J. Child Lang. **36** (2009), 23–52. © 2008 Cambridge University Press
doi:10.1017/S0305000908008854 Printed in the United Kingdom

Coherent discourse solves the pronoun interpretation problem*

JENNIFER SPENADER, ERIK-JAN SMITS AND PETRA HENDRIKS

University of Groningen

(*Received* 14 *June* 2007. *Revised* 25 *January* 2008. *First published online* 2 *September* 2008)

ABSTRACT

Many comprehension studies have shown that children as late as age 6;6 misinterpret object pronouns as co-referring with the referential subject about half the time. A recent review of earlier experiments testing children's interpretation of object pronouns in sentences with quantified subjects (Elbourne, 2005) also suggests that there is a 'Pronoun Interpretation Problem'. In contrast, two experiments addressing English children's pronoun production (Bloom, Barss, Nicol & Conway, 1994; de Villiers, Cahillane & Altreuter, 2006) show almost perfect usage. The aim of this study is to verify this asymmetry between pronoun production and pronoun comprehension for Dutch, and to investigate the effects of coherent discourse and topicality on pronoun production and comprehension. Employing a truth-value judgment task and an elicited production task, this study indeed finds such an asymmetry in 83 Dutch children (age range 4;5–6;6). When object pronouns were clearly established as the topic of the target sentence, the Pronoun Interpretation Problem dissolved entirely. These results are compatible with the asymmetrical grammar hypothesis of Hendriks & Spenader (2005/2006) and suggest, contrary to many previous claims, that children are highly proficient at using pragmatic clues in interpretation.

[*] We are especially grateful to the children, teachers and parents of the Haydnschool and the Nassauschool in Groningen for participating in this study. We would also like to thank Jill de Villiers and the Acquisition Lab at the University of Groningen for discussion of the experiment, Robbert Prins for making the drawings used in the experiment, John Hoeks for his help with the statistical analysis of the data, and the editors and two anonymous reviewers for their valuable comments. Petra Hendriks gratefully acknowledges the Netherlands Organization for Scientific Research (grant no. 277-70-005) for their financial support. Address for correspondence: Jennifer Spenader, Artificial Intelligence, University of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands. tel: +31-50-3636327; fax: +31-50-3636687; e-mail: J.Spenader@ai.rug.nl.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

INTRODUCTION

Many experiments in numerous languages have established that children who correctly interpret reflexives from the age of four or five, have trouble interpreting pronouns correctly until past the age of 6;6 (e.g. Chien & Wexler, 1990). Consider the following example, based on Chien & Wexler's materials:

(1) This is Mamma Bear. This is Goldilocks. Is Mamma Bear washing herself?
(2) This is Mamma Bear. This is Goldilocks. Is Mamma Bear washing her?

Children consistently correctly interpret *herself* as referring to Mamma Bear in (1). At the same time, when presented with (2), children incorrectly choose Mamma Bear as the referent for *her* as often as they choose Goldilocks, about 50% of the time. This delay of correct interpretation of pronouns, or the 'Pronoun Interpretation Problem', has serious implications for Binding Theory, the theory of how pronouns and reflexives are used. The Pronoun Interpretation Problem has been attributed to underdeveloped pragmatic skills (Chien & Wexler, 1990), insufficient working memory capacity (Reinhart, 2004, to appear) or children's inability to take into account the speaker's perspective as a hearer (Hendriks & Spenader, 2004, 2005/2006).

Two recent studies with conflicting conclusions motivate re-examining previous explanations of the Pronoun Interpretation Problem. Elbourne (2005) reviewed earlier experimental work that attributed the Pronoun Interpretation Problem found in sentences with referential subjects to pragmatic factors unrelated to grammatical knowledge of pronouns. In this earlier work, experimental data indicating that children correctly interpreted object pronouns in sentences with quantified subjects was used to argue that there is no delay in pronoun mastery.[1] Elbourne's re-examination however presents several plausible alternative interpretations of the data and suggests that children do display a delay in acquiring pronouns. Thus the challenge of explaining why reflexives are mastered early, and pronouns late, remains.

[1] For example, Chien & Wexler (1990) compared children's interpretation of pronouns and reflexives with referential subjects and quantified subjects in four experiments. In their first three experiments, children's comprehension of object pronouns and reflexives was studied in sentences with referential subjects. Children were found to be significantly better at interpreting reflexives compared to pronouns. In experiment four of their study, children's interpretation of object pronouns in sentences with referential subjects was compared to children's interpretation of object pronouns in sentences with quantified subjects. The results indicated that children were better at interpreting object pronouns correctly in sentences with quantified subjects.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

On the other hand, a recent experimental study by de Villiers, Cahillane & Altreuter (2006) shows that children's pronoun production in English is nearly perfect, strongly suggesting children know how to use pronouns.

The current study considers production as well as comprehension, and combines classic experimental questions with new questions regarding the topic structure of the materials. Children's and adults' comprehension of pronouns and reflexives in Dutch were studied with a truth-value judgment task. Three different versions of the test materials were used, allowing the influence of discourse coherence and topicality on pronoun interpretation to be studied.

The paper proceeds as follows. First, we give a brief overview of earlier findings on children's comprehension and production of pronouns versus reflexives, including an overview of the two recent papers mentioned above, Elbourne (2005) and de Villiers *et al*. (2006). Two new theories that offer an explanation of the Pronoun Interpretation Problem, Reinhart's (2004, 2006) processing account and Hendriks & Spenader's (2004, 2005/2006) grammatical account, are introduced and related to the available production data. We then look more closely at previously tested materials in terms of their discourse coherence, by using insights from Centering Theory (Grosz, Weinstein & Joshi, 1995). The conclusion is that many of the materials used for testing pronoun comprehension are inherently awkward, either failing to establish any topic or directly contradicting pragmatic rules for topics in their pronominalization. This may explain the dramatic differences between results depending on the materials used: not all studies find a 50% error in pronoun interpretation, and error rates vary roughly between 25% and 70%. When pronominalization goes counter to topic information, or when no topic information is available, children must rely on grammatical knowledge of pronoun use and the Pronoun Interpretation Problem emerges. When topic constraints are followed, children employ pragmatic knowledge and achieve results similar to reflexives. After introducing the methods used in the current study of Dutch children and adults on pronoun interpretation and production, we present the results of our study. We end with a discussion of how the results affect our understanding of Binding Theory, the relation between production and comprehension, and the influence of discourse coherence and topicality on acquisition.

The Pronoun Interpretation Problem

The pronoun interpretation delay is quite robust, and has been observed in many languages, including English (Chien & Wexler, 1990; McDaniel, Smith Cairns & Hsu, 1990; McDaniel & Maxfield, 1992; McKee, 1992; see also Grimshaw & Rosen, 1990, for a review), French and Danish

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

(Jakubowicz, 1984, 1991; Hamann, Kowalski & Philip, 1997) and Dutch (Koster & Koster, 1986; Koster, 1993; Philip & Coopmans, 1996). The delay is found with different experimental methods, including truth-value judgment tasks, picture selection tasks, and act-out tasks. The results all strongly suggest that children know how to interpret reflexives from a young age but simply guess when interpreting a pronoun until at least the age of 6;6.

The Pronoun Interpretation Problem poses a serious challenge to Binding Theory, because the principles that are argued to govern the usage of reflexives and pronouns, Principle A and Principle B, are interdefinable:

Principle A: A reflexive must be bound within its governing category.
Principle B: A pronoun may not be bound within its governing category.

A governing category is a syntactic domain within which certain syntactic relations can be defined. For our purposes, the governing category can be considered the clause in which the pronoun or reflexive occurs. Binding is the relation between an element like a reflexive or a pronoun, and another element which c-commands it, and on which it depends to fix its reference. As is apparent from the definitions given above, Principle A applies when Principle B does not, and vice versa. Therefore it is difficult to account for how children could correctly comprehend one form, yet consistently fail to comprehend the other.

Chien & Wexler (1990) consider the delay to be pragmatic because of the existence in adult language of apparent violations of Principle B, such as: 'It must be John. He_i looks like him_j'. Here, *he* and *him* are argued to display what is termed 'accidental co-reference'. If syntax correctly tags each pronoun with a different index, say *i* and *j*, then it is still possible that both these indices can be resolved to the same real-world referent under certain pragmatic conditions. An acceptable example of this would be: 'Everyone voted for him. Even he voted for him', where *he* and *him* are intended to refer to the same referent. Chien & Wexler (1990) claim that children get confused because they encounter examples of accidental co-reference. This leads them to interpret non-accidental co-reference cases incorrectly. Thornton & Wexler (1999) attribute children's errors to a similar pragmatic deficiency related to their contact with cases of accidental co-reference. Experimental work showing that children seem to correctly interpret object pronouns when the subject is a quantified phrase seemed to support this analysis. In a context with Goldilocks and the three bears, where the three bears are touching themselves, children tend to correctly answer 'no' to questions like the following:

(3) Is every bear touching her?

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Thus, in a context where there is no accidental co-reference possible, such as when the subject is quantified, children correctly interpret pronouns. Only in contexts where accidental co-reference could be possible, such as (2), do they make errors. In this analysis, what adults have and children lack is the pragmatic skill necessary to distinguish the exceptional accidental co-reference contexts from normal contexts.

In a recent survey of this and other earlier experiments, however, Elbourne (2005) convincingly argues that there is no evidence that children did show knowledge of Principle B when interpreting sentences using quantified subjects like (3). One alternative hypothesis Elbourne considers is that children perhaps simply choose the most salient referent. Careful study of the original unpublished experimental materials of Chien & Wexler, among others, showed that salience might indeed have been a confounding factor. For example, in one truth-value judgment task for sentences like (3) the drawing of Goldilocks was nearly three times as large as the drawings of each of the bears. Children's interpretation of (3), and similar sentences, can be motivated by this salience hypothesis when the entire experimental context is considered.

There are two major conclusions we could draw from this work. First, it seems quite plausible that salience, a pragmatic feature, may offer a good explanation for children's pronoun resolution strategies. This is unexpected, given that one of the major works on children's knowledge of binding principles, Chien & Wexler (1990), explicitly attributes children's delayed comprehension of pronouns to a lack of pragmatic knowledge. The second conclusion is that Elbourne's survey effectively undermines the experimental evidence that appeared to show children's knowledge of Principle B. The reinterpreted experimental data instead all point to a lack of knowledge of Principle B.

However, this latter conclusion may be premature. Two studies have examined children's pronoun production: the fairly neglected Bloom, Barss, Nicol & Conway (1994) corpus study and a more recent experimental study by de Villiers *et al.* (2006). In direct contradiction to the results we just summarized, both studies showed that children correctly produce pronouns even from a young age on, suggesting they do have knowledge of Principle B.

Comprehension versus production

Bloom *et al.* (1994) studied children's production of first person pronouns and reflexives in the CHILDES database. They looked at 75 *myself* tokens and 2,834 *me* tokens, like (4) and (5).

(4) I hit myself.
(5) Give it to me.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Children between 2;3 and 3;1 correctly produced *myself* 93·5% of the time, while they correctly produced *me* 99·8% of the time. Note that because this study only dealt with first person forms, there is never any ambiguity in the choice of a referent, which outside of embedded sentences is always the speaker.[2] This makes these results less comparable to the comprehension studies, which have focused on third person singular forms.

A recent study by de Villiers *et al.* (2006), however, looked at children's production of third person pronouns and reflexives. In this study, 37 children between 4;6 and 7;2 (mean age 6;3) were first tested for their comprehension in sequences like (6), which is very similar to Chien & Wexler's materials (the 'classic condition'), or with single sentences with an embedded clause like (7), both in a truth-value judgment task.[3]

(6) Here is Big Bird and Grover. Big Bird is touching him/himself.
(7) Papa Bear says Baby Bear is touching him/himself.

Their results for comprehension were consistent with earlier results showing a delay in pronoun interpretation relative to reflexive interpretation, with two additional results. First, pronoun interpretation with a classic Chien & Wexler type condition was only 6·6%, much lower than chance would predict. Note however that de Villiers *et al.* (2006) only tested false pronoun cases, balanced with true reflexive cases, so the results are not totally comparable with other studies. De Villiers *et al.* (2006) attribute this result to the (lack of) referential salience being established for the referent the pronoun referred to. Second, the embedded sentences improved pronoun comprehension relative to the classic condition, as they predicted, but the same condition surprisingly made comprehension of reflexives harder. The first result can be accounted for in terms of referential salience but the second is harder to explain.

Production was tested by asking children to describe the pictures in the same way as presented in the comprehension task, producing several unexpected results. First, production was significantly better than comprehension for all tasks. The production of pronouns was nearly perfect. Children never used a reflexive when a pronoun was required in the classic condition, and did so only 2·8% of the time in embedded sentences. Testing production, however, means that it becomes possible for participants to offer a range of referring expressions, not just pronouns and reflexives. Proper names or full NPs are viable alternatives to pronouns, and children

[2] Bloom *et al.* (1994) restricted the study to first person forms because the second person form in English is ambiguous between plural and singular, and the intended referent a child has for a third person form is also difficult to determine from transcripts of recordings alone.
[3] De Villiers *et al.* (2006) also tested two further conditions with quantified subjects.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

made use of these alternatives in the de Villiers *et al*. study.[4] Interestingly, the proportion of proper names produced differed radically by condition. In the classic condition, children produced a proper name in 62% of the cases. In the embedded sentences, no proper names were produced. De Villiers *et al*. (2006) again account for the difference in terms of the salience of the antecedents: Pronouns must refer to salient referents, and the embedded sentence sets up the subject of the main clause as being more salient.

This recent experimental study also leads to two major conclusions. First, similar to Elbourne, de Villiers *et al*. (2006) suggest that children are sensitive to the salience of referents. Elbourne highlights children's possible use of salience in interpretation, while the de Villiers study suggests that it plays a major role in production in the choice between a pronoun or a full NP. Second, the children's nearly flawless production of pronouns is strong evidence that they do have knowledge of Principle B, and can use it.

Thus we have two recent studies, one theoretical and one experimental, that both appeal to the idea of salience but make contradictory predictions as to children's knowledge of pronoun use. It seems that children do not show knowledge of Principle B in comprehension but do show it in production. Interestingly, there are two recent accounts of children's grasp of Binding Theory that can deal with these seemingly contradictory conclusions. Both differ from earlier accounts in that they accept all the experimental data and believe that the problem has to do with particular qualities of Principle B.

Reinhart's account (Reinhart, 2004, to appear) argues that children possess knowledge of Principle A and Principle B. Principle B under her account pertains to variable binding only, not to co-reference. Variable binding is possible only if the potential antecedent c-commands the pronoun or reflexive. Principle B determines that, in this syntactic configuration, pronouns cannot be bound in the local domain of the antecedent. In sentence (2), Principle B rules out variable binding of the pronoun *her* by the subject *Mamma Bear* because *her* is c-commanded by *Mamma Bear* in its local domain. However, there is another possibility for obtaining an anaphoric interpretation, namely co-reference. Co-reference occurs when a free pronoun picks up its reference from a referential expression in the sentence. To determine whether co-reference is permitted between the pronoun and its potential antecedent, another derivation must be constructed with a bound variable. Co-reference is only permitted between an object pronoun and a potential antecedent if the interpretation obtained by

[4] Full NPs and proper names are not usually possible alternatives to reflexives, except in special cases such as the following example taken from Fodor (1975: 134):

(i) Only Churchill remembers Churchill giving the speech about blood, sweat, toil, and tears.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

co-reference in the given context is not equivalent to the interpretation obtained by variable binding. This is not true for sentence (2), so co-reference between *Mamma Bear* and *her* is disallowed in (2), but it is true for cases of accidental co-reference. In these cases, the co-reference interpretation is clearly distinguishable from the interpretation that would be obtained by variable binding. As a result, co-reference is permitted.

According to Reinhart, co-reference is in fact a superfluous interpretive procedure (Reinhart, 2004: 117). Variable binding also yields an anaphoric interpretation, while being more economical. That superfluous operations can apply at all is an imperfection of the grammar, Reinhart claims, but these operations are needed to meet the requirements of the context. The application of apparently superfluous operations triggers what Reinhart terms reference-set computation by the parser. By constructing a reference set of pairs consisting of a derivation and an interpretation, the parser determines whether the derivation involving the superfluous operation is appropriate in the given context, or whether the resulting interpretation can be obtained more economically (for example, by a derivation involving variable binding). If a more economical (but possibly ungrammatical) derivation exists for the same interpretation, the co-referential interpretation is blocked.

Whereas adults have no problems with such reference-set computations, the processing load posed by reference-set computation is argued to exceed children's processing capacities. Because of insufficient working memory capacity, children resort to a guessing strategy when reference-set computation is needed. Reinhart argues that reference-set computation not only accounts for the adult interpretation of pronouns (and children's comprehension problems), but also for the adult interpretation of contrastive stress (and children's comprehension problems), thus offering important insights into the parallels between these phenomena. The demanding process of reference-set computation is only required for the comprehension of marked forms such as pronouns and shifted stress, and not for the comprehension of unmarked forms such as reflexives or neutral stress, or for the production of marked or unmarked forms.[5] Since children have the relevant grammatical knowledge, Reinhart's account predicts no delays in production because, as Reinhart (2004: 135) points out, language users know which meaning they intend to express. Children's

[5] 'Unmarked' forms are those forms which are considered more basic, simpler, less derived or less complex in terms of information content coded on the form, while 'marked' forms are derived, more complex and often less common forms. Determining what forms are marked and what forms are unmarked from a set of forms with similar functions is often controversial. Reinhart and Hendriks & Spenader both consider pronouns marked and reflexives unmarked, although reflexives are morphologically more complex.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

comprehension delays are argued to be caused by the demands of processing reference sets. Consequently, children will begin to correctly interpret pronouns when they have developed sufficient working memory capacity.

Hendriks & Spenader (2004, 2005/2006), in contrast, argue that the Pronoun Interpretation Problem is a problem of the acquisition of the grammar rather than a problem of the use of the parser. The delay comes about because Principle B is derived from Principle A by combining the hearer's perspective with the speaker's perspective. Young children are incapable of combining the two perspectives. As a result, they have not yet derived Principle B. Hendriks & Spenader's work is formalized in Optimality Theory (OT), which differs from many other formal architectures of grammar in that it distinguishes production and comprehension. Production and comprehension are modeled as different directions of optimization: from meaning to form, and from form to meaning, respectively. The output in either direction is the output that satisfies the violable constraints of the grammar best. In their explanation, Hendriks & Spenader (2004, 2005/2006) combine a violable version of Principle A with a revised version of a constraint family originally proposed by Burzio (1998) to account for the distribution of reflexives and pronouns cross-linguistically.

Constraints:
PRINCIPLE A: A reflexive must be bound locally.
REFERENTIAL ECONOMY: Avoid full NPs >> Avoid pronouns >> Avoid reflexives

These constraints taken together simply capture the idea that hearers and speakers prefer to interpret and use reflexives as if they were locally bound, and speakers prefer to use referentially weaker forms over referentially more informative forms. This means they will prefer reflexives over pronouns.

The violable constraint PRINCIPLE A is stronger than REFERENTIAL ECONOMY, so it is more important to satisfy PRINCIPLE A than it is to satisfy REFERENTIAL ECONOMY. As a result, reflexives will be optimally interpreted as co-referential with the subject of the clause. But the constraints say nothing about the interpretation of pronouns. REFERENTIAL ECONOMY simply encourages the speaker to prefer referentially less informative forms. With these two constraints alone, reflexives are predicted to be correctly understood and produced, yet pronouns are only predicted to be correctly produced. In comprehension, with a pronoun as input and an interpretation as output, PRINCIPLE A has no effect, as it only directs hearers to associate reflexive forms with reflexive meaning. REFERENTIAL ECONOMY also has no effect, because it only has something to say about a choice between different forms, and has no effect on interpretation. Thus pronouns are ambiguous between an

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

interpretation according to which the object is co-referential with the subject and an interpretation according to which the object is disjoint with the subject. Children's errors are argued to be the result of this ambiguity in pronoun interpretation due to the grammar. Like Reinhart, Hendriks & Spenader (2004, 2005/2006) predict errors similar to guessing for pronoun interpretation. Note that their analysis depends on the fact that some constraints have an effect in one direction of optimization only, and have no effects in the other direction. This is an inherent property of OT, which can therefore be characterized as an asymmetrical grammar.

According to Hendriks & Spenader (2004, 2005/2006), the difference between adults, on the one hand, and children until at least the age of 6 or 7, on the other hand, is that adults are able to optimize bi-directionally (cf. de Hoop & Krämer, 2005/2006). Bi-directional optimization is a formal procedure modeling the dependence of the hearer's interpretation of a given form on the speaker's choices for this form, and vice versa. Informally, bi-directional optimization proceeds as follows: Adults hear a pronoun, consider that the speaker could have produced a reflexive and recognize that a reflexive is unambiguously co-referential with the subject. They realize that if the speaker intended a co-referential interpretation, the speaker would have used a reflexive. Since the speaker did not, the speaker must have intended a disjoint interpretation. So Hendriks & Spenader's proposal does away with Principle B by deriving it as an epiphenomenon of two constraints in combination with bi-directional optimization. What on the surface seems to be a separate principle, Principle B, is argued to be derived on the basis of the grammar. How children progress from unidirectional optimization to bi-directional optimization is still unknown, but several possibilities have been put forward. The development from unidirectional to bi-directional optimization may emerge as a result of increased memory capacity (comparable to what Reinhart argues), achievement of sufficient speed of processing (Hendriks, van Rijn & Valkenier, 2007) or perhaps the development of the ability to apply second order Theory of Mind reasoning.

With respect to acquisition, Hendriks & Spenader predict that in comprehension, pronouns are ambiguous and require hearers to consider alternative forms and their meanings in order to disambiguate the meaning of the pronoun. In production, in contrast, the grammar requires that reflexives are associated with co-referential meaning, so Principle A will be violated when a disjoint input is associated with a reflexive, and a pronoun will be preferred. As a result, production of pronouns is independent of bi-directional reasoning but comprehension is not. If children are not yet capable of taking into account the opposite perspective of the speaker, it is predicted that they will have problems with the comprehension of pronouns, but not with their production. Hendriks & Spenader's account

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

thus explicitly predicts the production data. As we saw earlier, Reinhart's account makes similar predictions, but for different reasons.

Thus, these two recent theories seem to be the only theories that are able to account for the bulk of the experimental data, accepting the comprehension delay in pronouns yet allowing for correct production from an early age. However, neither account has anything to say about the role of salience in pronoun interpretation. Since the recent results from Elbourne (2005) and de Villiers *et al.* (2006) both suggest that pragmatic information such as salience might be playing a key role in pronoun use, we should look more closely at how salience relates to earlier studies of pronoun interpretation and production.

Topic continuity and natural contexts

Salience is related to coherence. Very salient items are very accessible and thus are the most natural referents to be referred to with reduced forms like pronouns. One of the most well known formal theories of local discourse coherence and its relation to referent salience is Centering Theory (CT; Grosz *et al.*, 1995). CT makes predictions about coherence in two ways. First, coherent sequences will promote the pronominalization of utterance topics, because of the so-called 'pronoun rule'. Second, CT classifies transitions between utterances based on the way in which the topic is maintained or updated, which is believed to relate to the inference load on the hearer and ultimately the perceived coherence of a discourse.

In a CT analysis, all the referents referred to in an utterance are ranked according to their prominence in an ordered set of 'forward looking centers' (Cf). Prominence in the original CT was determined by grammatical role, with subjects being more prominent than direct objects, direct objects more prominent than indirect objects, etcetera. The most prominent referent of the utterance is termed the preferred center (Cp). Further, the referent referred to in the current utterance that was most prominent among referents referred to in the previous utterance is identified as the topic of the current utterance, or backward looking center (Cb). Consider the following sequence of two utterances:

(8) Baby Bear is with Mamma Bear. Mamma Bear is washing her.

First Sentence:	Cb = ?
	Cf = [Baby Bear, Mamma Bear]
	Cp = Baby Bear
Second Sentence:	Cb = Baby Bear
	Cf = [Mamma Bear, Baby Bear]
	Cp = Mamma Bear

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

In the first sentence, Baby Bear and Mamma Bear are referents, or 'centers', and Baby Bear is the most prominent referent because it is the subject. Thus it is the preferred center (Cp) in the ordered set of forward looking centers (Cf). No topic (Cb) can be identified because this is the first sentence of a sequence. In the second sentence, Mamma Bear is the most prominent referent because it is the subject. Since both Mamma Bear and Baby Bear were referred to in the previous sentence, Baby Bear is the most prominent referent from the previous sentence that is also realized in the current sentence. This makes Baby Bear the topic of the second sentence, i.e. the Cb. The pronoun rule says that if anything is pronominalized, then the topic (Cb) is also pronominalized. We can see that this is true since only the topic is pronominalized. Other valid realizations of the second sentence would be 'Mamma Bear is washing Baby Bear' and 'She is washing her', but 'She is washing Baby Bear' would violate the pronoun rule since Baby Bear would fail to be pronominalized even though Mamma Bear was.

The sequence in (8) fulfils the pronoun rule, but not all materials that have been used in experiments to test children's knowledge of Binding Theory have been so coherent. In the original Chien & Wexler experiment, each actor is introduced in its own sentence. Whether or not the subject of the third, key sentence was introduced first or second does not seem to have been considered relevant and is not reported in the data. However, the two orders result in very different analyses according to CT. Consider (9):

(9) This is Goldilocks. This is Mamma Bear. Is Mamma Bear washing her?

First Sentence:	Cb = ?
	Cf = [Goldilocks]
	Cp = Goldilocks
Second Sentence:	Cb = ?
	Cf = [Mamma Bear]
	Cp = Mamma Bear
Third Sentence:	Cb = Mamma Bear
	Cf = [Mamma Bear, Goldilocks]
	Cp = Mamma Bear

In a CT analysis of the third sentence, Mamma Bear is the preferred center, as well as the topic (Cb). This means that if the pronoun rule is followed, hearers should expect *her*, being the only pronoun, to refer to Mamma Bear, an interpretation that would violate Principle B. If a child does not know Principle B and relies on something like the pronoun rule to aid interpretation, this sentence encourages the wrong interpretation.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Cases like (10), where Mama Bear is introduced first, get a different analysis:

(10) This is Mamma Bear. This is Goldilocks. Is Mamma Bear washing her?

First Sentence:	Cb = ?
	Cf = [Mamma Bear]
	Cp = Mamma Bear
Second Sentence:	Cb = ?
	Cf = [Goldilocks]
	Cp = Goldilocks
Third Sentence:	Cb = Goldilocks
	Cf = [Mamma Bear, Goldilocks]
	Cp = Mamma Bear

In the third sentence in (10), Goldilocks is the topic (Cb), and resolving *her* to Goldilocks would, in contrast with (9), obey the pronoun rule.

Thus it seems that at least some of the original Chien & Wexler (1990) sentences are somehow unnatural with regard to pronominalization. De Villiers *et al.* (2006) used different materials, presenting both actors in the same sentence as a conjunction of two objects. The original CT algorithm would treat both referents as equally prominent, so there should be less confusion. Strictly speaking, CT transitions can only be determined given at least a two sentence context, so we cannot evaluate (7) or the embedded sentences tested by de Villiers *et al.* (2006). But CT is not entirely clear on what is considered an utterance, with some researchers treating subordinate clauses as their own discourse segment. If each clause is separately treated as having its own centers, then the embedded sentence does obey the pronoun rule.[6]

This reinterpretation of the materials does not prove that children do or do not have knowledge of Principle B. It merely shows that an alternative explanation is possible for some of the variation found in experiments using different stimuli, namely, that children are helped by coherent materials and hindered by less natural materials that disobey constraints on discourse coherence. This may account for the range of experimental results with pronoun interpretation, where individual studies have sometimes found

[6] This is illustrated with the following example.

(i) The elephant says that the alligator is hitting him.

First clause:	Cb = ?
	Cf = [the elephant]
	Cp = the elephant
Second clause:	Cb = the elephant
	Cf = [the alligator, the elephant]
	Cp = the alligator

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

error rates lower and higher than the 50% often found. If so, then previous experimental results may not accurately reflect children's knowledge of Principle B.

AIMS OF THE PRESENT STUDY

Based on the conflicting results for pronoun comprehension versus pronoun production and the new emphasis placed on the importance of coherent discourse structure, the present study aims to investigate two main issues: (1) whether there is an asymmetry between production and comprehension of third person pronouns in Dutch children; and (2) whether experimental materials where the referent of the pronoun is a clearly established topic will improve pronoun comprehension. Additionally, we are interested in whether establishing a clear topic influences pronominalization in production.

Under the assumption that Dutch and English are similar with respect to the constraints governing the use of pronouns and reflexives (see the materials section below for a discussion of this issue), we expect Dutch children to behave like English children, displaying a delay in pronoun comprehension, yet showing perfect pronoun production. The second question addresses the effects of topic structure (contributing to discourse coherence) on comprehension. Does manipulating the topic structure improve performance in comprehension? The pronoun rule only needs one context sentence to be evaluated. If we establish the intended antecedent for the pronoun so that it becomes the topic of the target sentence, we predict this will improve children's interpretation of the pronoun. An example minimally different from (9) is (11):

(11) This is Goldilocks. Is Mamma Bear washing her?

First Sentence:	Cb = ?
	Cf = [Goldilocks]
	Cp = Goldilocks
Second Sentence:	Cb = Goldilocks
	Cf = [Mamma Bear, Goldilocks]
	Cp = Mamma Bear

In this example, Goldilocks is established in the first sentence as a potential topic of the second, target sentence. We predict that children will make fewer errors interpreting pronouns in this type of discourse context.

Coherent discourse structure may also be fundamental for natural pronoun production. Specifically, how does the clear establishment of topic influence pronominalization? If production and comprehension draw on the same body of knowledge of grammar and discourse (as is assumed by Reinhart), the production results should be similar to the comprehension results. If, on the other hand, constraints may have different effects on production and

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

comprehension (as Hendriks & Spenader argue), we may expect differences between the production results and the comprehension results.

METHODS

Participants

We tested 100 children from two local elementary schools. From this group 17 children were excluded from further analysis because they incorrectly answered two or more than two out of six control items, strongly suggesting they did not pay attention or did not understand the task. The 83 children remaining (38 boys and 45 girls) ranged in age from 4;5 to 6;6 (mean age 5;5).

Materials

Dutch, in contrast to English, has two reflexive items: *zich* 'SE' (for Simplex Expression) and *zichzelf* 'SE-self'. A recent corpus study by Smits, Hendriks & Spenader (2007) has shown that the choice between *zich* and *zichzelf* for a reflexive action is strongly correlated with how often the verb is used with a reflexive versus a non-reflexive event. Verbs that frequently present reflexive events occur with *zich*, while verbs that tend to describe events performed towards someone else occur more frequently with *zichzelf* when used reflexively, although this is not a categorical distinction.

To select our materials, we did an online test asking adult speakers of Dutch to make a forced choice between *zich* and *zichzelf* for 45 verbs.[7] We then selected six verbs where participants preferred *zichzelf* more than 86% percent of the time, and which were easy to draw as both a reflexive and a non-reflexive action. These six verbs were: *achtervolgen* 'follow', *slaan* 'hit', *kietelen* 'tickle', *tekenen* 'draw', *wijzen naar* 'point to' and *bijten* 'bite'. By using verbs with a preference for *zichzelf* we can be assured that we are choosing those verbs that do not unfairly bias children towards an interpretation that is reflexive, since *zichzelf* is used with verbs that more frequently occur with non-reflexive actions than with reflexive actions (Smits *et al.*, 2007). In fact, we are biasing children towards an interpretation that is non-reflexive, which should make any errors in pronoun interpretation more striking. Drawings were produced with one reflexive and one non-reflexive action for each verb (see Figure 1). Each drawing had two characters, both animals. In the drawing displaying a reflexive action the second character watched what the first character did.

[7] The majority of the participants were students and staff from the Artificial Intelligence Department at the University of Groningen and were recruited by email.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

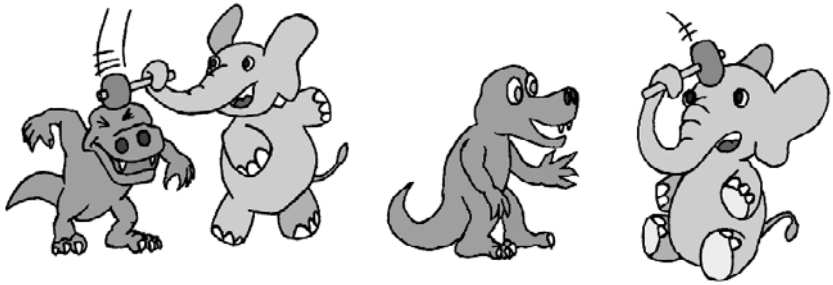


Fig. 1. Left: 'The elephant is hitting him'. Right: 'The elephant is hitting himself'.

We tested three different conditions:

Condition 1: Classic Condition

(12) Hier zie je een olifant en een krokodil. De olifant slaat hem/zichzelf.
'Here you see an elephant and an alligator. The elephant is hitting him/himself.'

Condition 2: Single Topic Condition

(13) Hier zie je een krokodil. De olifant slaat hem/zichzelf.
'Here you see an alligator. The elephant is hitting him/himself.'

Condition 3: Embedded Condition

(14) De krokodil zegt dat de olifant hem/zichzelf slaat.
'The alligator says that the elephant is hitting him/himself.'

For condition 3, we changed the pictures slightly by drawing a text balloon next to the animal speaking in the test items.

Next to these test items, control items were used to make sure that the child was paying attention, e.g. (15):

(15) Hier zie een schildpad en een tijger. De tijger zit in de kooi.
'Here you see a turtle and a tiger. The tiger is inside the cage.'

The accompanying picture either displayed this situation (three items) or not (three items), targeting respectively a 'yes' and a 'no' answer.

For the production task, the same animals from the comprehension task were used, but, crucially, displayed performing a different action than the action they performed in the comprehension test.

Procedures

Each child was tested on one of the three conditions, first for comprehension and then for production, and all six verbs were used in both parts of

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

TABLE 1. *Number of participants and their ages for each condition*

Experiment	Age group	N	Mean	Minimum	Maximum
Classic Condition	4	12	4;9	4;7	4;11
	5	11	5;5	5;0	5;11
	6	4	6;2	6;0	6;4
	Total	27	5;3	4;7	6;4
Single Topic Condition	4	9	4;9	4;7	4;10
	5	12	5;5	5;0	5;11
	6	7	6;3	6;0	6;6
	Total	28	5;5	4;7	6;6
Embedded Condition	4	8	4;7	4;5	4;11
	5	12	5;8	5;3	5;11
	6	8	6;3	6;1	6;6
	Total	28	5;6	4;5	6;6
Total		83	5;5	4;5	6;6

the experiment. The type of experimental condition the child was tested on was counterbalanced across four-, five- and six-year-olds (a one-way ANOVA with age group chosen as dependent variable and experimental condition chosen as independent variable did not show a significant effect ($F(2, 80) = 1{\cdot}62$, $p = 0{\cdot}205$; $\eta_p^2 = 0{\cdot}039$)). The number of participants and their ages for each condition are given in Table 1.

Comprehension task

The comprehension task consisted of eighteen pictures, with three 'no' control items, three 'yes' control items, six reflexive items and six pronoun items. The pictures were presented on a laptop computer with an add-on touch screen, with prerecorded narration in Powerpoint. Children were brought individually from their class to a room with the main experimenter and an assistant. The child sat in front of the test screen next to the main experimenter. The responses were audio-recorded and the assistant noted responses as well during the task. The methodology followed the methodology of de Villiers *et al*. (2006). The child was told by a puppet that the computer had been built by the experimenter and the assistant, but that the puppet believed that the computer was built wrong. The child was asked to help check if the computer was built right by looking at the pictures presented on the screen and listening to the recorded stimuli. If the recording correctly described the picture, the child could respond by saying 'yes' or 'no' and/or touching a smiley face on the computer screen. For correct sentences they were to touch a green smiley face. If the computer was wrong the child should touch the red frowning face. The children were first presented with two examples to introduce the child to the set-up. The

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

experimental stimuli, including the six control stimuli, were then presented in a semi-random order.

Production task

After the comprehension task, the child was told that because a number of the sentences did not match the pictures, it would be helpful if the child could help 'fix' the computer by recording new narrations in the child's own voice. In order to 'practice' and to show the child how she had to describe a displayed picture, four pictures were displayed accompanied by prerecorded narrations (in terms of the same condition the child was tested on in the comprehension experiment): two pictures displayed a reflexive action, and two displayed a non-reflexive action. Some children started repeating the recordings, while other children just listened. Thereafter the child was shown ten new pictures, five reflexive-target items and five pronoun-target items, and asked to describe the pictures for the computer in the microphone. The child sometimes needed prompting. For example, in the Embedded Condition the experimenter sometimes had to say 'The cat says that …' and encourage the child to begin speaking. This methodology followed de Villiers *et al.* (2006) as closely as possible to make our results comparable.

Coding

Answer possibilities in the production task are more varied than just pronouns or reflexives. A full NP is also a possible way to express a disjoint meaning, even though the training session tried to encourage pronoun production. Further, for pronouns, Dutch has both stressed and unstressed object pronominal forms that are lexically different (*hem* 'him' and its reduced version *'m* 'him', respectively). Also, besides *zichzelf* and *zich* Dutch also has a number of non-standard reflexives marked for gender, in particular *hemzelf* 'himself' and its reduced version *'mzelf* 'himself'. In order to keep track of all these answer possibilities, all answers were, as well as being recorded on the laptop as part of the experiment, also noted down by the assistant.

RESULTS

Comprehension task results

Figure 2 shows the mean percentage of correct interpretations of reflexives (white bars) and pronouns (black bars) across the three different experimental conditions: Classic Condition, Single Topic Condition and Embedded Condition.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

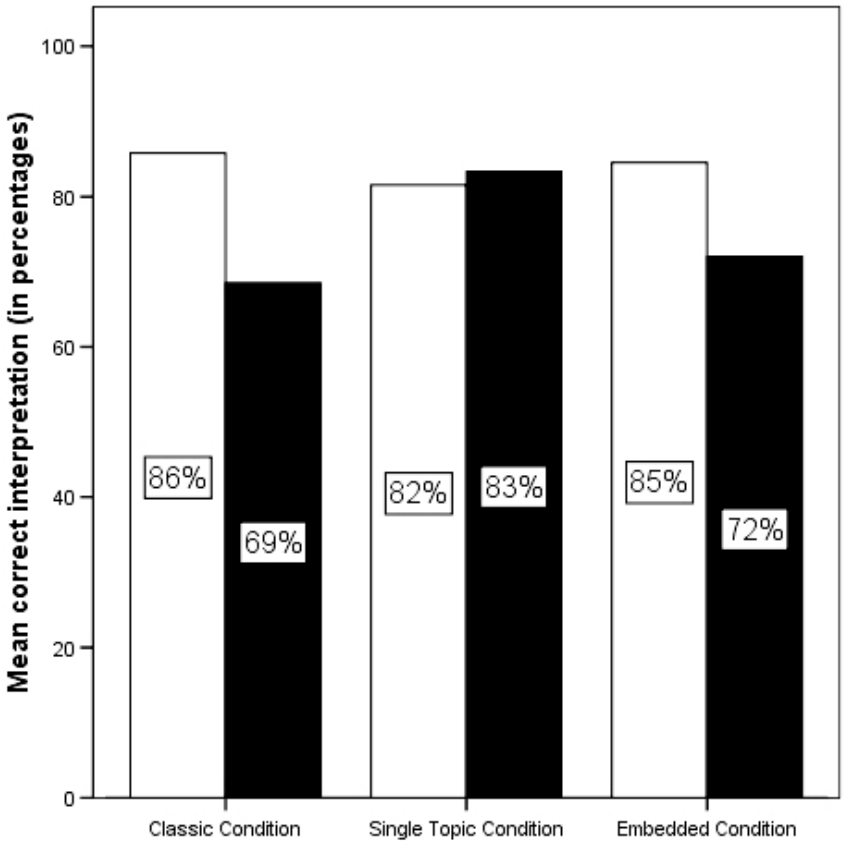


Fig. 2. Correct interpretation of reflexives (white) and pronouns (black) per experimental condition. The correct interpretation of a reflexive is a co-referential interpretation. The correct interpretation of a pronoun is a disjoint interpretation.

The children's answers were analyzed in a (2) × (3) repeated measures ANOVA. Referring expression (reflexive or pronoun) was chosen as within-subject factor and experimental condition (Classic Condition, Single Topic Condition, Embedded Condition) was chosen as between-subjects factor. This analysis shows an effect of referring expression ($F(1, 80) = 15{\cdot}43$, $p < 0{\cdot}001$; $\eta_p^2 = 0{\cdot}162$) and an interaction effect between referring expression and condition ($F(2, 80) = 5{\cdot}82$, $p = 0{\cdot}004$; $\eta_p^2 = 0{\cdot}127$).[8]

[8] To check whether proportions were appropriately used in the present and subsequent ANOVAs without violating assumptions, we did an arc-sine transformation of the data. This did not result in any different outcomes for the current and subsequent ANOVAs.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Overall, children are better at comprehending reflexives than pronouns, consistent with earlier results. But this effect was not equally strong across conditions. A one-way ANOVA shows there was a significant difference for experimental condition for sentences containing a pronoun ($F(2, 80) = 4{\cdot}74$, $p = 0{\cdot}011$; $\eta_p^2 = 0{\cdot}106$), but no significant difference for reflexives ($F(2, 80) = 0{\cdot}59$, $p = 0{\cdot}555$; $\eta_p^2 = 0{\cdot}015$). This means that the difference between conditions lies in children's comprehension of pronouns, and further that reflexives are comprehended similarly regardless of condition. Post hoc tests reveal no significant difference between the Classic Condition and the Embedded Condition, or between the Single Topic Condition and the Embedded Condition. However, there was a significant difference between the Classic Condition and the Single Topic Condition (Bonferroni with adjustment of alpha/3; $p = 0{\cdot}013$).

Following up the interaction effect between referring expression and experimental condition, the specific differences across experimental conditions become clear. Applying again Bonferroni adjustment of alpha/3 ($= 0{\cdot}16$) to correct for a possible Type I error due to the application of several *t*-tests to the same data, the results show that, for the Classic Condition, comprehension of reflexives was significantly better than comprehension of pronouns ($t(26) = 4{\cdot}52$, $p < 0{\cdot}001$). For the Embedded Condition, given the conservativity of the Bonferroni test, only a trend in this same direction was found (2-tailed *t*-test $t(27) = 2{\cdot}41$, $p = 0{\cdot}023$). There was no statistically significant difference between the comprehension of pronouns and the comprehension of reflexives for the Single Topic Condition ($t(27) = -0{\cdot}59$, $p = 0{\cdot}558$). Pronouns and reflexives were interpreted equally well in this condition, at the high level at which reflexives are interpreted in other conditions.

Comparing each age group across all conditions we found no interaction between the age of the children and their correct comprehension of either pronouns or reflexives. A $(2) \times (3) \times (3)$ repeated measures ANOVA (within-subjects factor was referring expression, between-subjects factors age group (four-year-olds, five-year-olds and six-year-olds) and experimental condition (Classic Condition, Single Topic Condition or Embedded Condition)) showed a non-significant interaction effect for referring expression, age group and experimental condition ($F(1, 74) = 0{\cdot}50$, $p > 0{\cdot}05$; $\eta_p^2 = 0{\cdot}026$).

Since visually inspecting the data in Tables 2 and 3 also suggests that the use of percentages did not result in any distortion, here and in the remainder of the paper we will only report results from analyses where the dependent variable was in percentages.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

TABLE 2. *Number (percent) of various forms produced in reflexive contexts (target : reflexive) and non-reflexive contexts (target : pronoun) per experimental condition*

Condition	Reflexive *zichzelf*	Pronoun *hem*	Full NP	Reflexive *zich*	Reflexive *hemzelf/ 'mzelf*	Pronoun *'m*	Other
Target: reflexive							
Classic Condition	117/135 (86·67)	2/135 (1·48)	3/135 (2·22)	6/135 (4·44)	3/135 (2·22)	0/135 (0)	4/135 (2·96)
Single Topic Condition	106/140 (75·71)	8/140 (5·71)	4/140 (2·86)	10/140 (7·14)	1/140 (0·71)	6/140 (4·23)	5/140 (3·57)
Embedded Condition	78/140 (55·71)	12/140 (8·57)	11/140 (7·86)	14/140 (10·00)	3/140 (2·14)	12/140 (8·57)	10/140 (7·14)
Total	301/415 (72·53)	22/415 (5·30)	18/415 (4·34)	30/415 (7·23)	7/415 (1·69)	18/415 (4·34)	19/415 (4·58)
Target: pronoun							
Classic Condition	2/135 (1·48)	12/135 (8·89)	94/135 (69·63)	1/135 (0·74)	0/135 (0)	10/135 (7·41)	16/135 (11·85)
Single Topic Condition	4/140 (2·86)	56/140 (40·00)	47/140 (33·57)	3/140 (2·14)	0/140 (0)	13/140 (9·29)	17/140 (12·14)
Embedded Condition	37/140 (26·43)	39/140 (27·86)	17/140 (12·14)	9/140 (6·42)	5/140 (3·57)	10/140 (7·12)	23/140 (16·43)
Total	43/415 (10·46)	107/415 (25·78)	158/415 (38·07)	13/415 (3·13)	5/415 (1·20)	33/415 (7·95)	56/415 (13·49)

Production task results

Table 2 presents each response type in the production task, and the number of response tokens produced for each type out of the total number of possible answers.

In the further analysis we decided to collapse responses with the standard reflexive form *zichzelf*, the simplex reflexive *zich* and the non-standard but clearly reflexive *hemzelf* and *'mzelf* into one reflexive category. We also combined the full standard pronoun response *hem* with the unstressed form *'m* and full NPs to one non-reflexive category, since any of these forms must be considered a valid choice, as confirmed by the adult study (see below for a discussion of the results of the adult study). Considering a reflexive form in a reflexive context (i.e. for a picture displaying a co-referential meaning) to be a correct production response, and a pronoun or full NP in a non-reflexive context (i.e. for a picture displaying a disjoint meaning) to be a correct production response, we found that children significantly more often produced correct reflexive forms in a reflexive context (78·55%, $SD = 20{\cdot}43$, $n = 83$) than correct non-reflexive forms in a non-reflexive context (71·81%, $SD = 29{\cdot}64$, $n = 83$) ($t(82) = 2{\cdot}16$, $p = 0{\cdot}34$). Figure 3 shows the mean percentage of correct production responses for co-referential

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

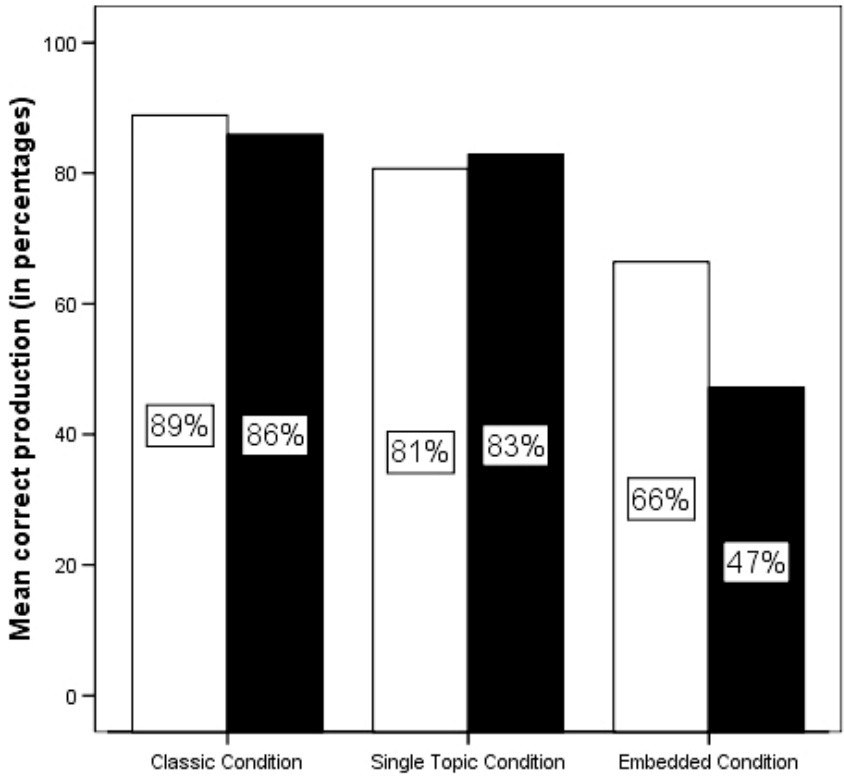


Fig. 3. Correct production of a co-referential meaning (white) and a disjoint meaning (black). The correct form for expressing a co-referential meaning is a reflexive. The correct form for expressing a disjoint meaning is a pronoun or a full NP.

meanings (white bars) and disjoint meanings (black bars) across the three different experimental conditions (Classic Condition, Single Topic Condition and Embedded Condition).

A (2) × (3) repeated measures ANOVA (referring expression (reflexive or pronoun) was chosen as within-subject factor and experimental condition (Classic Condition, Single Topic Condition, Embedded Condition) was chosen as between-subjects factor)) showed a significant interaction effect between referring expression and experimental condition ($F(2, 80) = 4{\cdot}697$, $p = 0{\cdot}012$; $\eta_p^2 = 0{\cdot}105$). Additionally, a main effect of referring expression was found ($F(1, 80) = 5{\cdot}001$, $p = 0{\cdot}028$; $\eta_p^2 = 0{\cdot}059$). No significant interactions were found with age additionally chosen as between-subject factor.

A one-way ANOVA showed an effect of experimental condition on children's production of correct reflexive forms ($F(2, 80) = 10{\cdot}53$, $p < 0{\cdot}001$; $\eta_p^2 = 0{\cdot}208$) and non-reflexive forms ($F(2, 80) = 22{\cdot}37$, $p < 0{\cdot}001$; $\eta_p^2 = 0{\cdot}359$).

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Post hoc analysis of the mean correctly produced reflexive forms between the three conditions showed a significant difference between the Classic Condition and the Embedded Condition (Bonferroni, $p<0{\cdot}001$) and between the Single Topic Condition and the Embedded Condition (Bonferroni, $p=0{\cdot}014$). Also for children's correct production of non-reflexive forms, a significant difference was found between the Classic Condition and the Embedded Condition (Bonferroni, $p<0{\cdot}001$) and between the Single Topic Condition and the Embedded Condition (Bonferroni, $p<0{\cdot}001$). Thus for production the Embedded Condition differed greatly from both other conditions.

When comparing the children's responses in non-reflexive contexts with their responses in reflexive contexts across experimental conditions, children turned out to give different responses in the Embedded Condition than in the Classic Condition and the Single Topic Condition. For the Classic Condition and the Single Topic Condition they produce correct responses in non-reflexive contexts almost as frequently as they do in reflexive contexts. However, in non-reflexive contexts in the Embedded Condition we find that children make a substantial number of errors producing a reflexive form instead. They produced an incorrect reflexive form 36·34% ($SD=21{\cdot}81$) of the time, a correct non-reflexive form 47·14% ($SD=27{\cdot}33$) of the time, and an 'other' form 16·43% ($SD=18{\cdot}10$) of the time.

Comparing production and comprehension by condition and by age

A (2) × (2) × (3) repeated measures ANOVA showed a significant interaction effect between within-subjects factors task type (comprehension or production) and referring expression (reflexive or pronoun/full NP) and the between-subject factor experimental condition (Classic Condition, Single Topic Condition or Embedded Condition) ($F(2, 80)=3{\cdot}15$, $p=0{\cdot}048$; $\eta_p^2=0{\cdot}073$). No significant interactions were found with age additionally chosen as between-subject factor.

Zooming in on each of the experimental conditions (see Table 3 for mean percentages and standard deviations) we can see where the differences lie.

Post hoc analysis with Bonferroni adjustment (alpha = 0·016) shows that, for reflexives in the Classic Condition, there was no significant difference between comprehension and production. But for pronouns, comprehension is significantly worse than production in the Classic Condition ($t(26)=-3{\cdot}59$, $p=0{\cdot}001$). In the Embedded Condition, reflexives were correctly comprehended significantly more often than they were produced ($t(27)=4{\cdot}28$, $p<0{\cdot}001$). For pronouns, comprehension was also significantly better than production ($t(27)=4{\cdot}19$, $p<0{\cdot}001$).

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

TABLE 3. *Mean percentages correct answers (and SDs) for target reflexive items and target pronoun items across experimental conditions and task type (comprehension and production)*

		Comprehension	Production
Classic Condition	Target: reflexives	85·80 (11·97)	88·89 (12·81)
	Target: pronouns	68·52 (20·84)	85·93 (22·06)
Single Topic Condition	Target: reflexives	81·55 (15·93)	80·71 (16·67)
	Target: pronouns	83·33 (17·57)	82·86 (22·25)
Embedded Condition	Target: reflexives	84·52 (16·31)	66·43 (23·76)
	Target: pronouns	72·03 (17·60)	47·14 (27·33)

We conclude that children comprehend and produce pronouns differently in the Single Topic Condition than they do in the Classic Condition and the Embedded Condition. Children comprehend and produce pronouns differently than reflexives in the Classic Condition and the Embedded Condition, but in the Single Topic Condition pronouns and reflexives are produced and comprehended equally well. In sum, we can say that children are equally good at producing and comprehending reflexives and pronouns if a coherent discourse is provided. The Embedded Condition turns out to be an anomaly in that the production of both reflexives and pronouns is affected negatively by the embedding structure.

Adults

We also tested thirty-four adults on the same production and comprehension tasks, distributed over the Classic Condition (twelve adults), the Single Topic Condition (ten adults) and the Embedded Condition (twelve adults). Correct comprehension of reflexives was 94% ($SD = 9{\cdot}07$) and of pronouns 98% ($SD = 5{\cdot}45$) across conditions (a number of errors had to do with one picture). The adults confirm that production possibilities for a disjoint meaning are much more extensive than just pronouns. Besides pronouns and reflexives adults produced, similarly to the children, full NPs. Analysis by means of a one-way ANOVA shows that the production of both pronouns and full NPs is affected by experimental condition (respectively $F(2, 31) = 4{\cdot}97$, $p = 0{\cdot}013$; $\eta_p^2 = 0{\cdot}243$ and $F(2, 31) = 16{\cdot}50$, $p < 0{\cdot}001$; $\eta_p^2 = 0{\cdot}516$). Post hoc analysis (*t*-test with Bonferroni adjustment) shows that significantly more full NPs were produced for a disjoint meaning in respectively the Classic Condition (mean = 78·33%; $SD = 30{\cdot}10$%) and the Single Topic Condition (mean = 52%; $SD = 44{\cdot}44$%) compared to the Embedded Condition (mean = 5%; $SD = 17{\cdot}32$%) ($p < 0{\cdot}001$ and $p = 0{\cdot}005$), mirroring the children's production responses. Adults also produced

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

significantly more pronouns for a disjoint meaning in the Embedded Condition (mean = 55%; *SD* = 17·32%) compared to the Classic Condition (mean = 16·67%; *SD* = 29·34) ($p = 0{\cdot}012$). No differences were found in adults' production of pronouns between the Embedded Condition and the Single Topic Condition.

DISCUSSION

The present study had two primary aims: to determine if the asymmetry between production of pronouns and comprehension of pronouns found for English children is also found for Dutch children, and to determine whether the discourse coherence of the experimental materials influences how children perform on the comprehension task. Additionally, we were also interested in whether topicality cues or their lack influence how children perform on the production task.

Asymmetrical comprehension and production

We found that children's correct production of Dutch pronouns is significantly better than their correct comprehension of these pronouns when the experimental materials fail to follow pragmatic constraints on topic pronominalization, as in the Classic Condition. This asymmetry is in accordance with the results found by numerous pronoun interpretation studies and the production studies of Bloom *et al*. (1994) and de Villiers *et al*. (2006) for English, and is also in line with the predictions made by Hendriks & Spenader's (2004, 2005/2006) hypothesis that the grammar is asymmetrical and has different effects in comprehension and production.

In the production part of our study, several forms were considered correct responses for expressing disjoint meanings, i.e. full NPs as well as pronouns, unlike in comprehension, where only pronouns and reflexives were tested. We collapsed the category of pronouns with other non-reflexive forms, such as full NPs. However, this does not mean we cannot compare production and comprehension: we can compare incorrect production with incorrect comprehension and note that in the Single Topic Condition and the Embedded Condition the children knew how to express a disjoint meaning, as they did not in general make errors by using a reflexive form for a disjoint meaning. This latter error would be plausible, given that in comprehension children make errors by treating the pronoun as if it were a reflexive form.

In production, the children in our study frequently produced full NPs rather than pronouns. Because our adult controls also produced full NPs in 78·33% (*SD* = 30·10) of the time in the Classic Condition and 52% (*SD* = 44·44%) of the time in the Single Topic Condition, we counted these

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

as correctly produced forms. These results seem to be consistent with the results of de Villiers *et al.* (2006), who also found that their Classic Condition led to avoidance of pronouns but that in the Embedded Condition no proper names were produced. What could be the reason that speakers choose to produce a full NP rather than a pronoun? And how should the effects of context be explained? One possibility is that the Classic Condition, in which the two potential referents are introduced as a conjunction, establishes some kind of contrast between the two potential referents. Because pronouns do not like to carry (contrastive) stress, this may result in the preference of a full NP over a pronoun in the Classic Condition. No such contrast is established in the Embedded Condition, which may explain why pronouns are not avoided here. Because there is an implicit contrast between the topic (Cb) and the preferred center (Cp) in the Single Topic Context, this could explain why this condition patterns somewhere in between the other two conditions. However, we did not find a more frequent use of the unstressed variant of the pronoun (*'m*) rather than the full pronoun (*hem*) in the Embedded Condition as compared to the other conditions, which would be expected if the avoidance of contrastive stress were the crucial factor.

Influence of discourse coherence on comprehension

Comparing the results of the Classic Condition with the Single Topic Condition in comprehension, we found that a natural single topic context improves the comprehension of pronouns significantly. The influence of discourse coherence and topicality on comprehension is quite strong and has the effect that the Pronoun Interpretation Problem is dissolved completely. Only if both possible referents for a pronoun are equally salient (as in the non-natural context used in Chien & Wexler's classic condition, and in many other studies) do we find the frequent incorrect selection of a co-referential meaning that is illustrative of the Pronoun Interpretation Problem. Since topicality has such a large effect on pronoun interpretation, it is conceivable that differences in discourse coherence and topicality of the experimental materials used may explain some of the variation in error rates that has been found in earlier studies of the Pronoun Interpretation Problem. Whereas the comprehension of pronouns is highly dependent on the discourse, the comprehension of reflexives, on the other hand, is entirely independent of discourse coherence and topicality.

It is interesting to note that the Pronoun Interpretation Problem dissolves when very simple measures to establish a topic are used. An elaborate context does not seem to be required. This is a different conclusion than was reached in a recent study by Conroy, Takahashi, Lidz & Phillips (2006), where it is argued that an elaborate context is required, in which the

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

accessibility and felicity of the co-referential interpretation and the disjoint interpretation are balanced.

Influence of discourse coherence on production

A comparison of the results of the Classic Condition with the Single Topic Condition in production reveals that the production of pronouns and reflexives is independent of topicality and discourse coherence. This suggests that pragmatic structure aids pronoun comprehension but does not affect pronoun production.

How can this selective effect of topic structure be explained? A possible explanation, which is in line with the asymmetrical grammar hypothesis of Hendriks & Spenader (2004, 2005/2006), is that the effects of discourse are caused by a constraint saying that pronouns refer to topics (let us call this constraint PRO→TOP). Such a constraint will guide pronoun interpretation towards the topic in single topic contexts. PRO→TOP would of course encourage the hearer only to consider pronouns with features compatible with the topic. If no unique topic can be identified (as in the Classic Condition), candidate meanings cannot be evaluated with respect to this constraint and a guessing pattern emerges.

But why does PRO→TOP not have any effects in production? Let us look at the relevant constraints of the grammar in more detail. If the constraint PRO→TOP is weaker than the constraint family REFERENTIAL ECONOMY, which prefers reflexives to pronouns and pronouns to full NPs, the constraint PRO→TOP is predicted not to have any effect in production. If the meaning to be expressed is a disjoint meaning, reflexives are ruled out as a result of the strongest constraint PRINCIPLE A. In this situation, the constraint REFERENTIAL ECONOMY will be decisive in production. As a result of the application of this constraint, pronouns will be preferred to full NPs. Because the constraint PRO→TOP is weaker than the constraint REFERENTIAL ECONOMY, it is more important to satisfy REFERENTIAL ECONOMY than it is to satisfy PRO→TOP. Hence, pronouns will be produced irrespective of their topicality.

In comprehension, on the other hand, the constraint REFERENTIAL ECONOMY does not have any effects because this constraint merely selects a form from a set of potential forms and does not influence their interpretation (see also Hendriks & Spenader's explanation of the Pronoun Interpretation Problem above). As a result, if a pronoun is encountered, the weaker constraint PRO→TOP will be decisive and select the discourse topic as the antecedent for the pronoun.

The constraint PRO→TOP thus explains children's dependence on topicality for their comprehension of pronouns as well as children's lack of dependence on discourse structure for their production of pronouns.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Children's comprehension of pronouns is sensitive to discourse structure because the constraints PRINCIPLE A and REFERENTIAL ECONOMY underdetermine the interpretation of pronouns. For adults, in contrast, pronouns are not ambiguous as a result of bi-directional optimization. Consequently, their comprehension of pronouns is not dependent on discourse structure.

In our explanation of the data, we simply stipulated the constraint PRO→TOP. However, this constraint has been independently motivated and has been argued to explain why children display a production delay with respect to the choice between a pronoun and a full NP as the subject of the sentence (Hendriks, Englert, Wubs & Hoeks, to appear). Until the age of 6, children tend to overuse pronouns in subject position and also frequently produce unrecoverable pronouns (Karmiloff-Smith, 1985). Only if language users optimize bi-directionally and take into account the hearer as a speaker will they be able to avoid unrecoverable forms. This explanation is in line with the explanation presented here, but seems incompatible with Reinhart's (2004, to appear) processing account, where it is explicitly stated that reference-set computation does not apply in production.

Thus, children's and adults' pattern of pronouns and reflexives follow from a straightforward extension of Hendriks & Spenader's (2004, 2005/6) asymmetrical grammar account. It is unclear whether Reinhart's (2004, to appear) processing account, which attributes children's errors to their failure to perform reference-set computation rather than to lack of knowledge of the grammar or lack of pragmatic skills, can be extended to be able to handle the observed effects of discourse coherence equally well.

The results of the Classic Condition and the Single Topic Condition nicely follow from the preference for pronouns to refer to the topic. The Embedded Condition results are more challenging to analyze. As mentioned in the discussion of the background to this study, whether or not this condition is considered to establish a clear topic is dependent on the still disputed question of what should be considered an utterance in Centering Theory. Specifically, can a subordinate clause be its own utterance? With respect to the comprehension results, the Embedded Condition seems to pattern with the Single Topic Condition in that there was no significant difference between the interpretation of pronouns and the interpretation of reflexives. In production, the accuracy of non-reflexive and reflexive forms in the Embedded Condition were again comparable, but both were surprisingly poor when compared to production in the other two conditions. It seems fairly clear that the Embedded Condition is in some way more complex than the Classic Condition and the Single Topic Condition. Further analysis is only possible with more data, so we leave this for future research.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

CONCLUSIONS

The present study investigated Dutch children's production and comprehension of pronouns and reflexives. It found a clear asymmetry between pronoun production and pronoun comprehension, with production being more or less perfect whereas comprehension is significantly worse. This asymmetry confirms Hendriks & Spenader's hypothesis that the grammar is asymmetrical and has different effects in comprehension and production. Furthermore, the present study found that the presence of a clear topic influences children's comprehension of pronouns, in effect dissolving the Pronoun Interpretation Problem. In object pronoun production, on the other hand, which was already adult-like in most children, topicality did not have any significant effects. The strong and selective effects of topicality emphasize the need to take discourse coherence seriously in acquisition studies.

REFERENCES

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

Burzio, L. (1998). Anaphora and soft constraints. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky (eds), *Is the best good enough? Optimality and competition in syntax*, 93–13. Cambridge, MA: MIT Press.

Chien, Y.-C. & Wexler, K. (1990). Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* **13**, 225–95.

Conroy, A., Takahashi, E., Lidz, J. & Phillips, C. (2006). Equal treatment for all antecedents: How children succeed with Principle B. Unpublished manuscript, University of Maryland.

de Hoop, H. & Krämer, I. (2005/2006). Children's optimal interpretations of indefinite subjects and objects. *Language Acquisition* **13**, 103–123.

de Villiers, J., Cahillane, J. & Altreuter, E. (2006). What can production reveal about Principle B? In K. U. Deen, J. Nomura, B. Schulz & B. D. Schwartz (eds), *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America*, 89–100. University of Connecticut Occasional Papers in Linguistics **4**.

Elbourne, P. (2005). On the acquisition of Principle B. *Linguistic Inquiry* **36**, 333–65.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Grimshaw, J. & Rosen, S. T. (1990). Knowledge and obedience: The developmental status of the Binding Theory. *Linguistic Inquiry* **21**, 187–222.

Grosz, B., Weinstein, S. & Joshi, A. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* **21**, 203–225.

Hamann, C., Kowalski, O. & Philip, W. (1997). The French 'Delay of Principle B' Effect. In E. Hughes, M. Hughes & A. Greenhill (eds), *Proceedings of the 21st annual Boston University Conference on Language Development*, 205–219. Somerville, MA: Cascadilla Press.

Hendriks, P., Englert, C., Wubs, E. & Hoeks, J. (to appear). Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information*.

Hendriks, P. & Spenader, J. (2004). A bidirectional explanation of the pronoun interpretation problem. In P. Schlenker & E. Keenan (eds), *Proceedings of the ESSLLI '04 Workshop on Semantic Approaches to Binding Theory*. Nancy, France. http://www.linguistics.ucla.edu/people/schlenker/BindingWorkshop.pdf

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press

Hendriks, P. & Spenader, J. (2005/2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* **13**, 319–48.
Hendriks, P., van Rijn, H. & Valkenier, B. (2007). Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua* **117**, 1879–96.
Jakubowicz, C. (1984). On markedness and binding principles. *Proceedings of the North Eastern Linguistics Society* **14**, 154–82.
Jakubowicz, C. (1991). L'acquisition des anaphores et des pronoms lexicaux en français. In J. Guéron & J. Y. Pollock (eds), *Grammaire générative et syntaxe comparée*, 229–53. Paris: Editions du CNRS, Sciences du Langage.
Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes* **1**, 61–85.
Koster, C. (1993). *Errors in anaphora acquisition*. Unpublished doctoral dissertation, Utrecht University.
Koster, J. & Koster, C. (1986). The acquisition of bound and free anaphora. Paper presented at the 11th Annual Boston University Conference on Language Development, Boston, MA.
McDaniel, D. & Maxfield, T. L. (1992). Principle B and contrastive stress. *Language Acquisition* **2**, 337–58.
McDaniel, D., Smith Cairns, H. & Hsu, J. R. (1990). Binding principles in the grammar of young children. *Language Acquisition* **1**, 121–39.
McKee, C. (1992). A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition* **2**, 21–54.
Philip, W. & Coopmans, P. (1996). The double Dutch Delay of Principle B Effect. In A. Stringfellow, D. Cahana-Amitay, E. Hughes & A. Zukowski (eds), *Proceedings of the 20th Boston University Conference on Language Development*, 576–87. Somerville, MA: Cascadilla Press.
Reinhart, T. (2004). The processing cost of reference-set computation: Acquisition of stress shift and focus. *Language Acquisition* **12**, 109–155.
Reinhart, T. (2006). Interface strategies: *Optimal and costly computations*. Cambridge, MA: MIT Press.
Reinhart, T. (to appear). Processing or pragmatics? Explaining the coreference delay. In T. Gibson & N. Pearlmutter (eds), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.
Smits, E. J., Hendriks, P. & Spenader, J. (2007). Using very large parsed corpora and judgment data to classify verb reflexivity. In A. Branco (ed.), *Anaphora: Analysis, algorithms and applications*, 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, LNAI (Lecture Notes in Artifical Intelligence) #4410, 77–93. Berlin Heidelberg: Springer-Verlag.
Thornton, R. & Wexler, K. (1999). *Principle B, VP ellipsis and interpretation in child grammar*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S0305000908008854 Published online by Cambridge University Press