# Comprehension and production of subject pronouns: Evidence for the asymmetry of grammar

*Charlotte Koster, John Hoeks and Petra Hendriks*

## 1. Introduction[1]

One central property of human language is that, in general, adult speakers can understand whatever they produce and adult listeners can produce whatever they understand. This observed symmetry between production and comprehension might not, however, be an inherent property of grammar. It is well-known that children sometimes understand meanings that they do not yet correctly produce. Recent studies have also provided evidence that children sometimes correctly produce forms that they do not yet understand. Such delays in comprehension have been found in areas as diverse as object pronouns, indefinite noun phrases, prosody and contrastive stress, word order and structural attachment (see Hendriks and Koster 2010, for discussion). Many of these delays occur relatively late in acquisition, after age 5 or even later, resulting in a gap between correct production and correct comprehension that can span several years.

Such asymmetries in language acquisition present a real challenge to rule-based systems of grammar. If children know a rule of grammar, they should be able to use this rule in production and comprehension alike. So, how can asymmetries between comprehension and production in child language be explained? Taking object pronouns as an example, previous accounts of children's acquisition have attempted to explain comprehension errors as resulting from a lack of pragmatic knowledge necessary to distinguish exceptional cases from the standard pattern (Thornton and Wexler 1999), from insufficient working memory capacity for the parser to compare alternative forms and meanings (Reinhart 2006), or as an experimental artifact due to an unbalanced context (Conroy et al. 2009). One common denominator in all these accounts is that they fail to provide a detailed explanation of children's successful *production* of object pronouns.

The general solution of arguing that asymmetries arise as a result of difficulties at the interface with other linguistic modules and do not reflect core properties of the grammar also makes it difficult to explain why certain delays occur only in particular syntactic environments or only in par-

ticular languages. The pronoun interpretation problem, for example, does not occur in Italian (McKee 1992) and, in Spanish, it occurs only in Exceptional Case Marking constructions (Baauw and Cuetos 2003). The account of object pronoun acquisition given by Hendriks and Spenader (2005/6), based on the Asymmetric Grammar Hypothesis, is capable of predicting when delays will occur and when correct production and comprehension will be achieved. Delays are argued to be a direct result of the grammar itself being inherently asymmetric and may occur in either production or comprehension, depending on local constraint properties of the grammar. Adult language tends to be symmetric because adult speakers take into account their potential listeners in their use of grammar, and vice versa, and adapt their forms and meanings accordingly.

The aim of the present study is to extend the Asymmetric Grammar Hypothesis to anaphoric subjects in discourse and to test the prediction that the acquisition of subject pronouns may be delayed because of asymmetries in the grammar. The asymmetric grammar predicts a comprehension delay with object pronouns, but a production delay with subject pronouns. Speakers should avoid using a subject pronoun when this pronoun will be resolved by a listener as referring to the incorrect antecedent, for example following a topic shift. Whether or not children produce unrecoverable subject pronouns is investigated in a discourse production task using picture storybooks. In addition, the same participants are given a comprehension task to investigate their understanding of pronouns at the end of discourses that either do or do not involve a topic shift. As listeners, will they correctly interpret the topic shift marking in the discourses? This study also investigates the role of working memory in discourse production and comprehension by including a working memory task.

## 2. Asymmetric Grammar Hypothesis

According to the Asymmetric Grammar Hypothesis, asymmetries in acquisition are the result of inherent properties of the grammar. Although not compatible with rule-based systems of grammar, this explanation straightforwardly follows from constraint-based systems such as Optimality Theory (e.g., Prince and Smolensky 2004). In Optimality Theory, production is viewed as a process of optimization starting with a particular input meaning and yielding the optimal output form to express that meaning. Comprehension proceeds in the opposite direction and starts with a particu-

lar input form to yield the optimal output meaning for that form. The grammar is made up of constraints which are potentially conflicting. Resolution of conflicts is achieved by ranking constraints in a particular order of strength and permitting a stronger constraint to have priority over a weaker one. The form or meaning that best satisfies the set of constraints is considered the optimal form or meaning and hence is realized in the language. Because constraints are output-oriented and compare competing output candidates without necessarily considering the input, constraints can be direction-sensitive. Consequently, different constraints may be effective in production and comprehension (Smolensky 1996). For example, a constraint preferring reflexive forms to pronoun forms irrespective of their meanings will be effective in production only. In comprehension, this constraint does not have any effect because comprehension does not involve choosing between candidate forms (the form is already given as produced by the speaker), but only between candidate meanings. Crucially, even with the constraints properly ranked, asymmetries may still arise between production and comprehension.

In order to achieve the symmetry between production and comprehension observed in adult language, a person must consider both the speaker as a listener and the listener as a speaker. This is formally modeled as bidirectional optimization (Blutner 2000). Bidirectional optimization combines the two directions of unidirectional optimization: From meaning to form, and from form to meaning. Consequently, in comprehension, a listener may need to select a different meaning if the initially selected meaning would have been better expressed by a speaker with another form. In production, a speaker may need to select a different form if the initially selected form would be interpreted by a listener differently than the speaker had intended.

In the framework of bidirectional Optimality Theory, language acquisition is two-fold, including both constraint ranking and bidirectional optimization. First, a very young child must discover the proper constraint ranking involved in a particular linguistic phenomenon. For example, Cannizzaro (2010) investigated subject-object word order in young Dutch children and argues that their errors in comprehension but not production may be the result of constraint mis-ranking.

Even when constraints of a particular language phenomenon are properly ranked, production and comprehension can still be asymmetric in the child's language. The second step towards adult symmetry involves achieving bidirectional optimization over both form and meaning. This is not a general cognitive skill that is achieved at a specific moment in development

and then applicable to all aspects of language. Rather, it is dependent upon the constraints involved in a particular language phenomenon and the amount of exposure the child has to this phenomenon. For example, regarding comprehension of object pronouns, Hendriks, van Rijn and Valkenier (2007) demonstrated, using computational simulations, that this transition from unidirectional to bidirectional optimality can be achieved through sufficient exposure to the relevant forms. For further discussion of language acquisition in Optimality Theory, see Fikkert and De Hoop (2009).

In this paper, the focus is on this second step of acquisition in which children are assumed to have acquired the correct constraint ranking but are not yet adult-like because they are not capable of bidirectional optimization. The acquisition of object pronouns illustrates this second step. Hendriks and Spenader (2005/6) argue that children's original unidirectional and asymmetric performance on object pronouns follows from the interaction of two violable constraints. The first constraint expresses a preference for reflexives to be bound by the local subject, and can be seen as a violable version of Principle A of Binding Theory. This constraint pertains to reflexives only and is not sensitive to the direction of use, having the same effect on production and comprehension. A second, weaker constraint is direction-sensitive because it is formulated as a constraint on forms, and only has an effect on production. This constraint expresses a preference for referentially more economical forms over referentially less economical forms (see Burzio 1998 and Gundel, Hedberg and Zacharski 1993, for variants of this constraint hierarchy). So, reflexives are preferred over pronouns, and pronouns are preferred over full NPs. Because Principle A is the stronger of the two constraints, reflexives are never interpreted as disjoint to the local subject in comprehension, and a disjoint meaning is never expressed by a reflexive in production. Furthermore, the Referential Economy constraint has the effect that a reflexive, rather than a pronoun, is used to express a coreferential meaning. These two constraints predict children's correct production of both reflexives and object pronouns. With respect to the comprehension of reflexives and pronouns, however, the same two constraints predict ambiguity for object pronouns. Because Principle A and Referential Economy do not distinguish between possible meanings for pronouns, both the coreferential and the disjoint meaning come out as optimal. Because no Principle B specifically pertaining to pronouns is assumed, based solely on these two constraints, children are predicted to correctly interpret reflexives but to display a guessing pattern with respect to the interpretation of object pronouns

Under this account, asymmetries between comprehension and production arise as a result of the particular constraints of the grammar. These asymmetries disappear when speakers and listeners optimize bidirectionally. A listener who encounters a sentence with an object pronoun must not only select the optimal meaning for this form, but should also check whether he, as a speaker, would have produced this pronoun when intending to express this meaning. If a listener selects a coreferential meaning for a potentially ambiguous pronoun, he will then discover that this coreferential meaning is best expressed using a reflexive. Because this form (reflexive) is different from the heard form (pronoun), the coreferential meaning cannot be the meaning the speaker intended to express. As a result, the listener should block the coreferential meaning for the object pronoun and select the disjoint meaning instead (see Hendriks and Spenader 2005/6, for a more formal account of bidirectional optimization applied to object pronouns). Thus, the pattern that object pronouns receive a disjoint meaning (i.e., Principle B of Binding Theory) is a derived effect resulting from bidirectional optimization, and is not represented directly in the constraints.

## 3. Acquisition of Subject Pronouns

The acquisition of referring expressions and topic structure in discourse has been the focus of research for quite some time. In production, a pioneer in the field is Karmiloff-Smith (1981) who investigated storybook narratives by French and English children between 4 and 9 years old. She found that the youngest children use pronouns indiscriminately, with no consideration for their listeners. As the children grow older, pronominalizations become governed by a broader thematic organization of their narratives. The children have to first achieve a higher level of intrasentential discourse organization before the role of discourse topic can begin to develop. The production of pronouns in discourse has been extensively investigated by other researchers as well (for example, Hickmann and Hendriks 1999; Wigglesworth 1997). In comprehension, there is also a tradition of discourse pronoun research in children of various ages (for example, Tyler 1983; Song and Fischer 2005). How children interpret pronouns in discourse context has been shown to be affected by their developing perception of discourse structure and their interpretation of the prominence of possible antecedents.

The present paper investigates both production and comprehension of discourse pronouns in the same group of children. The study has a narrower

scope than general discourse structure in that it is specifically oriented toward a speaker's choice of form in relation to a listener's interpretation of this form. The Asymmetric Grammar Hypothesis is extended here beyond object pronouns and reflexives at sentence level to subject pronouns in discourse. This is possible because in Optimality Theory, constraints express general statements about language: Their ranking and application is not construction-dependent. If Referential Economy constrains the possible forms of a language by selecting the correct object pronoun form in potential binding environments within sentences, this constraint is expected to be effective in other areas of the language as well, for example with respect to subject pronouns at the discourse level. If a speaker wishes to select the optimal form for referring to a previously mentioned referent in subject position, based on the Referential Economy constraint alone the speaker should prefer reflexives to pronouns and pronouns to full NPs. But because the stronger constraint, Principle A, also applies in this case, a reflexive is actually a suboptimal form. A reflexive in subject position cannot be bound by the local subject. As a result, a speaker will resist using a reflexive and will prefer using a pronoun over a full NP in subject position (cf. Hendriks et al. 2008).

In discourse, an additional constraint is needed to explain the reference of subject pronouns. If the speaker intends the subject pronoun to refer to the most salient referent in the discourse, that is, the current discourse topic, it will be interpreted correctly by the listener. But if the pronoun does not refer to the most salient referent, its meaning will be unrecoverable for the listener. To explain why listeners are selective in their interpretation of pronouns in discourse, a force must be present in the language that restricts the interpretation of subject pronouns to the most salient entity in the immediate linguistic discourse, that is, to the discourse topic. This requirement has been proposed in the literature in several guises, for example as the pronoun rule in Centering Theory (Grosz, Joshi and Weinstein 1995). In this paper, we implement this requirement as a constraint of grammar and refer to it as ProTop: "Pronouns refer to the discourse topic" (cf. Beaver 2004; Hendriks et al. 2008; see also Hendriks et al. 2010). As listeners will preferably interpret a pronoun as referring to the discourse topic, bidirectionally optimizing speakers will take the listener's perspective into account and use a referentially more expensive, but recoverable full NP when they need to refer to a non-topic referent. Similarly, as speakers will preferably use a pronoun, bidirectionally optimizing listeners will interpret the speak-

er's choice of a suboptimal form such as a definite NP as signaling reference to a non-topic.

The Asymmetric Grammar Hypothesis makes predictions in relation to children and their production and comprehension of subject pronouns and topics in discourse. Since bidirectional optimization combines the speaker's direction of optimization with the listener's, it is more complex than unidirectional optimization. Achieving bidirectional optimization may be dependent on cognitive resources such as processing speed or working memory. In the latter case, the size of a child's working memory may be linked to whether the child is able to apply bidirectional optimization or not.

If children apply the constraints Referential Economy and ProTop unidirectionally, the following predictions about production and comprehension can be made (see Hendriks et al. 2008, and Wubs et al. 2009, for a more formal account). In production, a unidirectionally optimizing speaker will use referentially more economical subject pronoun forms for previously mentioned referents, regardless of their current discourse topic status. Specifically, children are predicted to overproduce subject pronouns. In comprehension, a unidirectionally optimizing listener will fail to interpret a full NP as signaling reference to an entity which is not at that moment the discourse topic. For this listener, a full NP can equally well refer to the discourse topic as to a non-topic, because neither of these meanings violates the two constraints.

Whereas the constraints of the grammar predict that children correctly produce object pronouns, the same grammar predicts children to make errors when producing subject pronouns. So, the Asymmetric Grammar Hypothesis not only predicts the existence of asymmetries, but also the direction of these asymmetries. In the case of subject pronouns, the bidirectional adult grammar is characterized by symmetry: Adults consider the linguistic perspective of others and pair subject pronouns with discourse topics and full NPs with non-topics

In this study, whether or not children produce unrecoverable subject pronouns is investigated in a discourse production task using picture storybooks, each featuring two characters. The storybooks are structured in such a way that it encourages speakers to switch the topic twice. The first character starts out as the most likely initial topic of the speaker's discourse, but halfway through the story the second character should become the new discourse topic. The critical question is how speakers reintroduce the first character at the end of the storybook. Will they use a full NP or will they use a pronoun? Crucially, since until that moment, the second character is the current discourse topic, a pronoun will be

interpreted incorrectly by a listener as continuing to refer to this second character.

In addition to the production task, participants are given a comprehension task based on recorded discourses that either do or do not involve a topic shift. The topic shift is marked by a full NP in subject position. The experimental question here is whether listeners will be able to use this linguistic marking of topic shift to interpret a potentially ambiguous pronoun at the end of the discourse.

In order to investigate the role of working memory capacity in relation to the participants' success or failure to optimize bidirectionally, a simple word-repetition memory task is also administered. The experimental question here is whether there is a relation between working memory and success/failure to bidirectionally optimize, that is, to consider the linguistic needs or intentions of a conversational partner.

Summed up, the predictions of the Asymmetric Grammar Hypothesis tested in this study are:

1) Children prefer to produce subject pronouns over full NPs, even when referring to non-topics.
2) Children do not interpret full NPs as a topic shift signal and may therefore interpret a subsequent subject pronoun incorrectly.
3) Adult-like pronoun use results from bidirectional optimization, which in children is related to a larger working memory capacity.

## 4. Method

### 4.1. Participants

This study included 31 children (15 girls, 16 boys), all of whom were typically developing native speakers of Dutch, attending kindergarten classes in a public school (mean age = 5;6 years, range = 4;3 – 6;5). Two additional children were not included in the analyses because they produced almost no full utterances. The study also included an adult control group of 23 native Dutch speakers (11 women, 12 men, mean age = 24;7 years, range = 20;7 – 30;9).

### 4.2. Materials production task

While looking at a storybook with one picture per page and six pictures per story, participants described what was happening in the picture story. The

storybook format, with individual pictures, was chosen over films. With the presentation and description of one picture at a time, the child is encouraged to describe each scene in order. Also the overall memory load of the child's task is lighter than if the child would first see a complete cartoon and later describe it, or if the child's description had to match a cartoon's presentation tempo. In this study, each participant saw the same four storybooks, all with the same internal structure designed to elicit topic shifts from the storyteller. One of the storybooks is shown in Figure 1.

*Figure 1.* An example of a picture storybook in the production task.

The first and second pictures show the first character only. A speaker has to decide how to introduce this character and how to continue referring to it. Because the first character is the only possible referent at this point, it is the most likely initial topic of the discourse. In the third picture, the second character enters the story. In the next two pictures, this character is shown performing an action while the first character is passive. Again, a speaker must decide how to introduce the second character and how to continue referring to it. At this point in the story, the second character has become highly prominent and the speaker is likely to initiate a topic shift and establish the second character as the new topic. The final picture of the story only shows the first character. As a consequence, a speaker cannot maintain the second character as the topic and is expected to initiate a topic shift again, switching back to the original topic, the first character.

Although much can be said about the introduction and maintenance of referents and topics in discourse, this study focuses on the data with respect

to topic shift only. In particular, it concentrates on the second shift from the new topic (second character) back to the previous topic (first character), which should occur at the end of the storytelling. A speaker who takes into account the listener's perspective will use a full NP. On the other hand, a speaker who does not, or cannot, take into account the listener's perspective will simply use the more economical, but unrecoverable pronominal form.

### 4.3. Materials comprehension task

In the comprehension task, all participants listened to the same pre-recorded stories about two characters of the same gender. The final sentence of each story contained a potentially ambiguous subject pronoun, matching the gender of the two characters. At the end of each story, participants were asked a question about which of the two characters had done something. In total, there were eight stories composed of six sentences each. The structure of the stories differed: Four included a topic shift halfway through the story and the other four did not. The body of the stories with topic shift was designed in such a way that they were parallel in structure to the production stories (all designed to elicit a topic shift). An example of a comprehension story with topic shift is given in Figure 2.

1 *Een schoonmaakster wil de eendjes gaan voeren.*
'a cleaning-lady wants to go feed the ducks'
2 *Ze haalt het oude brood uit de broodtrommel.*
'she gets the old bread out of the breadbox'
3 *Ze vraagt aan een juf om mee te gaan.*
'she asks a teacher(FEM) to go along'
4 *De juf scheurt de broodjes van de schoonmaakster in stukjes.*
'the teacher(FEM) tears the cleaning-lady's bread into pieces'
5 *En dan geeft de juf het brood van de schoonmaakster aan de eendjes.*
'and then the teacher(FEM) gives the cleaning-lady's bread to the ducks'
6 *Ze vindt eendjes hele lieve diertjes.*
'she thinks ducks are very sweet animals'

Comprehension question:
*Wie vindt eendjes hele lieve diertjes?*
'who thinks ducks are very sweet animals?'

*Figure 2.* An example of a recorded comprehension story with topic shift.

In the topic shift condition, the first character is introduced in the first sentence with an indefinite subject NP and is referred to with a subject pronoun in the next two sentences. The second character is introduced in the third sentence by an indefinite (prepositional) object NP. This character is the actor in the fourth and fifth sentence and is referred to with a definite subject NP, while the first character is referred to here with a definite non-subject NP. The final sentence of the story contains a potentially ambiguous subject pronoun, which matches the two characters in gender.

The comprehension task involves answering a question about the final sentence. To answer correctly, the listener must resolve the pronoun in that sentence. If participants assume pronouns to refer to the discourse topic, their answer will clarify who they think the topic is at the end of the story. If participants think the second character became the new topic halfway through the story, they will resolve the pronoun as this second character.

Four stories without topic shift are also included to compare them with the topic shift stories. An example of a comprehension story without topic shift is given in Figure 3.

1 *Een clown heeft net zijn eigen gezicht geschminkt.*
'a clown has just painted his own face'
2 *Hij wil wel eens iemand anders schminken.*
'he wants to paint someone else'
3 *Hij komt in de keuken een kok tegen.*
'he comes across a cook(MASC) in the kitchen'
4 *De clown besluit de kok te schminken.*
'the clown decides to paint the cook(MASC)'
5 *En dan schminkt de clown een heel stoer gezicht bij de kok.*
'and then the clown paints a real tough face on the cook(MASC)'
6 *Hij vindt dat het prachtig is geworden.*
'he thinks it turned out great'

Comprehension question:
*Wie vindt het prachtig geworden?*
'who thinks it turned out great?'

*Figure 3.* An example of a recorded comprehension story without topic shift.

In these stories without topic shift, the first character remains the actor throughout the whole story. This character is introduced with an indefinite subject NP in the first sentence and is referred to by a pronoun in the second and third sentence. In order to keep these stories similar to the stories with topic shift, the first character is mentioned by a definite NP in the

fourth and fifth sentence, although as a subject rather than an object. The second character is introduced in the third sentence by an indefinite (prepositional) object NP and is referred to in the fourth and fifth sentence with a definite NP. There is no topic shift: The second character is never referred to with a pronoun and also never occurs in subject position. The final sentence of the story again contains a potentially ambiguous pronoun.

To answer the comprehension question correctly, the participant should again use the structure of the preceding discourse to determine the topic referred to by the pronoun in the last sentence. If the participant considers the first character to be the topic throughout the whole story, the participant will resolve the pronoun as this first character.

### 4.4. Materials memory task

The Auditory Memory Test, a subpart of the Schlichting Test for Language Production (Schlichting et al. 1995a), was also included in the test battery. This task includes word lists of increasing length that are read aloud by the tester and repeated by the participant. The words are mainly one-syllable in length, with a CVC-structure. All the words originate from the Lexilijst, a vocabulary test for 1;9 – 2;3 year old children (Schlichting et al. 1995b).

### 4.5. Procedure

Each child was tested individually in a quiet room at school. The session took roughly 20 minutes, with the production task preceding the comprehension task. Between these two language tasks, the memory task was administered. Two testers were present for testing the child. One tester sat across from the child and turned the storybook pages (production) or listened to the story recordings with the child (comprehension). The second tester noted responses to the final pictures during the production task and, during the comprehension task, played the pre-recorded stories and noted the answers to the final questions. This person sat further away from the child, behind a computer screen. During the production task, it was made clear to the child that the second tester wanted to understand the story but could not see the storybook. The child's job was to make sure that this second person understood what was happening in the stories.

The first tester began the production task by showing the child an introductory page including all the storybook characters and asking the child to name them. This introductory page was used to check the child's know-

ledge of the names of the story characters. Then the tester told a story based on a two-page storybook (one sentence per page) and asked the child to also tell a story, based on another two-page storybook. After the practice session, the child was once again reminded that the second tester could not see the storybook and wanted to know what was happening in the story. The child was then asked to describe four storybooks of six pictures each.

The memory task was presented as a word repetition game. First, for practice, the tester read one word aloud and the child repeated it. Then the tester read aloud and the child repeated increasingly longer sets of word lists. The test ended when the child failed two lists in a row or refused to continue.

The comprehension task included eight pre-recorded stories, four with topic shift and four without topic shift. These stories were presented via a computer, in two different orders. At the end of each story, the recording was stopped and the tester asked the child the question about the ambiguous pronoun in the last sentence. If necessary, the child was allowed to hear the story another time.

At several moments during the testing session and upon conclusion of the whole session, the children were rewarded with colorful stickers. In general, the children gave the impression that they had no trouble understanding what was expected of them and that they enjoyed doing the tasks.

Test sessions for adults were equivalent to the children's sessions, with a few exceptions. Adults were also tested individually, but with only one tester present. They were warned that, later on, someone else would have to listen to their recordings and understand their stories. Their session lasted roughly 10-15 minutes. During the introduction to the production task, the adults did not have to name the figures on an introductory page. The adults were explicitly requested to produce only one or two sentences per storybook page, since pilot testing showed that some adults tell long and detailed stories. The adults received no rewards.

The total test sessions were recorded and later transcribed. All transcriptions were controlled for accuracy. As an additional check during the actual test sessions, a tester noted responses to target pictures, answers to comprehension questions and successfully repeated memory lists.

### 4.6. Scoring and coding

In the production task, the focus was on how speakers refer back to a previously introduced topic after they have established a new topic. Derived from theoretical accounts of reference in discourse (Beaver 2004; Grosz,

Joshi and Weinstein 1995), the following three rules were adopted for coding discourse topics: 1) the topic of an utterance is a referring expression that has been mentioned in the preceding utterance, 2) if the utterance contains one pronoun, this pronoun is the topic, and 3) if the utterance contains no pronouns or more than one pronoun, the topic is the subject of the preceding utterance.

A topic shift was coded if the topic at a certain point in the discourse was different from the previous topic. Two coders (not the authors) independently scored each transcript for topic shifts and referential forms used to re-introduce the first character. Their judgments with respect to topic shift agreed for 95.5%. Over all the topic shift productions, the judges agreed on the expression used to re-introduce the first character for 98.6%. In the cases of judgment differences or obvious errors, the authors made the final decisions as to how to code the data.

The goal of the comprehension task was to investigate whether participants were sensitive to the difference between stories with and without topic shift. To investigate this sensitivity, answers to the question about the ambiguous pronoun were scored as to whether they referred to the first-mentioned character or the second-mentioned character in the story. A third category 'other' included any answers that did not fit into the first two response categories. Participants gave a variety of 'other' responses. For example, they sometimes mentioned a referent from a previous story or answered "both people" or "I don't know".

## 5. Results[2]

### 5.1. Results production task

Elicitation of a first topic shift halfway through the storybooks was quite successful. Children realized a topic shift (+TS) from the first character to the second character when talking about the third (or sometimes fourth) picture of the storybook 84% of the time (104 out of 124 stories) and adult speakers 98% of the time (90 out of 92 stories). Failure to realize a topic shift was caused by a participant either focusing too strongly on only one character, resulting in no shift, or alternating the two characters without establishing a topic. Ten children failed to realize a first topic shift one time, three children two times and one child three times. In the adult group, the two unrealized shift productions came from two different participants.

Only in the productions that realize a first topic shift is it necessary for a speaker to re-introduce the first character with an NP and, therefore, further analyses include only these productions. The score of main concern was whether participants used a full NP or a pronoun when referring back to the first character in describing the final picture. For example, in the pirate story (see Figure 1), one child's description of the last three pictures concluded with a re-introduction of the pirate using a full NP in the final utterance: *de ridder heeft een visnet / hij pakt de bal d'ruit / en dan is de Piet Piraat d'r blij mee* ('the knight has a fish net / he gets the ball out / and then the Pete-Pirate is happy with it'). In contrast, another child concludes with a pronoun: *dan gaat de ridder 'm vangen / en hij heeft de bal in een net gevangen / nu heeft ie z'n bal weer terug* ('then the knight is going to get it / and he has got the ball in a net / now he has his ball back again').

To see whether children and adults differed in number of full NPs used to refer back to the previous topic, two t-tests were conducted: An independent-samples t-test on the basis of mean percentages full NPs per participant (t1-analysis), and a paired-samples t-test on the basis of mean percentages full NPs per item (t2-analysis). All percentages were normalized using an arcsine transformation. The difference between the two age groups was significant on both analyses ($t1(52)=-8.5$; $p<.001$, $t2(3)=-4.7$; $p<.05$).

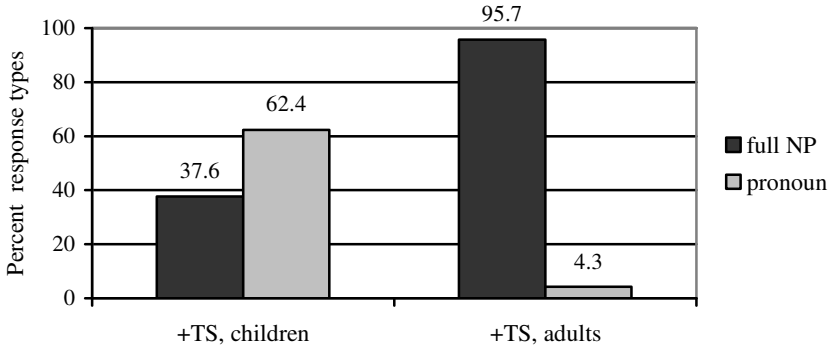


*Figure 4*. Referring expression used in production task to refer to previous topic after topic shift (+TS).

Figure 4 gives the percentages, aggregated over participants, of full NPs and pronouns produced by children and adults when re-introducing the first character. Children used 37.6% full NPs (SE=9%) and 62.4% pronouns (SE=9%). Individually, eight children used pronouns exclusively, three

used NPs exclusively and the other 20 children showed a mixed pattern. In contrast with the children, the adults used 95.7% full NPs (SE=4%) and only 4.3% pronouns (SE=4%). The adult pronoun responses were produced by different participants.

5.2. Results comprehension task

Children and adults were also compared as to how they answered the question at the end of the two types of comprehension stories. The children's and adults' response percentages (aggregated over participants) for these two story types are reported separately in Figure 5 below. For the topic shift stories, children answered 52.4%, 26.6% and 21%, respectively, with the first character, second character and 'other' responses. For the stories without topic, they answered 54%, 21% and 25%, respectively, with the first character, second character and 'other' responses. Individually, the children again showed a mixed pattern, with some children giving many correct answers, some giving mixed responses and others giving many incorrect responses. In contrast, the adult answers for the topic shift stories were 34.8%, 65.2% and 0% respectively, for the first character, second character and 'other' responses. For the stories without topic shift, the adults answers were 91.3%, 6.5% and 2.2% respectively, for the first character, second character and 'other' responses. Their patterns of responses were also mixed.

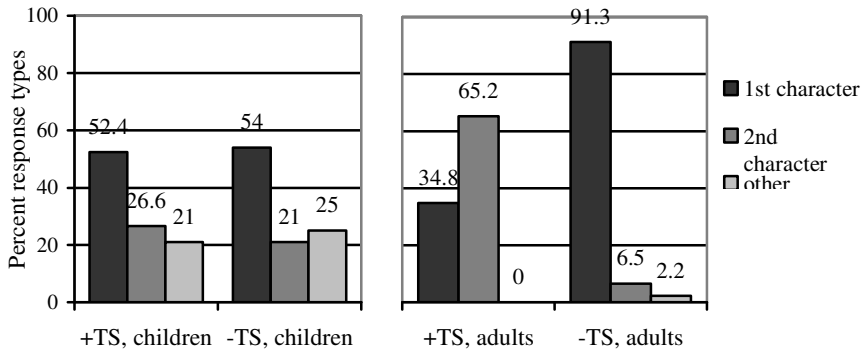


*Figure 5.* Reference assigned to pronoun in comprehension task, in stories with (+TS) and without (-TS) topic shift.

Repeated measures ANOVAs were applied to the response percentages per participant (F1, averaged over items) and to percentages per item (F2, aver-

aged over participants). These analyses included three factors: Response Type (1st character, 2nd character, and 'other' response), Story Type (with topic shift (+TS) and without topic shift (-TS)) and Age Group (children and adults). Response Type was considered a within-participants and within-items factor, Story Type was treated as within-participants and between-items, and Age Group as between-participants but within-items. As in the previous analyses, an arcsine transformation was used for all percentages. To guard against possible violations of the statistical assumption of sphericity, the Huyn-Feldt correction was used whenever the factor Type of Response was involved (Stevens 1992). We report the actual degrees of freedom that were used in the statistical test.

The main effect of Response Type was significant ($F1(1.8,104)=73.1$; $p<.001$; $F2(2,12)=45.5$; $p<.001$), as were the two-way interactions of Response Type and Age Group ($F1(1.8,104)=12.8$; $p<.001$; $F2(1.3,12)=6.1$; $p<.05$) and of Story Type and Response Type ($F1(2,104)=38.4$; $p<.001$; $F2(2,12)=17.7$; $p<.001$). These effects, however, were qualified by a significant three-way interaction of Response Type, Story Type and Age Group ($F1(2,104)=29.8$; $p<.001$; $F2(1.3,12)=7.1$; $p<.05$).

Follow-up analyses investigating the nature of this three-way interaction showed that for children, there was no interaction between Story Type and Response Type (both F-values<1). There was only a main effect of Response Type ($F1(1.5,60)=13.6$; $p<.001$; $F2(1.9,12)=8.6$; $p<.01$). Whether the children heard a story with or without topic shift, in either case they most often answered the comprehension question by naming the first character as the answer to the question (53%; SE=9%), as opposed to choosing the other options (all p-values equal .06); the remaining responses were equally distributed over the second character (24%; SE=8%) and the 'other' responses (23%; SE=8%) (both p-values equal 1.00).

In contrast, the adult data showed a highly significant interaction effect on Response Type and Story Type ($F1(1.3,44)=86.3$; $p<.001$: $F2(1.4,12)=15.6$; $p<.01$). Follow-up analyses showed that adults had a qualitatively different response pattern for stories with a topic shift compared to stories without a topic shift. Adults favored answering the question in topic shift stories with the second character (65.2%; SE=9%) over the first character (34.8%; SE=9%) ($p1< .01$; $p2=1.00$) or the 'other' response (0%; SE=0%) ($p1<.001$; $p2<.05$), and with the first character over the 'other' responses ($p1<.001$; $p2=.10$), although these differences were not always significant in the analysis by items. For stories without topic shift, the answer pattern was completely reversed: Here, adults preferred answering the

question by mentioning the first character (91.3%; SE=5%) over either the second character (6.5%; SE=4%) (p1<.001; p2<.05) or the 'other' response (2.2%; SE=3%) (p1<.001; p2<.005). The difference between using the second character versus the 'other' response was not significant (p-values>.30). As in the previous analyses, Bonferroni correction was applied to each set of these pairwise comparisons.

### 5.3. Results memory task

Participants' score on the memory task was the total number of word lists that they correctly repeated. For children, the mean score was 7.7 lists, with a range of 4-10 lists. For adults, the mean was 14 lists, with a range of 10-18 lists. Pearson correlation coefficients were calculated between the variables Memory Score and available measures of language comprehension and production. Age was also included in the correlations to track its effects in relation to memory and language development (all correlations were subjected to a two-sided test of significance). For adults, there were no significant correlations between Age and Memory Score (r= -.17; p=.44), between Age and language scores (absolute r-values<.19; p-values>.39), or between Memory Score and language scores (absolute r-values<.25; p-values>.24).

Significant correlations were found in the children's group: The correlation between Age and Memory Score was marginally significant (r= +.33; p=.07), indicating that there was a trend for memory scores to increase with increasing age. Age as well as Memory Score was significantly correlated with a number of language measures.

In production, children with higher memory scores were more prone to use a full NP instead of a pronoun when re-introducing the first character at the end of their storytelling (r= +.47; p<.01). Age predicted production performance somewhat less well (r= +.30; p=.10).

In comprehension, correlations were found between children's adult-like answers to the final question and their Memory Score. In recordings with a topic shift, children with higher memory scores were more likely to give the correct, second referent answer (r= +.37; p<.05). Memory score was not significantly correlated with the tendency to give the incorrect first referent answer (r= +.29; p=.12). However, the relation between Memory Score and the likelihood to give 'other' responses was negative, and highly significant (r= -.56; p<.001). This pattern of correlations was approximately the same for the relation between Age and measures of pronoun comprehension: Age

and correct second referent answer were strongly correlated (r= +.53; p<.01), as were Age and 'other' response, which had an inverse relation (r= -.57; p<.001). Age and incorrect first referent answers were not correlated (r= +.14; p=.46).

No other correlations were found between Memory Score, Age, and comprehension responses There were also no significant correlations between production and comprehension measures in either of the two age groups (absolute r-values<.27; p-values>.14).

## 6. Discussion

In this study, children and adults were tested on both production and comprehension of stories in which they demonstrated their use and interpretation of anaphoric subjects in ongoing discourse. In accordance with the Asymmetric Grammar Hypothesis, three predictions were formulated about production, comprehension and their relation to working memory capacity.

The *prediction in relation to production* was that children would prefer to produce subject pronouns over full NPs, even when referring to non-topics. The results of the production task support this prediction. In the present study, when adult speakers re-introduce the previous topic at the end of the story, they use a full NP. This choice of form is determined by bidirectional optimization on the basis of the constraints of the grammar. Children generally do not use full NPs. Instead, they tend to use referentially more economical subject pronoun forms for previously mentioned referents. Their use of pronouns does not discriminate between referents which are the current discourse topic and referents which are not. So, as speakers, they do not seem to optimize bidirectionally: They do not take their listener's perspective into account. The result is that a listener may be unable to recover the intended meaning from the used form .

The *prediction about comprehension* was that children would not interpret NPs as a topic shift signal and might therefore interpret a subsequent subject pronoun incorrectly. Because only the topic shift stories signaled a topic shift, the questions in the two story types should be answered differently. For an adult listener, the correct interpretation of a full NP requires bidirectional optimization. After listening to stories without a topic shift, the adults correctly answered the question by naming the first character. Also, they generally correctly answered the question in the topic shift stories by naming the second character. For these topic shift stories, however,

the adults also gave an unexpected number of answers referring to the first character. In comparison, the children's pronoun resolution did not seem to be affected at all by the presence or absence of topic shift marking. They answered the questions equivocally for both story types, with a general preference for the first character. Their remaining answers were divided between the second character and 'other' responses. As in production, we conclude that children do not seem to optimize bidirectionally: They fail to notice when the speaker signals a topic shift by using a full NP rather than a pronoun.

Why do the adults, who do differentiate between story types, seem to be more successful answering questions about stories without topic shift? For experimental purposes, the structure of the two comprehension stories was made as parallel as possible to each other and to the production storybooks. This structure, however, might have made the topic shift stories more opaque for listeners. One aspect of discourse structure which may be playing a role here is referent prominence, involving factors such as frequency of mention, first mention, subject status and pronominalization (Song and Fischer 2005). In the present comprehension stories, the first character might be considered more prominent, even in the topic shift stories, which could make it more difficult for listeners to note the topic shift in this story type. This could explain why the adults, who did differentiate between the two stories, were nevertheless less successful with the topic shift stories.

The children's response pattern was different from the adults' pattern in that the children reacted almost identically to both story types. Differentiation between stories with and without topic shift depends upon two things: 1) the assumption that pronouns refer to discourse topic, and 2) the ability to interpret the full NP form and non-subject status of the first character halfway through the story as an indication of topic shift. The conclusion of the present study is that the children's difficulties with answering the question about the subject pronoun are rooted in their failure to note the topic shift markings, and not in their lack of knowledge that pronouns refer to prominent antecedents. This standpoint is supported by Song and Fischer (2005), who demonstrated how discourse prominence factors of possible antecedents already affected three year old children's comprehension of subject pronouns. For the comprehension of object pronouns in a short discourse, Spenader, Smits and Hendriks (2009) presented evidence that when there is a clearly established discourse topic, children do know that object pronouns refer to the discourse topic. They are able to use this dis-

course structural information to interpret a pronoun that is otherwise ambiguous for them.

A complete account of children's development of grammar must also explain why the ability of taking into consideration the other person's perspective in communication is not fully functional in children as old as 6. In the present study, the *prediction pertaining to memory* stated that adult-like subject pronoun use and interpretation in discourse is the result of the completion of bidirectional optimization, which may require a larger working memory capacity. This study did not include a Theory of Mind test since children of this age are expected to already possess first order Theory of Mind abilities (for an overview, see De Villiers 2007). Focusing on working memory capacity, the first step in investigating the prediction about memory is to determine the children's working memory capacity. As expected, results showed that the children were not able to correctly repeat as many word lists as the adults. Within the group of children, there was variation in memory scores with a tendency for memory scores to increase with age.

The second step is to investigate the prediction that a larger memory capacity is related to successful bidirectional optimization in production and comprehension. A significant correlation was found between children's higher memory scores and adult-like production of NPs to re-introduce a previous topic in discourse. Recall the production task: At the beginning of their discourse, speakers introduce a first character. Later, they shift the topic to introduce a second character. At introduction, these two characters are new to the speaker and the listener. Towards the end, however, the children must shift from the second character back to the previously introduced first character. To successfully apply bidirectional optimization, a speaker will have to keep in mind what the listener already knows and what he needs to know in order to follow this second topic shift. The correlations showed that children with higher working memory scores are thus able to consider the perspective of their listeners. Like adult speakers, when re-introducing the first character, these children more often opt for a full NP and not a referentially more economical pronoun.

Two significant correlations between memory and comprehension were also found, both in the stories with topic shifts. Children with higher memory scores were able to attend better to the speaker's story structure and recognize the topic shift, thereby giving the correct, second character as the answer to the pronoun question. Children with lower memory scores seemed to lose track of the story more often and come up with 'other' an-

swers that pertained to neither story character. For the stories with no topic shift, there were no significant correlations.

In the present study, working memory scores are positively correlated with children's more adult-like performance in both production and comprehension. The presumption is that as children have more memory capacity they will be able to successfully apply bidirectional optimization. On the basis of data presented here, however, it cannot be conclusively determined whether limited working memory capacity hinders children in moving from unidirectional to bidirectional optimization or in maintaining the necessary representation of the discourse.

Summarizing, the evidence presented in this study suggests that the ability to take into account the listener as a speaker and the speaker as a listener is a complex skill that children still have to develop for different linguistic phenomena during the course of language acquisition. Such bidirectional optimization is, however, crucial for obtaining a symmetric system of language in which a speaker's preference for economy is optimally balanced with a listener's need for clarity. In production, because children often do not bidirectionally optimize, they are overly economical and often produce unrecoverable subject pronouns. Even though they have a specific referent in mind for the pronoun, they do not consider whether the listener will be able to identify that referent. In comprehension, when children are not capable of bidirectional optimization, they do not recognize the speaker's use of a full NP as signaling a topic shift and, therefore, do not correctly interpret a subsequent pronoun. The problem is not a local misinterpretation of the pronoun, but rather an inability to understand the topic shift marking in the prior discourse.

In conclusion, the results of this study confirm the predictions made by the Asymmetric Grammar Hypothesis. In this account, asymmetries between comprehension and production arise as a result of the particular constraints of the grammar. Based on the constraints involved, children may show selective delays in production and comprehension, whenever they are not yet able to take into account the opposite perspective in communication. These asymmetries disappear when children, as speakers and listeners, optimize bidirectionally.

## 7. Notes

1. This investigation was supported by a grant from the Netherlands Organization for Scientific Research, NWO, awarded to Petra Hendriks (grant no. 277-

70-005) for the VICI project "Asymmetries in Grammar". The authors thank Ellis Wubs for her assistance in carrying out the experiment and an anonymous reviewer for useful comments.

2. The results of the present analyses differ slightly from those reported in Wubs et al. 2009, due to the use of an improved scoring method.

## References

Baauw, S. and Cuetos, F.
2003 The interpretation of pronouns in Spanish language acquisition and breakdown: Evidence for the "Principle B Delay" as a non-unitary phenomenon. *Language Acquisition* 11, 219-275.

Beaver, D. I.
2004 The optimization of discourse anaphora. *Linguistics and Philosophy* 27, 3-56.

Blutner, R.
2000 Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17, 189-216.

Burzio, L.
1998 Anaphora and soft constraints. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (eds.), *Is the Best Good Enough? Optimality and Competition in Syntax.* 93-113. Cambridge, MA: MIT Press.

Cannizzaro, G.
2010 Comprehension delay of word order in preschoolers. *31 Deutschen Gesellschaft für Sprachwissenschaft*, March 4-6, 2009, Osnabrück. Article to appear in this volume.

Conroy, S., Takahashi, E., Lidz, J. and Phillips, C.
2009 Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry* 40, 446-486.

De Villiers, J.
2007 The interface of language and Theory of Mind. *Lingua* 117, 1858-1878.

Fikkert, P. and de Hoop, H.
2009 Language acquisition in Optimality Theory. *Linguistics* 47-2, 311-357.

Grosz, B. J., Joshi, A. K. and Weinstein, S.
1995 Centering: A framework for modeling the logical coherence of discourse. *Computational Linguistics* 21, 203-225.
Gundel, J., Hedberg, N. and Zacharski, R.
1993 Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307.
Hendriks, P., Banga, A., van Rij, J., Cannizzaro, G., and Hoeks, J.
2010 Adults' on-line comprehension of object pronouns in discourse. *31 Deutschen Gesellschaft für Sprachwissenschaft*, March 4-6, 2009, Osnabrück. Article to appear in this volume.
Hendriks, P., Englert, C., Wubs, E. and Hoeks, J.
2008 Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information* 17, 443-466.
Hendriks, P. and Koster, C.
2010 Production/comprehension asymmetries in language acquisition. *Lingua* 120, 1887-2094.
Hendriks, P. and Spenader, J.
2005/6 When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* 13, 319-348.
Hendriks, P., Van Rijn, H., & Valkenier, B.
2007 Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua 117* (11), 1879-1896.
Hickmann, M. and Hendriks, H.
1999 Cohesion and anaphora in children's narratives: A comparison of English, French, German and Mandarin Chinese. *Journal of Child Language 26, 291-452.*
Karmiloff-Smith, A.
1981 The grammatical marking of thematic structure in the development of language production. In W. Deutsch (ed.), *The Child's Construction of Language.* 121-147. London: Academic Press.
McKee, C.
1992 A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition* 2, 21-54.
Prince, A. and Smolensky, P.
2004 *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell.
Reinhart, T.
2006 *Interface Strategies*. Cambridge, MA: MIT Press.
Schlichting, J.E.P.T., van Eldijk, M.C.M., Lutje Spelberg, H.C., van der Meulen, S. and van der Meulen, B.F.
1995a *Schlichting Test voor Taalproductie*. Nijmegen: Berkhout.
1995b *Lexilijst*. Nijmegen: Berkhout.

Smolensky, P.
1996 On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27, 720-731.
Song, H, and Fischer, C.
2005 Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language* 52, 29-57.
Spenader, J., Smits, E. J. and Hendriks, P.
2009 Coherent discourse solves the Pronoun Interpretation Problem. *Journal of Child Language* 36, 23-52.
Stevens, J.
1992 *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
Thornton, R. and Wexler, K.
1999 *Principle B, VP Ellipsis, and Interpretation in Child Grammar*. Cambridge, MA: MIT Press.
Tyler, L.K.
1983 The development of discourse mapping processes: The on-line interpretation of anaphoric expressions. *Cognition* 13, 309-341.
Wigglesworth, G.
1997 Children's individual approaches to the organization of narrative. *Journal of Child Language* 24, 279-309.
Wubs, E., Hendriks, P., Hoeks, J. and Koster, C.
2009 Tell me a story!: Children's capacity for topic shift. *Proceedings Generative Approaches to Language Acquisition North America 3 (GALANA 3)*, Cascadilla Press.