

A bidirectional explanation of the pronoun interpretation problem

Petra Hendriks and Jennifer Spenader[□]

1 Acquisition of binding phenomena

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 intrasentential pronouns even up to the age of 6;6 ([13], [6]). For example, sentences like (1) are correctly understood from a young age (95% of the time according to some studies), but the *him* in (2) is misinterpreted as coreferring with the subject about half the time, which seems to be the result of chance performance. During the same period, children's interpretation of reflexives is adult-like.

- (1) The boy saw himself.
- (2) The boy saw him.

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

There are several strategies to deal with this data. [12] and [4] revise Principle B so that (2) is no longer governed by it, making a distinction between syntactic coindexing and pragmatic coreference. As a result, another explanation has to be found for the interpretation of the pronoun in (2). One of the main arguments for this approach is that children seem to correctly interpret pronouns in the scope of quantified noun phrases. This shows knowledge of syntactic coindexing. However, experimental results here are unclear, with some experiments finding children able to interpret these pronouns ([7]), while other experiments have found that children have trouble with these pronouns as well ([8],[9]). Also, it is theoretically a matter of dispute whether pronouns in the scope of quantified noun phrases pattern with reflexives ([7]) or with pronouns as in (2) ([4]).

An alternative strategy followed by [6] is to argue that (2) is governed by Principle B but that children don't always obey this principle in an experimental setting. A problem with such an account is it is unable to explain why children behave so differently with respect to reflexives and pronouns. Additionally, they disconnect Principle A from Principle B, arguing that "Knowledge of Principle A is logically independent of Principle B" ([6]:197).

Children's production data complicates the picture. [1] studied the spontaneous production of the English pronoun *me* and the reflexive *myself* in data

[□] Center for Language and Cognition, University of Groningen, The Netherlands, {p.hendriks,j.k.spenader}@let.rug.nl

from the CHILDES database. By age 2;3-3;1 the children consistently used the pronoun to express a disjoint meaning (99.8% correct), while they used the reflexive to express a coreferential interpretation (93.5%). They thus conclude from this production data that children have competence in binding principles, even if their performance with them may sometimes indicate otherwise.¹

We instead offer a novel explanation arguing that children's lag in pronoun interpretation is due to the late acquisition of the ability to reason bidirectionally. This explanation handles both examples (1) and (2) and is consistent with the majority of the experimental results of children's production and comprehension of reflexives and pronouns.

2 Anaphora and soft constraints

Languages display a vast range of anaphoric devices. According to [3], the behavior of these elements is more adequately described in terms of implicational hierarchies than in terms of their morphological class, because the function of anaphoric expressions is affected by what other referential devices are present in a given language. Because hierarchies can be straightforwardly translated into soft constraints, which are characteristic of Optimality Theory (henceforth OT, see [11]), OT seems well suited for explaining the distribution and interpretation of reflexives and pronouns. [3](93-4) proposes a soft-constraint alternative to the Principles A, B and C of Binding Theory which is based on the following two constraints:

Principle A: A reflexive must be bound locally

Referential Economy: $a \gg b \gg c$

- a. bound NP = reflexive
- b. bound NP = pronoun
- c. bound NP = R-expression

Although [3] doesn't do so himself, we will refer to the first constraint as *Principle A*, since its effect is similar to that of Principle A of Binding Theory. For the sake of simplicity, we will not specify the term 'locally' here but assume that, for English, this constraint is satisfied if the reflexive is coreferential with a c-commanding co-argument of the same predicate. The second constraint, which [3] terms *Referential Economy*, actually consists of three soft 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 [3] considers "reflexives to have no inherent referential content, pronouns to have some, and R-expressions to have full referential content"([3]:93), the effect of this constraint is that reflexives are preferred to pronouns as bound NPs, and pronouns are preferred to R-expressions in these cases.

We adopt these two constraints but because we are concerned with the distribution as well as the interpretation of reflexives and pronouns, we revise them to

¹ This doesn't rule out the possibility of a specific third-person delay. See section 4 for discussion.

distinguish the effects of these constraints on the form of linguistic expressions from the effects they have on their interpretation. We generalize the constraint Referential Economy in such a way that it applies to the form of an expression only:

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. 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 Principle A. Under [3]’s formulation, which we adopt, Principle A establishes a relation between a specific form (a reflexive) and a specific interpretation (a coreferential meaning). Hence, Principle A is a constraint on forms as well as meanings. Because Principle A is stronger than Referential Economy, 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 [3]’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 The Pronoun Interpretation Problem and bidirectionality

In this section, we will show that the interaction between Principle A and Referential Economy explains the child language data discussed in section 1. In production, tableau 1 predicts that a reflexive is preferred for a coreferential meaning.² For a disjoint meaning, a pronoun is preferred over a reflexive, which violates *Principle A*.

INPUT: coreferential meaning	<i>Principle A</i>	<i>Referential Economy</i>
☞ reflexive		
pronoun		*!

INPUT: disjoint meaning	<i>Principle A</i>	<i>Referential Economy</i>
reflexive	*!	
☞ pronoun		*

Tableau 1 Unidirectional optimization for production

Tableau 2 gives the results of interpretation. Because *Referential Economy* is a constraint on forms it isn’t relevant here. Thus based on *Principle A*, it is predicted that the optimal interpretation of a reflexive is coreferential. Because *Principle A* only has

² We assume in English that constraints that govern gender and number features are ranked higher in the constraint hierarchy than Principle A and Referential Economy. But because these constraints are irrelevant to the Pronoun Interpretation Problem, we omit them.

an effect when a reflexive is the input or a candidate output, it isn't relevant when the input form is a pronoun. The result of optimization is thus that both interpretations are equally preferred. This could account for the results that children perform at chance levels.

INPUT: reflexive	<i>Principle A</i>	<i>Referential Economy</i>
☞ coreferential meaning		
☞ disjoint meaning	*!	

INPUT: pronoun	<i>Principle A</i>	<i>Referential Economy</i>
☞ coreferential meaning		
☞ disjoint meaning		

Tableau 2 Unidirectional optimization for interpretation

If children begin with unidirectional optimization, what then leads them to acquire the adult-like disjoint interpretation for pronouns at a later stage? Unidirectional learning strategies using constraint re-ranking will not lead to pronouns being optimally interpreted as disjoint. However, bidirectional optimization, which considers production and comprehension simultaneously, shows the adult pattern. Formally, weak-bidirectional Optimality Theory is defined by [2] in the following way, where “harmonic” means roughly “better”:

Bidirectional Optimality (weak version)

A form-meaning pair, $\langle f, m \rangle$ is bidirectionally optimal iff:

- a. there is no distinct pair $\langle f', m \rangle$ such that $\langle f', m \rangle$ is more harmonic than $\langle f, m \rangle$ and $\langle f', m \rangle$ satisfies b.
- b. there is no distinct pair $\langle f, m' \rangle$ such that $\langle f, m' \rangle$ is more harmonic than $\langle f, m \rangle$ and $\langle f, m' \rangle$ satisfies a.

For our data, given the two meanings and the two forms there are four logically possible form-meaning pairs. The form-meaning pair $\langle \text{reflexive}, \text{coref} \rangle$ is an optimal pair, marked in our bidirectional tableau 3 with the symbol ♠. It satisfies both constraints under consideration. Bidirectional OT then allows a further round of optimization. The second and third candidate pairs won't be considered in this further optimization because they each incorporate a form or a meaning that is part of the already identified optimal pair, i.e. $\langle \text{reflexive}, \text{coref} \rangle$. The pair $\langle \text{pronoun}, \text{disj} \rangle$ will then be identified as the second optimal pair, marked by ♠ in our tableau.

	<i>Principle A</i>	<i>Referential Economy</i>
☞ <reflexive, coref>		
<reflexive, disj>	*!	
<pronoun, coref>		*!
☞ <pronoun, disj>		*!

Tableau 3 Bidirectional analysis of reflexives and pronouns.

Thus a bidirectional analysis predicts normal adult usage of pronouns and reflexives. We propose that children begin with unidirectional optimization, and only later acquire the ability to optimize bidirectionally. A child must, when hearing a pronoun, reason about what other non-expressed forms are associated with the potential interpretations of pronouns, realize that a coreferential meaning is better expressed with a reflexive, and then by a process of elimination realize that because this potential meaning is already better expressed with a reflexive, the pronoun should be interpreted as disjunct. Optimizing bidirectionally inherently involves reasoning about alternatives not present in the current situation, which may be a skill acquired very late, thus explaining the lag in acquisition. This analysis is compatible with the ideas in [7], and parallels [5]’s analysis of children’s acquisition of the interpretation of indefinites in Dutch.

4 Predictions of our analysis

Our analysis makes a number of interesting predictions. First, we expect the correct production of the third person pronouns *him* and *her* to precede their correct interpretation. This prediction could be tested in production experiments. In particular, we expect third person pronouns to be produced correctly well before the age of 6;6. With respect to first and second person pronouns, however, there may not be a similar delay in comprehension because there are generally no alternative antecedents. If a child has acquired the knowledge that first person forms refer to the speaker, there will not be any ambiguity with respect to the possible referent of the pronoun *me*.

A second prediction of our analysis is that, even if there is a specific third person delay, it will only account for part of the four year gap during which the comprehension of third person pronouns lags behind the production of first person pronouns, since the transition from a unidirectional optimization strategy to a bidirectional optimization strategy is expected to take time.

Finally, we predict that in other cases where reasoning about alternative forms seems crucial, such as the interpretation of scalar implicatures, there will also be a comparable delay in comprehension.

5 Conclusions

Our analysis accounts for the Pronoun Interpretation Problem 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 disassociation between the system for comprehension and the system

for production. We predict that lags in acquisition occur in cases where comprehension involves reasoning about alternatives, and that it is this bidirectional optimization, and not the principles themselves, that are acquired late.

Acknowledgments

The authors thank Helen de Hoop, Irene Krämer and Emar Maier for their valuable comments on an earlier version of this paper. Jennifer Spenader (grant no. 355.70.005) and Petra Hendriks (grants no. 051.02.070 and 015.001.103) gratefully acknowledge the financial support of the Netherlands Organization for Scientific Research (NWO).

References

- [1] Bloom, P., A. Barss, J. Nicol and L. Conway. Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language* 70(1):53-71, 1994.
- [2] Blutner, R. Some Aspects of Optimality in Natural Language Interpretation, *Journal of Semantics* 17(3):189-216, 2000.
- [3] Burzio, L. Anaphora and Soft Constraints. In: P. Barbosa et al. eds, *Is the best good enough? Optimality and competition in syntax*, pages 93-116. MIT Press, Cambridge, MA, 1998.
- [4] Chien, Y.-C. and K. Wexler. Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* (13):225-295, 1990.
- [5] de Hoop, H. and I. Krämer, Children's optimal interpretations of indefinite subjects and objects, Unpublished manuscript, University of Nijmegen, 2004.
- [6] Grimshaw, J. and S.T. Rosen. Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry* 21:187-222, 1990.
- [7] Grodzinsky, Y. and T. Reinhart. The innateness of binding and the development of coreference. *Linguistic Inquiry*, 24:69-101, 1993.
- [8] Jakubowicz, C. Binding principles and acquisition fact revisited. Paper presented at the 21st Annual Meeting of the North East Linguistics Society (NELS), Amherst: University of Massachusetts. 1991.
- [9] Koster, C. *Errors in Anaphora Acquisition*. Ph.D. Dissertation, Utrecht University. 1993.
- [10] Koster, C. Problems with pronoun acquisition. In *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives: Vol. 2. Binding, Dependencies and Learnability*. Lust, B., Hermon, G., & Kornfilt, J., eds, Hillsdale, N.J.:Lawrence Erlbaum, 1994.
- [11] Prince, A. and P. Smolensky. *Optimality Theory: Constraint interaction in generative grammar*, Ms., Rutgers University and University of Colorado. To appear. Cambridge, MA: MIT Press., 1993.
- [12] Reinhart, T. Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy*, 6:47-88, 1983.
- [13] Wexler K. and Y.-C. Chien The development of lexical anaphors and pronouns. *Papers and Reports on Child Language Development*. Vol. 24, Stanford University. 1985.