Coreference in Child Russian:
Distinguishing Syntactic and Discourse Constraints

Nina Kazanina and Colin Phillips
University of Maryland

1. Introduction

In English, as in many other languages, sentences like (1), (2) and (4) allow coreference between the subject NPs *Pooh* and *he*, whereas (3) does not. Coreference in (3) is blocked by Condition C of the Binding Theory, which prevents a pronoun from c-commanding its antecedent (Chomsky, 1981):

(1) While Pooh$_i$ was reading a book, he$_i$ ate the apple.
(2) Pooh$_i$ ate the apple while he$_i$ was reading a book.
(3) * He$_i$ ate the apple while Pooh$_i$ was reading a book.
(4) While he$_i$ was reading a book, Pooh$_i$ ate the apple.

In the Russian counterparts of (1-4), coreference possibilities are the same as English, except that coreference is *disallowed* in (4R): the *while*-sentence with backwards anaphora is possible in English, but impossible in Russian.

(1R) Poka Pooh$_i$ chital   knigu,   on$_i$ s'el   yabloko.
      while Pooh   was reading.imp the book   he   ate.perf the apple
(2R) Pooh$_i$ s'el   yabloko, poka on$_i$ chital   knigu.
      Pooh ate.perf the apple while he   was reading.imp the book
(3R) * On$_i$ s'el   yabloko, poka Pooh$_i$ chital   knigu.
      he ate.perf the apple   while Pooh was reading.imp the book
(4R) * Poka on$_i$ chital   knigu,   Pooh$_i$ s'el yabloko.
      while he was reading.imp the book   Pooh ate.perf the apple

Here we examine the nature of this contrast between Russian and English in coreference possibilities, with a focus on what Russian-speaking children know about coreference possibilities, and how they come to know it.

The unavailability of coreference in (4R) cannot be due to linear order alone, since Russian does allow a pronoun to corefer with a following NP in examples like (5), in which the pronoun is inside a relative clause.

(5) Rasskaz, kotoryy ona$_i$ prochitala, rasstroil devochku$_i$.
      story   which she$_i$ read.perf upset.perf   girl$_i$.
     The story which she$_i$ read upset the girl$_i$.

Furthermore, the ban on coreference in (4R) between the embedded subject pronoun and the matrix subject NP cannot be explained in terms of c-command relations as an effect of Condition C, because in (4R) the subject of the matrix

clause does not c-command the embedded subject. If it did, and if coreference in (4R) was blocked by Condition C, then we would incorrectly predict that neither the matrix subject nor the matrix object could be the antecedent for the pronoun in (6).

(6)   Poka on$_{*i/j}$ razgovarival   po telefonu, Ia$_i$ gotovil Winnie-Poohu$_j$ uzhin.
        while he was talking.imp on phone, Eeyore cooked   Pooh   dinner
        While he$_{*i/j}$ was talking on the phone, Eeyore$_i$ was cooking dinner for Pooh$_j$.

The unavailability of coreference in (4R) does not involve c-command, and cannot be attributed to Principle C or to a simple linear order constraint. Also, since Russian does not allow null subjects in sentence-initial *while*-clauses, we cannot attribute the impossibility of coreference in (4R) to the availability of null-subjects Therefore, some independent constraint must be at work in Russian. As we will describe in more detail below, the constraint in question is a non-configurational constraint which specifically targets agent NPs.[1]

In what follows, we first confirm the independence of the *poka*-constraint and Condition C by showing a developmental dissociation between the two constraints. We then discuss in more detail the nature of the constraint and how it might be learned.

2.   Experiment

We used a Truth Value Judgment Task (TVJT) (Crain & McKee, 1985; Crain & Thornton 1998) to test Russian children's knowledge of coreference possibilities in sentences like (1R-4R). English sentences very similar to (1-4) were tested by Crain & McKee (1985), also using the TVJT. They showed that 3-year old English-speaking children allow the same coreference possibilities as adults in all 4 sentences. Importantly, English children correctly reject backwards anaphora in (3), but correctly allow backwards anaphora in (4), thereby indicating that they know Condition C.

Methods and design. We tested coreference judgments in a 2x2 between subjects design, in which we independently varied the name-pronoun order (forwards vs. backwards anaphora) and the main-subordinate clause order, thus testing the full paradigm in (1R-4R). The use of a between-subjects design allowed us to test more than one token of each condition per child. We tested 44 children aged 2;8 to 4;11 in day-care centers in Moscow. We chose this age range because previous literature, primarily on English, has shown that children in this age range respect Condition C (Crain & McKee, 1985, Crain & Thornton, 1998) but still allow certain Condition B violations (Chien & Wexler, 1990; Avrutin & Wexler, 1992). An additional 12 children were tested, but were excluded from the analyses, because they gave incorrect answers on one or more filler trials. A number of additional children, predominantly 3-year olds, were

excluded from the experiment because they were unable to perform the task. A further 6 older children (age 5-6) were tested in a within-subjects design. Each session lasted for approximately 25 minutes. The parents of all children gave informed consent for their participation.

Participants were assigned randomly to each of the 4 conditions. Each child saw the same condition across four stories, interspersed with an equal number of simple filler stories. Fillers were assigned after each test item in such a way that the child should give the opposite response to the preceding test item. The filler stories were used to prevent the children from giving identical responses across all trials, and as an independent measure of whether each child understood the task.

The stories were acted out by one experimenter using toy figures, while a second experimenter manipulated a puppet. Following each story, the puppet described one thing that he thought happened in the story. The child's task was to reward the puppet with a candy if he was correct, or to feed him a candy wrapper if he was incorrect. When the child said that the puppet was wrong, we asked the child to explain why, in order to assess the child's understanding of the story. For the same reasons, we asked the child to say and explain again what the puppet said, even if the child agreed with the puppet's statement.

Crucially, each of the four stories was used in exactly the same form in all four experimental conditions, with the only variation across conditions being in the word order of the puppet's final statement. The following is one of the two stories in which the test sentence contained the conjunction *poka* 'while'; the two remaining stories used the conjunction *kogda* 'when'.[2]

EXAMPLE STORY (translated from Russian)

SETTING: *A room. There is a table in the corner of the room, and there is an apple and two bananas on the table. Eeyore is in the other corner of the room reading a book.*
POOH: "Hi, Eeyore. You are reading a book, I see. I wonder what I should do? – *Pooh walks around the room, notices the apple. -* Oh, what a nice apple! I shall eat it right now."
EEYORE: "No, Pooh, you can't eat it: that's my apple." *Eeyore continues to read the book.*
POOH: "OK, I can't eat the apple, because it's Eeyore's apple. Then I shall have to eat a banana instead." *Pooh eats a banana.*
EEYORE: "OK, Pooh, I've finished reading the book, so you can read the book now." *Pooh starts reading the book.*
EEYORE: *Eeyore walks across the room to the table with the apple.* "Here is my apple. I think I shall eat it right now." *Eeyore takes the apple to his mouth to eat it, but just before biting into it he stops and says*: "I shouldn't be such a greedy donkey! Pooh wants the apple and so I think I should give it to him. As for me, I can have a banana instead." *Eeyore drags the table*

with the apple to Pooh who is reading the book. "Pooh, here is the apple and you can have it."

POOH: "Oh, I'm such a happy bear! I have a book to read and an apple to eat! I am going to read the book and eat the apple!"

FINAL SETTING: *Pooh continues reading the book. At the end of the story, there is an apple leaf next to Pooh and the book, to remind the child that it was Pooh who ate the apple, while reading the book.*

PUPPET: "That was a story about Eeyore and Pooh. First Eeyore was reading the book and then Pooh was reading the book. I know one thing that happened …

	While Pooh was reading the book, he ate the apple."	Forw. anaphora emb. clause first
or	Pooh ate the apple, while he was reading the book."	Forw. anaphora main clause first
or	He ate the apple, while Pooh was reading the book."	Backw. anaphora Condition C
or	While he was reading the book, Pooh ate the apple."	Backw. anaphora *poka*-constraint

In order to use exactly the same story across all four experimental conditions, while satisfying the requirements of the TVJT (Crain & Thornton, 1998), it was necessary to satisfy the Condition of Plausible Denial *twice*. Pooh nearly ate the apple while Eeyore was reading the book, and Eeyore nearly ate the apple while Pooh was reading the book, but in both instances the apple-eater chose a banana instead at the last minute. These circumstances made it felicitous for children to reject any of the test sentences under a disjoint reference reading. In all other respects, the stories were designed to bias children towards a coreferential interpretation of the pronoun. In this way, children's 'no' responses could be interpreted as clear evidence for a constraint blocking coreference.

Results: Figure 1 shows overall results for all conditions. As can be seen, children accepted coreference readings (by responding 'yes') on more than 80% of the control trials involving forwards anaphora. They rejected coreference readings on the Condition C trials at a rate of 83%, providing further cross-linguistic evidence for the claim that even very young children respect Condition C. Russian children reject coreference in the critical *poka*-conditions at a rate of 42% (recall that these in fact include examples with both *poka* 'while' and *kogda* 'when'). On the one hand, this is a significantly higher rejection score than in the two forwards anaphora conditions; on the other hand, this is a much higher rate of acceptance than is found in the Condition C trials. An ANOVA and pairwise comparisons of condition-scores show all main effects, interactions and pairwise contrasts between conditions to be reliable

(p<0.01), with the exception of the two forward anaphora control conditions, which were not reliably different, as expected.
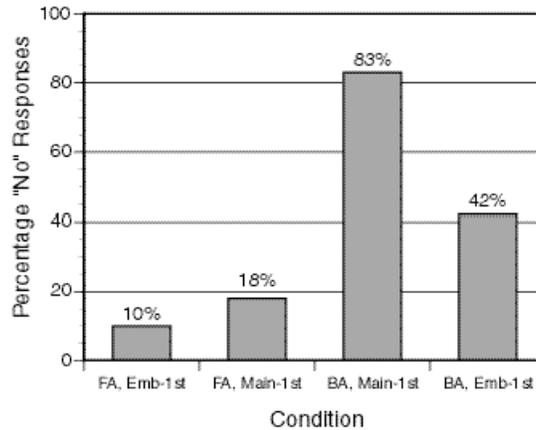

Figure 1:
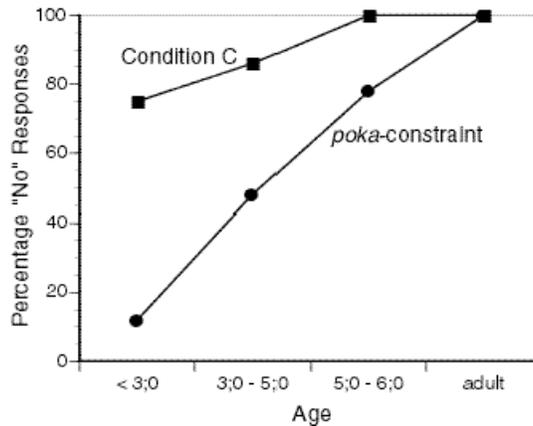Rejection of Coreference in 44 Russian children, aged 2;8-4;11 (mean age = 4;0)

Figure 2:
Rejection of Coreference in Backwards Anaphora Items, ages <3;0 to adult

Figure 2 shows a breakdown by age-group of rejection of Principle C and *poka*-constraint conditions in Russian children and adults. The figure shows that whereas rejection of Principle C violations is consistent at all ages tested, rejection of *poka*-constraint violations shows a clear change over time. Below age 3, children show a high level of acceptance of coreference in *poka*-constraint contexts; 3-4 year olds show 48% rejection of coreference. It is not

until after age 5 that children consistently disallow coreference in *poka*-constraint contexts. Our group of 5-year olds rejected coreference in both Condition C (100%) and *poka*-constraint contexts (78%).

The judgments of the youngest group of Russian children look strikingly like English: both groups block coreference only in the Condition C trials. The parallel with English is further supported by comparing our results with the results of Crain & McKee's study (Crain & McKee, 1985). The youngest group of children, aged below 3 years, are closely matched in Russian and English, as Figure 3a shows. They reject coreference in Condition C sentences, but allow it in *poka/while*-sentences. Although the size of the youngest group of Russian children is small, the difference between conditions is significant (Mann-Whitney U-test, $p<0.01$). The older groups of Russian and English-speaking children, aged 3-5 years, continue to give very similar judgments to Condition C sentences, but begin to reject coreference in *poka*-sentences more than in *while*-sentences. Note that within this older group of Russian children it is not the case that children either consistently accept or consistently reject coreference in *poka*-sentences: most children give a mix of 'yes' and 'no' responses.
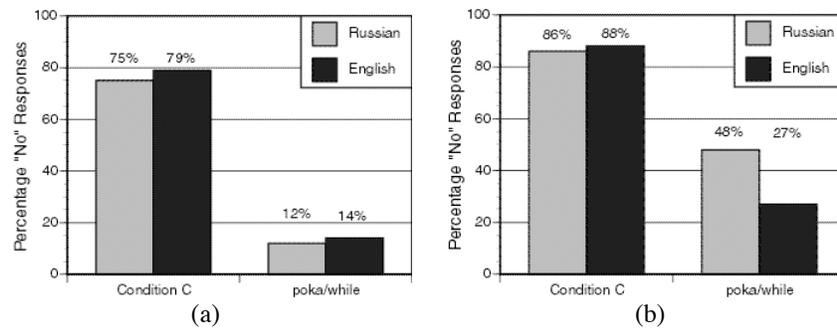

Figure 3: Backwards anaphora judgments of 3-year olds
(a) Russian (n=5, mean 2;9), English (n=7, mean 3;1, Crain & McKee, 1985)
(b) Russian (n=39, mean 4;2), English (n=62, mean 4;2, Crain & McKee, 1985)

Discussion: Our results show a clear developmental dissociation between Condition C and the *poka*-constraint. The delay lends further support to the notion that the two constraints are independent, but leaves open the question of what the *poka*-constraint actually is, and how Russian children succeed in learning it by around age 5. The remainder of the paper focuses on these questions.

The dissociation between the two constraints may be viewed as consistent with previous findings on English, where children are known to respect Condition C as young as age 3 (Crain & McKee, 1985; Crain & Thornton, 1998), but have been argued to show a delay in pragmatic or discourse-related constraints on coreference, e.g. in the literature on Principle B (Chien & Wexler, 1990; Grodzinsky & Reinhart, 1993; Thornton & Wexler, 1999). However, we

should also consider the possibility that the developmental dissociation results from the fact that Condition C is a universal constraint – which may not need to be learned at all – whereas the *poka*-constraint is a Russian-specific constraint, which must be learned.

First, however, we should consider whether the 'delay of *poka*-constraint' effect that we have observed could have the same cause as the well-studied 'delay of Principle B' effect.

A number of studies of children's coreference judgments have shown that children incorrectly allow Condition B violations involving referential antecedents (7a), but correctly reject Condition B violations involving quantificational antecedents (7b) (e.g., Chien & Wexler, 1990; Philip & Coopmans, 1996; Crain & Thornton, 1998; Thornton & Wexler, 1999; Avrutin & Wexler, 1999). A family of accounts of this effect run as follows. Adults reject both (7a) and (7b) because they correctly analyze both as cases of variable binding, which is regulated by Condition B. Young children incorrectly analyze (7a) as involving coreference, which is not regulated by Condition B, and hence escape Condition B in (7a). Since the pronoun in (7b) can only be linked to the quantificational subject by means of variable binding, Condition B cannot be escaped. In other words, children can use coreference (as opposed to variable binding) in configurations where adults cannot (either because they do not know the constraint, or because they are unable to compute its consequences).

(7) a. Mama Bear$_i$ scratched her$_i$.
    b. Every bear$_i$ scratched her$_i$.

We can ask whether Russian children's over-acceptance of coreference in *poka*-sentences also results from the use of coreference relations where variable binding is required. We consider this possibility unlikely. First, if this strategy were available to Russian children in *poka*-sentences, it should also be available to them in Condition C contexts (in which the pronoun-name order is identical to *poka*-sentences). But this clearly is not the case, since even the youngest children respect Condition C. Second, the issue of whether a bound variable representation is available is irrelevant in the critical *poka*-sentences, since the main clause subject neither c-commands nor is c-commanded by the subject of the *poka*-clause. Therefore, we consider it unlikely that the 'delay of the *poka*-constraint' effect and the 'delay of Condition B' effect have the same source.

Also, it is interesting to consider the Russian children's responses in light of the suggestion that young children have a constraint which disallows all backwards anaphora (Tavakolian, 1977; Lust et al., 1980; Solan, 1981; Lust, 1986). Strikingly, in Russian we find that younger children allow backwards anaphora more liberally than older children or adults. Thus we find no support for the notion that the youngest children have a linear constraint against backwards anaphora.

3. Nature of the Constraint

Next, we turn to a more detailed characterization of the nature of the constraint against coreference in *poka*-sentences, focusing in particular on the scope of the constraint, in terms of case, thematic relations, and verbal aspect. We also consider whether the constraint applies across a broader range of temporal conjunctions.

Position of the Pronoun: We first examine whether coreference is prohibited in all backwards anaphora sentences introduced by the conjunction *poka* 'while' in Russian. In the examples of *poka*-sentences discussed above, the embedded clause pronoun is a nominative agent. Changes to the case or thematic properties of the pronoun in the embedded clause lead to marked improvement in coreference possibilities.

(8) shows that a matrix clause subject can be the antecedent of an embedded clause object pronoun. (9) shows that when the pronoun is the possessor of the subject NP, coreference is also possible.

(8)   Poka Anna   smotrela   na nego$_i$, Misha$_i$ ulybalsya.
        while Ann was looking.imp at him   Mike was smiling.imp
        While Ann was looking at him$_i$, Mike$_i$ kept smiling.

(9)   Poka ego$_i$ mama   chitala   knigu,   Pooh$_i$ s'el   yabloko
        while his mother was reading.imp book   Pooh   ate.perf apple
        While his$_i$ mother was reading the book, Pooh$_i$ ate an apple.

These examples show that the constraint does not apply to non-subject NPs.

Some subject NPs are exempt from the constraint. (10) shows that when the embedded subject pronoun is a dative subject, coreference is also possible. (11) shows that even certain nominative subject pronouns allow backwards anaphora in *poka*-sentences. The embedded subject in (11a) is a nominative experiencer, and the embedded subject in (11b) is the subject of the raising predicate *kazhetsya* 'seem'. In both instances, the nominative subject is a non-agent, and coreference is possible.

(10)   Poka emu$_I$ bylo zhal' devochku, Pooh$_i$ stoyal   ryadom s ney.
        while he-Dat was.imp sorry   girl   Pooh was standing.imp next to her
        While he$_i$ felt sorry for the girl, Pooh$_i$ was standing next to her.

(11) a. Poka ona$_i$ byl   direktorom , Anna$_i$ mogla   delat' vse, chto hotela.
        while she was.imp director   Anna could.imp do   all   that wanted
        While she$_i$ was a director, Ann$_i$ could do whatever she wanted.
     b. Poka on$_i$ kazalas'   rebenkom, Ivan$_i$ ezdil   v avtobuse zaycem.
        while he seemed.imp child   Ivan rode.imp   bus   without ticket
        Since he$_i$ seemed to be a child Ivan$_i$ rode the bus for free.

Therefore, the constraint specifically regulates backwards anaphora among *agent* subjects in *poka*-sentences.

Other temporal conjunctions: Examination of a wider range of temporal subordinating conjunctions reveals a good deal of variation. For example, coreference is possible between nominative agents in sentences with the conjunctions *do togo kak* 'before' (12a) and *posle togo kak* 'after' (12b).

(12) a.   Do togo kak ona pereehala   v Rossiyu, Masha zhila   vo Francii.
        before   she moved.perf to Russia Masha was living.imp in France
        Before she moved to Russia, Masha lived in France.

    b.   Posle togo kak ona vypila   tabletku, Masha usnula   na divane.
        after   she took.perf   pill   Masha fell asleep.perf on sofa.
        After she took a pill, Mashe fell asleep on the sofa.

Sentences with the subordinating conjunction *kogda* 'when' provide minimal pairs with both possible and impossible coreference. In (13a) both clauses contain an imperfective verb, and coreference is possible. In (13b) the main clause contains an imperfective verb and the subordinate clause contains a perfective verb. Coreference is impossible in (13b).

(13) a.   Kogda ona$_i$ uezzhala iz goroda, Masha$_i$ plakala.
        when she left.imp   town Masha   cried.imp
        When she was leaving the town, Masha was crying.

    b. * Kogda ona$_i$ uehala iz goroda, Masha$_i$ plakala.
        when she left.perf town   Masha cried.imp
        When she left the town, Masha was crying.

In the examples in (13), then, coreference is only available when the main and subordinate clauses match in aspect. Furthermore, the sentences in our experiment which block backwards anaphora with *poka* 'while' and *kogda* 'when' had non-matching aspect in the two clauses. In example (12b) above, introduced by *posle togo kak* 'after', the two clauses match in aspect, and backwards anaphora is possible. This raises the possibility that some of the variation among temporal conjunctions in coreference possibilities reduces to an aspect-matching requirement.[3]

However, the aspect-matching requirement could not be a simple morphological requirement that the two verbs match in aspect. The requirement could only be stated at a semantic level. This is shown by examples like (12a), introduced by *do togo kak* 'before' and with mismatching aspect. In order to explain the availability of backwards anaphora, despite the formal aspectual mismatch, we might appeal to the fact that the imperfective VP 'living in

France' is understood as a bounded, completed event, due to the semantics of the sentence. We have encountered a number of other examples where the availability of a semantically matching aspectual interpretation seems to lead speakers to allow backwards anaphora, despite a formal aspectual mismatch.

Although aspect-matching can account for some of the variation in coreference possibilities, it remains to be seen whether this requirement can account for all of the variation. We find examples like (6), where there is an aspect match, but backwards anaphora between the matrix and embedded subjects is impossible. Such examples cannot, given our current state of understanding, be reduced to an aspect matching requirement. This suggests that the aspectual constraint on coreference may be a more complicated requirement on the combination of the events in the matrix and embedded clauses, than an aspect-match.

More work is needed in order to establish whether all of the variability across conjunctions in coreference possibilities can be explained in terms of aspect.

4. Learning

An important remaining question, which we can only begin to address here, is how and why Russian children's coreference judgments in *poka*-sentences change from English-like judgments at around age 3 to adult-like judgments by around age 5.

We can exclude the possibility that Russian children learn coreference possibilities conservatively, only assuming that backwards anaphora is possible once they observe positive evidence in their input. Quite independent of the difficulty of reliably tracking coreference in input, the experiment results show that Russian children do the opposite: they overgeneralize at a young age, and must retreat from this overgeneralization.

It is possible to appeal to a general delay in the acquisition of discourse or pragmatic constraints on coreference, similar to claims that have been made to explain the 'Delay of Principle B' effect. Although this is possible as a way of explaining why Russian children continue to violate the constraint for so long, it essentially restates the problem of how these children eventually come to obey the *poka*-constraint.

A possible role for aspect is promising. If young Russian children overgeneralize in *poka*-sentences because they bypass a certain aspectual requirement for some reason, perhaps because of incomplete knowledge of aspect, then they can retreat from their overgeneralization simply by learning about aspect. Nevertheless, such an approach leaves it a mystery why Russian is subject to this aspectual constraint, whereas English is not. An appeal to the intuition that aspect is more grammaticized in Russian than in English does not solve the problem completely, since we have already seen that the aspectual constraint applies to semantic aspect rather than to morphological aspect.

Clearly, then, the learning problem awaits more detailed investigation.

5. Conclusions

We have seen that adult speakers of Russian disallow backwards anaphora in sentences introduced by *poka* 'while' and certain other temporal conjunctions, but young Russian children overgeneralize, by allowing backwards anaphora in these contexts at a high rate. Russian children do not consistently reject backwards anaphora in *poka*-sentences until after age 5. This contrasts with Principle C, which is respected at all ages tested, even among children below age 3. The difference in the acquisition of the two constraints in Russian is particularly striking, because the sentences in which they come into play are minimally different ((4R) and (3R)).

Closer analysis of the contexts which disallow backwards anaphora in Russian shows that the constraint specifically applies to sequences of agent subjects, and may be restricted to contexts in which the main and subordinate events stand in a specific aspectual relationship to one another.

However, the question of why such a constraint should apply to a specific subset of backwards anaphora contexts in Russian, in particular why agenthood and aspect are singled out, remains unanswered, as does the question of how Russian children come to know this constraint.

Endnotes

1 Examples of constraints on coreference which cannot be reduced to c-command are familiar from many other languages, including English. For example, backwards anaphora between subjects of coordinated clauses is generally unacceptable, **He$_i$ opened the door and Pooh$_i$ went out into the street*.
2 Although Russian adults disallowed coreference 100% of the time in our test sentences, in general *kogda* seems to be less restrictive than *poka*.
3 We assume that the aspect-matching constraint is relevant only to sentences with agent pronouns in the embedded clause.

References

Avrutin, S. & K. Wexler. (1992). Development of Principle B in Russian: Coindexation at LF and coreference. *Language Acquisition, 4*, 259-306.

Avrutin, S. & K. Wexler. (1999). Children's knowledge of subjunctive clauses: obviation, binding and reference. *Language Acquisition, 8*, 69-102.

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

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

Crain, S. & C. McKee. (1985). *Acquisition of Structural Restrictions on Anaphora* Proceedings of NELS XV

Crain, S. & R. Thornton. (1998). *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, Mass.: MIT Press.

Grodzinsky, Y. & T. Reinhart. (1993). The innateness of binding and coreference. *Linguistic Inquiry, 24*, 69-103.

Lust, B. (1986). Introduction. In B. Lust (ed.), *Studies in the Acquisition of Anaphora, Vol. I: Defining the Constraints*. Dordrecht: Reidel, pp.3-103.

Lust, B., K. Loveland & R. Kornet. (1980). The development of anaphora in first language: Syntactic and pragmatic constraints. *Linguistic Analysis, 6*, 359-391.

Philip, W. & P. Coopmans. (1996). The double Dutch Delay of Principle B effect. In A. Stringfellow *et al.*, eds., *Proceedings of BUCLD 20*. Somerville, MA: Cascadilla Press.

Solan, L. (1981). The acquisition of structural restrictions on anaphora. In S. Tavakolian (ed.), *Language Acquisition and Linguistic Theory*. Cambridge, MA: MIT Press, pp.59-73.

Tavakolian, S. (1977). *Structural Principles in the Acquisition of Complex Sentences*. Ph.D. dissertation, University of Massachusetts, Amherst.

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