches to Binding Theory, Nancy (2004); and the KNAW Academy Colloquium on Cognitive Foundations of Interpretation, Amsterdam (2004). We thank the audiences for useful comments. We also thank Helen de Hoop, Jill de Villiers, Charlotte Koster, Irene Krämer, Erik-Jan Smits, Angeliek van Hout, the members of the Acquisition Lab of the University of Groningen, the editors of *Language Acquisition* and two anonymous reviewers for valuable suggestions and discussion.

References

Aissen, J. (1999) "Markedness and subject choice in Optimality Theory," *Natural Language and Linguistic Theory* 17, 673-711.

Aissen, J. (2003) "Differential object marking: Iconicity vs. economy," *Natural Language and Linguistic Theory* 21, 435-483.

Aloni, M., A. Butler and D. Hindsill (to appear) "Nuclear Accent in Bidirectional Optimality Theory," in M. Aloni, A. Butler and P. Dekker, eds., *Questions in Dynamic Semantics*, Crispi Publications.

Avrutin, S. (1999) *Development of the Syntax-Discourse Interface*. Kluwer, Dordrecht.

Baauw, S. and F. Cuetos (2003) "The Interpretation of Pronouns in Spanish Language Acquisition and Language Breakdown: Evidence for the 'Delayed Principle B Effect' as a Non-unitary Phenomenon," *Language Acquisition* 11(4), 219-275.

Badecker, W. and K. Straub (2002) "The Processing Role of Structural Constraints on the Interpretation of Pronouns and Anaphors," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, 748-769.

Bates, E., P.S. Dale and D. Thal (1995) "Individual Differences and their Implications," in P. Fletcher and B. MacWhinney, eds., *The Handbook of Child Language*, Blackwell, Oxford.

Benedict, H. (1979) "Early Lexical Development: Comprehension and Production," *Journal of Child Language* 6, 183-200.

Bloom, P., A. Barss, J. Nicol and L. Conway (1994) "Children's Knowledge of Binding and Coreference: Evidence from Spontaneous Speech," *Language* 70, 53-71.

Beaver, D. and H. Lee (2004) "Input-Output Mismatches in OT," in R. Blutner and H. Zeevat, eds., *Optimality Theory and Pragmatics*, Palgrave/Macmillan.

Blutner, R. (2000) "Some Aspects of Optimality in Natural Language Interpretation," *Journal of Semantics* 17, 189-216.

Blutner, R., H. de Hoop and P. Hendriks (to appear) *Optimal Communication*, CSLI Publications, Stanford, CA.

Buchwald, A., O. Schwartz, A. Seidl and P. Smolensky (2002) "Recoverability optimality theory: Discourse anaphora in a bidirectional framework," Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), Edinburgh, UK.

Burzio, L. (1998) "Anaphora and Soft Constraints," in P. Barbosa et al., eds., *Is the Best Good Enough? Optimality and Competition in Syntax*, MIT Press, Cambridge, MA.

Chien, Y.-C. and K. Wexler (1990) "Children's Knowledge of Locality Conditions on Binding as Evidence for the Modularity of Syntax and Pragmatics," *Language Acquisition* 13, 225-295.

Chomsky, N. (1981) *Lectures on Government and Binding*, Foris, Dordrecht.

Clark, E.V. (1993) *The Lexicon in Acquisition*, Cambridge University Press, Cambridge.

Cutler, A. and D.A. Swinney (1987) "Prosody and the development of comprehension," *Journal of Child Language* 14, 145-167.

De Hoop, H., M. Haverkort and M. van den Noort (2004) "Variation in Form versus Variation in Meaning," *Lingua* 114: 9/10, 1071-1089.

De Hoop, H. and I. Krämer (to appear) "Children's Optimal Interpretations of Indefinite Subjects and Objects," *Language Acquisition*.

De Villiers, J., J. Cahillane and E. Altreuter (to appear) "What can production reveal about Principle B?," *Proceedings of GALANA 2004*, Generative Approaches to Language Acquisition North America, December 2004, University of Hawai'i at Manoa.

Eisner, J. (1999) "Doing OT in a Straitjacket," Handout, U. of Rochester Linguistics Talk, June 1999.

Elbourne, P. (2005) "On the Acquisition of Principle B," *Linguistic Inquiry* 36:3, 333-365.

Fraser, C., U. Bellugi and R. Brown (1963) "Control of Grammar in Imitation, Production, and Comprehension," *Journal of Verbal Learning and Verbal Behavior* 2, 121-135.

Fischer, S. (2004) "Optimal Binding," *Natural Language and Linguistic Theory* 22, 481-526.

Goldin-Meadow, S., M.E.P. Seligman and R. Gelman (1976) "Language in the Two Year Old," *Cognition* 4, 189-202.

Grimshaw, J. and S.T. Rosen (1990) "Knowledge and Obedience: The Developmental Status of the Binding Theory," *Linguistic Inquiry* 21, 187-222.

Grimshaw, J. (1997) "Projections, Heads, and Optimality," *Linguistic Inquiry* 28:3, 373-422.

Grodzinsky, Y. and T. Reinhart (1993) "The Innateness of Binding and the Development of Coreference," *Linguistic Inquiry* 24, 69-101.

Hendriks, P., S. Hendrickx, R. Looije and C. Pals (2005) "Hoe perfect is ons taalsysteem? Een bidirectionele OT-analyse van de verwerving van klemtoonverschuiving," *Tabu* 34, 71-97.

Hendriks, P. and H. de Hoop (2001) "Optimality Theoretic Semantics," *Linguistics and Philosophy* 24:1, 1-32.

Huang, Y. (1994) *The Syntax and Pragmatics of Anaphora: A Study with Special Reference to Chinese*, Cambridge University Press, Cambridge.

Jäger, G. (2002) "Some Notes on the Formal Properties of Bidirectional Optimality Theory," *Journal of Logic, Language and Information* 11, 427-451.

Jakubowicz, C. (1991) "Binding Principles and Acquisition Fact Revisited," paper presented at the 21st Annual Meeting of the North East Linguistics Society (NELS), University of Massachusetts, Amherst.

Jakubowicz, C. (1984) "On Markedness and Binding Principles," *Proceedings of the North Eastern Linguistics Society* 14, 154-182.

Karmiloff-Smith, A. (1985) "Language and cognitive processes from a developmental perspective," *Language and Cognitive Processes* 1, 61-85.

Kaufman, D. (1988) *Grammatical and Cognitive Interactions in the Study of Children's Knowledge of Binding Theory and Reference Relations*, Doctoral dissertation, Temple University, Phildelphia, PA.

Kaufman, D. (1992) "Grammatical or Pragmatic: Will the Real Principle B Please Stand?," in B. Lust, J. Kornfilt, G. Hermon, C. Foley, Z. Nuñez del Prado, S. Kapur, eds., *Syntactic Theory and First Language Acquisition: Cross Linguistic Perspectives, vol. 2: Binding, Dependencies and Learnability: Principles or Parameters?*, Erlbaum, Hillsdale, NJ.

Kennison, S.M. (2003) "Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing," *Journal of Memory and Language* 49, 335-352.

Koster, J. and C. Koster (1986) "The Acquisition of Bound and Free Anaphora," paper presented at the 11th Annual Boston University Conference on Language Development, Boston, MA.

Koster, C. (1993) *Errors in Anaphora Acquisition*, Doctoral dissertation, Utrecht University.

Kuhn, J. (2001) *Formal and computational aspects of optimality-theoretic syntax.* Doctoral dissertation, Universität Stuttgart.

Layton, T.L. and S.L. Stick (1979) "Comprehension and Production of Comparatives and Superlatives," *Journal of Child Language* 6, 511-527.

Lee, H. (2001a) "Markedness and Word Order Freezing," in P. Sells, ed., *Formal and Empirical Issues in Optimality Theoretic Syntax*, CSLI Publications.

Lee, H. (2001b) *Optimization in Argument Expression and Interpretation: A Unified Approach*, Doctoral dissertation, Stanford University.

Levinson, S. (2000) *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, MIT Press, Cambridge, MA.

Levinson, S. (1987) "Minimization and Conversational Inference," in J. Verschueren and M. Bertuccelli-Papi, eds., *The Pragmatic Perspective*, John Benjamins, Amsterdam.

MacWhinney, B. and C. Snow (1985) "The Child Language Data Exchange System," *Journal of Child Language* 12, 271-296.

MacWhinney, B. and C. Snow (1990) "The Child Language Data Exchange System: An Update," *Journal of Child Language* 17, 457-472.

Maling, J. (1984) "Non-Clause-Bounded Reflexives in Modern Icelandic," *Linguistics and Philosophy* 7, 211-241.

Maling, J. (1986) "Clause-Bounded Reflexives in Modern Icelandic," In: J. Maling & A. Zaenen (eds.), *Modern Icelandic Syntax*, Syntax and Semantics 24, Academic Press, San Diego, 277-287.

Mattausch, J. (2004a) "Optimality Theoretic Pragmatics and Binding Phenomena," in R. Blutner and H. Zeevat, eds., *Optimality Theory and Pragmatics*, Palgrave/Macmillan.

Mattausch, J. (2004b) *On the Optimization and Grammaticalization of Anaphora*, Doctoral dissertation, Humboldt-University, Berlin.

McDaniel, D., H. Smith Cairns and J.R. Hsu (1990) "Binding Principles in the Grammar of Young Children," *Language Acquisition* 1, 121-139.

McDaniel, D. and T. L. Maxfield (1992) "Principle B and Contrastive Stress," *Language Acquisition* 2, 337-358

McKee, C. (1992) "A Comparison of Pronouns and Anaphors in Italian and English Acquisition," *Language Acquisition* 2, 21-54.

Nicol, J. and D. Swinney (1989) "The role of structure I co-reference assignment during sentence comprehension," *Journal of Psycholinguistic Research* 18, 5-19.

Perner, J., S.R. Leekam and H. Wimmer (1987) "Three-year Olds' Difficulty with False Belief: The Case for a Conceptual Deficit," *British Journal of Developmental Psychology* 5, 125-137.

Piñango, M.M., P. Burkhardt, D. Brun and S. Avrutin (2001) "The Architecture of the Sentence Processing System: The Case of Pronominal Interpretation," paper presented at SEMPRO (Cognitive Science Annual Meeting 2001).

Prince, A. and P. Smolensky (2004) *Optimality Theory: Constraint Interaction in Generative Grammar,* Blackwell. Also appeared as Technical Report CU-CS-696-

93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993.

Reinhart, T. (1983) "Coreference and Bound Anaphora: A Restatement of the Anaphora Questions," *Linguistics and Philosophy* 6, 47-88.

Reinhart, T. (2004) "The Processing Cost of Reference-Set Computation: Acquisition of Stress Shift and Focus," *Language Acquisition* 12.2, 109-155.

Reinhart, T. (to appear) "Processing or Pragmatics? Explaining the Coreference Delay," in T. Gibson and N. Pearlmutter, eds., *The Processing and Acquisition of Reference*, MIT Press, Cambridge, MA.

Reinhart, T. and E. Reuland (1993) "Reflexivity," *Linguistic Inquiry* 24, 657-720.

Richards, N. (1997) "Competition and Disjoint Reference," *Linguistic Inquiry* 28, 178-187.

Sekerina, I.A., K. Stromswold and A. Hestvik (2004) "How do adults and children process referentially ambiguous pronouns?," *Journal of Child Language* 31, 123-152.

Solan, L. (1987) "Parameter Setting and the Development of Pronouns and Reflexives," in T. Roeper and E. Williams, eds., *Parameter Setting*, Reidel, Dordrecht.

Thornton, R. and Wexler, K. (1991) "VP Ellipsis and the Binding Principles in Young Children's Grammars," paper presented at the 16th Annual Boston University Conference on Language Development, Boston, MA.

Thornton, R. and Wexler, K. (1999) *Principle B, VP Ellipsis and Interpretation in Child Grammar*, MIT Press, Cambridge, MA.

Van Rooy, R. (2004) "Relevance and Bidirectional Optimality Theory," in R. Blutner and H. Zeevat, eds., *Optimality Theory and Pragmatics*, Palgrave Macmillan, Houndmills, Basingstoke, Hampshire.

Vogel, R. (2004) "Remarks on the Architecture of Optimality Theoretic Syntax Grammars," in R. Blutner and H. Zeevat, eds., *Optimality Theory and Pragmatics*, Palgrave Macmillan, Houndmills, Basingstoke, Hampshire.

Wilson, C. (2001) "Bidirectional Optimization and the Theory of Anaphora," in G. Legendre, J. Grimshaw, and S. Vikner, eds., *Optimality-Theoretic Syntax*, MIT Press, Cambridge, MA.

Zeevat, H. (2000) "The Asymmetry of Optimality Theoretic Syntax and Semantics," *Journal of Semantics* 17, 243-262.

Footnotes

[1] De Villiers et al. also tested children on both sentence types with quantified subjects, obtaining comparable results except for the two-sentence condition where the target for production was a reflexive. For this condition children erroneously produced a pronoun 38.6% of the time.

[2] Though a recent reanalysis of these results in Elbourne (2005) concludes that other explanations are equally, or even more likely, arguing that they do not represent evidence that children know Principle B.

[3] An anaphoric element is considered to be bound when it is coindexed with another element in c-commanding position.

[4] Formulating the constraint in terms of material implication follows the design principles of Primitive OT (Eisner 1999), argued to make constraints and their patterns of violations more transparent. Thus just as the material implication $a \rightarrow b$ is false only when a is true and b is false, this constraint will only incur a violation when a reflexive form is associated with a non-locally bound interpretation. Note also that just as $a \rightarrow b$ does not entail that $b \rightarrow a$, the constraint does not mean that every locally bound interpretation must be expressed by a reflexive.

[5] R-expressions includes proper names and full NPs.

[6] Note that deriving one of the principles from the other is a common strategy. Rather than stipulating both principles (as, e.g. Chien and Wexler (1990) do), researchers such as Jakubowicz (1984, 174), Solan (1987, 201), Reinhart (1983) and Levinson (1987) have stipulated Principle A and derived Principle B. Alternatively, work by Levinson (2000), Huang (1994) and others have stipulated Principle B and derived Principle A as an effect of neo-Gricean reasoning. This latter analysis has later been argued to be describable in terms of bidirectional optimization (Mattausch (2004a), Mattausch (2004b)).

[7] In Optimality Theory the actual number of output candidates is usually assumed to be infinite, but generally only relevant candidates are presented in expositions of an analysis.

[8] We assume that, in English, constraints that govern gender and number features are ranked higher in the constraint hierarchy than PRINCIPLE A and REFERENTIAL ECONOMY. However, because these constraints are irrelevant for the examples under discussion, we have omitted them here.

[9] Jäger calls this notion of optimality 'x-optimal' rather than 'superoptimal', but we use the term 'superoptimal' because it is a more standard term.

[10] Reinhart (2004) generalizes Rule I to the process of reference-set computation, which Reinhart (to appear) interestingly enough terms "an optimality type procedure comparing two competing representations". She argues that reference-set computation not only is required in the area of coreference, but also in areas such as implicature, Quantifier

Raising, and stress shift for focus. However, no attempts are made by Reinhart to incorporate the process of reference-set computation into the grammar.

[11] Although Karmiloff-Smith's explanation of the observed patterns is different from ours, it bears some resemblance to the proposed OT account. According to Karmiloff-Smith, children's output is initially stimulus-driven (phase 1), whereas subsequently in development an internal control process constrains children's productions. In phase 1, which seems to correspond to our phase of unidirectional optimization, representations of form/function pairs are independently stored. In contrast, in phase 2, the representations of phase 1 are redescribed on the basis of analogies of form and function. The resulting system, which resembles our bidirectional system, controls children's further use of language. From this phase on, representations of form/function pairs are evaluated with respect to the content of other entries in long-term memory, Karmiloff-Smith argues.

[12] Avrutin (1999) claims that his processing account yields correct predictions with respect to production (p. 63ff and chapter 4, section 2.1). However, the inferential mechanisms which his account makes crucial use of are not fully specified, thus making it difficult to assess this claim.

[13] Cutler and Swinney (1987, 163-4) claim that the performance paradox with respect to prosody disappears if children's production of accent at age three to four is assumed to be merely a physiological reflex which is not symptomatic of underlying prosodic competence. This physiological reflex is drawn from the work of Bolinger, who argues

that the roots of accentual focus lie in primitive physiological mechanisms associated with level of speaker excitation.

[14] Our analysis can be straightforwardly extended in the direction indicated by Fischer (2004) to distinguish between binding domains of differing size. Replacing our PRINCIPLE A with a universal sub-hierarchy of constraints in which the binding domain differs in size, while maintaining our universal sub-hierarchy REFERENTIAL ECONOMY, yields the correct predictions for the sentences in (19) and (20) as well as the sentences we discussed earlier, assuming the following ranking: PRINCIPLE A (SD) ("A reflexive must be bound within its Subject Domain") >> PRINCIPLE A (TD) ("A reflexive must be bound within its Theta Domain") ° AVOID PRONOUNS >> AVOID REFLEXIVES. The ° symbol in between the second and third constraint indicates that these two constraints are tied (which implies that both orderings of the constraints are possible, see Fischer for discussion). Because a preposition theta marks its complement, the Theta Domain in (19) and (20) is formed by the prepositional phrase, whereas the Subject Domain in these examples is formed by the entire sentence. In standard examples such as (1) and (2) the Theta Domain and the Subject Domain coincide. As the reader can check for himself, the proposed constraint ranking yields the correct result for all examples presented here. Thus our analysis is able to capture the breakdown in complementarity between pronouns and reflexives.

A difference between our account and Fischer's (2004) is that Fischer uses a hierarchy which is the inverse of our REFERENTIAL ECONOMY, and hence derives Principle A effects from Principle B, rather than the other way around. Under such an

analysis, reflexives are the marked forms. Consequently, their interpretation is expected to be acquired later than that of pronouns. However, on the grounds of the acquisition facts discussed in this paper, pronouns should be considered the marked forms, suggesting that our approach is to be preferred.

[15] Note however that with this experiment it is impossible to distinguish evidence of knowledge of Principle B from knowledge that a pronoun cannot take a quantified noun phrase as an antecedent. Only the latter knowledge would be sufficient for children to correctly interpret sentences like (22).

[16] Additional experimental evidence for the coreference-coindexation distinction and revised Principle B was found by Thornton and Wexler (1991) in experiments looking at VP ellipsis, though for simplicity we only consider the data with quantified NPs here.

[17] The ranking of the constraint relative to PRINCIPLE A and REFERENTIAL ECONOMY is not relevant for unidirectional optimization in production or comprehension, but may be relevant in a more comprehensive set of constraints. However, if future experimental results show that children do not correctly interpret pronouns in the scope of quantified NPs, then the constraint must be low ranked.