Anaphora resolved

Roelofsen, F.

Publication date
2008

Chapter 2

Constraints on Binding and Coreference: State of the Art

Binding and coreference are constrained in interesting ways. This chapter discusses some of the most prominent existing accounts of such constraints: those of Reinhart (1983), Heim (1998), Fox (1999a), Büring (2005b), and Reinhart (2006). Before turning to the theories proper, however, let me first review the basic data that is to be accounted for.

2.1 Basic Data

Section 2.1.1 discusses Condition B effects, section 2.1.2 discusses Crossover effects, and section 2.1.3 discusses Dahl's puzzle.

2.1.1 Condition B Effects

Binding is subject to so-called *Condition B* effects:

2.1. GENERALIZATION. [Condition B effects on Binding][1]
Pronouns cannot be bound by their coarguments.

2.2. DEFINITION. [Coarguments]
Coarguments are NPs whose θ-role and/or case is assigned by the same predicate.

Here are some sentences which exhibit Condition B effects:

(2.1) a. Every girl loved her.
b. Every man sent a letter to him.

[1] Several versions of this generalization have been proposed in the literature. The present formulation, adapted from Büring (2005a, pp.55–56), is relatively theory-neutral and covers the relevant data. The term *Condition B* originates from Chomsky's (1981) binding theory.

c. Every woman believes her to be a great dancer.
d. Mary asked every boy to wash him.

Coreference is very often also subject to Condition B effects. That is, it is usually impossible for pronouns to corefer with their coarguments. For example, coreference is impossible in the following examples.

(2.2) a. Susan loved her. $\Rightarrow$ her $\neq$ Susan
b. Tom sent a letter to him. $\Rightarrow$ him $\neq$ Tom
c. Norah believes her to be a great dancer. $\Rightarrow$ her $\neq$ Norah
d. Mary asked John to wash him. $\Rightarrow$ him $\neq$ John

Early theories of pronominal anaphora such as (Chomsky, 1981) assumed that coreference is *always* subject to Condition B effects. But Reinhart (1983, p.169) pointed out that there are at least two kinds of environments in which coreference is not subject to Condition B effects. The first kind of environment involves focus-sensitive operators like *only*:

(2.3) Only Max himself voted for him.

The second kind of environment is one in which previous discourse makes it particularly clear that a coreferential interpretation is intended:[2]

(2.4) I know what John and Mary have in common.
John hates Mary and Mary hates her too.

(2.5) If everyone voted for Oscar, then certainly Oscar voted for him.

I must note here that the judgments of my informants do not always confirm those of Reinhart. Many of my informants find that coreference is very marginal in (2.3), (2.4) and (2.5), and emphasize that there are certainly much more natural ways to convey the intended messages. In the recent literature, several authors have acknowledged the controversial status of these data (cf. Schlenker, 2005; Grodzinsky, 2007; Heim, 2007). The theories to be discussed below, however, do take these data very seriously. In fact, they are in large part especially designed to deal with them. Thus, for the sake of the discussion, I will pretend throughout the first part of this dissertation that these data are undisputed. Eventually, in the second part of the dissertation, I will try to account for their borderline-status.[3]

[2] Very similar examples were used by Evans (1980) to show that Condition C effects are suppressed in certain environments, see section 2.1.2 below. Therefore, examples like (2.4) and (2.5) are often attributed to Evans, rather than to Reinhart. I will likewise refer to (2.4) and (2.5) as Evans' examples from now on.

[3] The following example, adapted from Heim (1998), is sometimes taken to instantiate a third set of examples in which coreference is insensitive to Condition B effects.

(i) How can you doubt that the speaker is Zelda? She praises her to the sky.

2.1.2 Crossover Effects

Apart from Condition B effects, binding is also subject to so-called *crossover* effects. The rough idea is this. By definition, a noun phrase A can bind a pronoun P only if A c-commands P at LF. But this does not seem to be enough. For example, the sentences in (2.6) do not have the readings represented by the logical forms in (2.7).

(2.6) a. He likes every man.
b. His mother likes every man.
c. Who does he like?
d. Who does his mother like?

(2.7) a. [every man]^1 [he$_1$ likes t$_1$]
b. [every man]^1 [his$_1$ mother likes t$_1$]
c. [who]^1 [he$_1$ likes t$_1$]
d. [who]^1 [his$_1$ mother likes t$_1$]

These cases are sometimes taken to exemplify the following more general pattern: a noun phrase A can only bind a pronoun P if A already c-commands P in its base position (cf. Koopman and Sportiche, 1982; Reinhart, 1983; Büring, 2005a). In our present terminology, this generalization can be formulated as follows.

2.3. Generalization. [Crossover Effects]
A pronoun can only be cobound with traces that c-command it.

Note that all the logical forms in (2.7) involve a pronoun which is cobound with a trace that does not c-command it. So the generalization correctly predicts that these logical forms are illicit.

However, there are also many counterexamples to the generalization. Higginbotham (1980), for instance, observed that binding is possible in sentences like the following, even though this would involve a pronoun being cobound with a non-c-commanding trace.

(2.8) a. Whose mother loves him?
b. Every senator's portrait was on his desk.
c. Somebody from every city despises its architecture.

These examples seem to show that in order for A to bind P it is not necessary for A itself to c-command P in its base position, but merely for A to *be contained in* another NP which c-commands P in its base position.

However, as Heim notes, even though *she* and *her* may be intended to refer to the same person, they are intended to do so *through different guises*. Technically, they are assigned two distinct individual concepts, which may refer to the same individual in the real world but to distinct individuals in other worlds in the context set. Thus, technically speaking, coreference does not obtain here, and the question whether or not it is subject to Condition B effects does not arise.

But even to this weaker generalization there are several counterexamples. Lasnik and Stowell (1991, p.690) for example, observed that it does not apply to pronouns which occur in *adjuncts*.

(2.9) Who did Joan say she admired in order to please him?

(2.10) Which book did you tell Bill to file without reading it?

And Reinhart (1983, p.180) reported the following type of counterexample, which she attributed to Ross:

(2.11) That people hate him disturbs every president.

So it is quite clear that generalization 2.3 cannot be right. The theories to be discussed below focus on a weaker generalization that does seem to be empirically adequate.

2.4. Generalization. [Strong Crossover Effects]
A pronoun cannot be cobound with a trace that it c-commands.

This generalization avoids the counterexamples just discussed. The downside of it is that it predicts *some*, but not all alleged crossover effects. For instance, it predicts that the logical forms in (2.7a) and (2.7c) are illicit, but it has nothing to say about the ones in (2.7b) and (2.7d). In section 3.4 I will return to this issue, but in the rest of this chapter, I will assume that generalization 2.4 is the one to account for.

Strong Crossover effects are sometimes considered to be a special case of so-called *Condition C* effects (see especially Chomsky, 1981). The generalization is supposed to be that R-expressions (traces, names, and descriptions) cannot be covalued (cobound or coreferential) with any expression that c-commands them. For example, in (2.7a) and (2.7c) the traces cannot be cobound with the pronouns that c-command them, and in (2.12) below, the name *Max* cannot corefer with the pronouns that c-command it.

(2.12) a. He loves Max. $\Rightarrow$ he $\neq$ Max
b. He called Max's mother. $\Rightarrow$ he $\neq$ Max
c. He says that Mary called Max's mother. $\Rightarrow$ he $\neq$ Max

However, there are many counterexamples to this generalization. For example, coreference is possible in:

(2.13) Whom did the candidates themselves vote for?
Not surprisingly, John voted for John and Bill voted for Bill.

(2.14) I know what John and Mary have in common:
John hates Mary, and Mary hates Mary as well. (cf. Evans, 1980)

(2.15) I think that this is exactly what happened:
Peter forced Tom to call Peter's girlfriend. (cf. Schlenker, 2004)

(2.16) He didn't give her a diamond ring because,
although he's madly in love with her,
Walter's just not ready to tie the knot. (McCray, 1980)

Some of the theories to be discussed below do offer an account of Condition C effects, but none of them deals with all these counterexamples.[4] I think that Condition C effects are the result of various mechanisms working in tandem. This point will be discussed briefly in appendix A. In the present chapter Condition C effects will be left out of consideration.

2.1.3 Dahl's Puzzle

The following type of sentence, first discussed by Östen Dahl (1974), represents a notorious puzzle concerning the interpretation of VP ellipsis.

(2.17) Max said that he called his mother. Bob did too.
a. ...Bob too said that Bob called Bob's mother.
b. ...Bob too said that Max called Max's mother.
c. ...Bob too said that Bob called Max's mother.
d. #...Bob too said that Max called Bob's mother.

The challenge is to account for the fact that (2.17a), (2.17b), and (2.17c) are possible readings of the target clause, while (2.17d) is not. The possible logical forms of the source clause, *Max said that he called his mother*, are given in (2.18) (we are only interested here in those logical forms in which both *he* and *his* are anaphorically related to *Max*). I have also indicated which reading of the target clause is associated with each of these logical forms, assuming that VP ellipsis is governed by VP Identity.

(2.18) Max said that he called his mother.

a.	[Max]^1 [t$_1$ said [he$_1$]^2 [t$_2$ called his$_2$ mother]]		(2.17a)
b.	[Max]^1 [t$_1$ said [he$_1$]^2 [t$_2$ called his$_1$ mother]]		(2.17a)
c.	[Max]^1 [t$_1$ said [he]^2 [t$_2$ called his mother]]	he=his=Max	(2.17b)
d.	[Max]^1 [t$_1$ said [he]^2 [t$_2$ called his$_2$ mother]]	he=Max	(2.17b)
e.	[Max]^1 [t$_1$ said [he$_1$]^2 [t$_2$ called his mother]]	his=Max	(2.17c)
f.	[Max]^1 [t$_1$ said [he]^2 [t$_2$ called his$_1$ mother]]	he=Max	(2.17d)

[4]Notice that (2.16) also falsifies a weaker version of Condition C, which says that R-expressions cannot be covalued with c-commanding *pronouns*. It is widely known that the original formulation of Condition C is problematic, but this weaker version is usually assumed to be valid. Counterexamples such as McCray's are not often acknowledged. I thank Anna Szabolcsi for pointing me to a manuscript by Peter Sells (1987), which pays special attention to McCray's example.

To account for Dahl's puzzle, a theory of anaphora must do two things: first, it must rule out (2.18f) as a logical form of the source clause, and thus (2.17d) as a possible reading of the target clause. Second, it must allow enough logical forms of the source clause to derive each of the legitimate readings of the target clause. In particular, it should not rule out both (2.18a) and (2.18b) (because this would make (2.17a) unavailable) or both (2.18c) and (2.18d) (this would make (2.17b) unavailable) or (2.18e) (this would make (2.17c) unavailable). This pattern cannot be accounted for in terms of Condition B effects or Strong Crossover effects, so there must be certain additional restrictions on binding and/or coreference.

2.1.4 Summary

The data can be summarized as follows. Binding is subject to Condition B and Strong Crossover effects. Coreference is usually also subject to Condition B effects, but there are some specific environments in which it is not. Finally, there must be certain additional restrictions on binding and/or coreference to account for the pattern found in Dahl's puzzle.

In the remainder of this chapter I will discuss some of the most prominent existing accounts of Condition B effects, Strong Crossover effects, and Dahl's puzzle. The first two proposals, those of Reinhart (1983) and Heim (1998), are primarily concerned with Condition B effects. Fox (1999a), Büring (2005b), and Reinhart (2006) are also concerned with Strong Crossover effects and Dahl's puzzle.

2.2 Reinhart's Coreference Rule

Tanya Reinhart (1983) argues that binding and coreference are fundamentally different. Binding relations, she assumes, are encoded in the syntax (by means of indices) and are subject to *grammatical* constraints. Coreference is not encoded in the syntax, but rather established *contextually.* Therefore, coreference cannot be subject to grammatical constraints. Rather, restrictions on coreference are of a *pragmatic* nature.

The first ingredient of Reinhart's theory of Condition B effects, then, is a grammatical constraint on binding. I will simply call this constraint *Condition B* here, and formulate it in a relatively theory-neutral way (cf. Büring, 2005a, pp.55–56):[5]

2.5. Definition. [Condition B]
Pronouns cannot be bound by their coarguments.

[5] There is an ongoing debate in the literature about how this constraint should be formulated exactly (cf. Pollard and Sag, 1992; Reinhart and Reuland, 1993; Büring, 2005a), and about whether it can be derived from more general syntactic principles (cf. Reuland, 2001, 2008). This debate is interesting in its own right, but perpendicular to the discussion here.

Next, Reinhart points out an interesting consequence of the assumption that binding is encoded in the syntax while coreference is not: it is always more *risky* for a speaker to use syntactic structures which contain referential elements than to use syntactic structures in which all the anaphoric links are already encoded. This is because referential elements always have to be resolved by the hearer, and this can go wrong. If all anaphoric relations are syntactically encoded, resolution does not come into play. Reinhart assumes that speakers generally want to avoid any risk of being misinterpreted, and thus always prefer to use syntactic structures which contain *bound* anaphoric elements rather than syntactic structures which contain *referential* anaphoric elements. Only if speakers cannot express the intended meaning using bound anaphora will they use referential elements. This idea can be implemented as follows:

2.6. Definition. [Coreference Rule]
A speaker will never use a logical form LF in a context C if LF is semantically indistinguishable from one of its binding alternatives in C.

2.7. Definition. [Binding Alternatives]
Let C be a context, let LF be a logical form, and let A and B be two noun phrases in LF, such that A and B corefer in C and such that A c-commands B in LF. Then the structure obtained from LF by:

- Quantifier raising A in case it has not been raised yet, and
- Replacing B with a (possibly reflexive) pronoun bound by A

is called a binding alternative of LF in C.

The idea is this: if a logical form LF containing a referential expression B expresses a certain meaning M, and that same meaning could also be expressed by a logical form LF′ which only differs from LF in that B is replaced by an element that is not coreferential with its antecedent but rather *bound* by it, then a speaker will always use LF′ rather than LF to express M.

Let us see how the Coreference Rule deals with the data discussed in section 2.1. First consider a typical Condition B effect:

(2.19) Max washed him.
 a. [Max]^1 [t$_1$ washed him] him = Max
 b. [Max]^1 [t$_1$ washed himself$_1$]

The Coreference Rule predicts that a speaker will never use (2.19a), in which *Max* and *him* corefer, because (2.19a) is semantically indistinguishable from its binding alternative (2.19b): both express the proposition that Max washed Max. A hearer will conclude from this that coreference cannot be intended in (2.19).

Now, let's see whether the Coreference Rule can deal with Condition B environments in which coreference is exceptionally permitted. First consider a focus construction:[6]

(2.20) Only Max himself voted for him.

a. [only] [[Max himself]^1 [t_1 voted for him]] him = Max himself
b. [only] [[Max himself]^1 [t_1 voted for himself$_1$]]

This time, the Coreference Rule does *not* rule out coreference, because the interpretation of (2.20a) *differs* from the interpretation of its binding alternative (2.20b): (2.20a) says that the others did not vote for *Max*, while (2.20b) says that the others did not vote for *themselves*. Finally, consider one of Evans' examples:

(2.21) I know what John and Mary have in common.
Mary voted for John and John voted for him too.

a. [John]^1 [t_1 voted for him] him = John
b. [John]^1 [t_1 voted for himself$_1$]

The Coreference Rule does *not* rule out coreference in (2.21a), because the interpretation of (2.21a) differs from the interpretation of its binding alternative (2.21b): (2.21a) says that John has the property of having voted for John, and this is indeed the property that John and Mary are supposed to have in common. (2.21b) on the other hand, says that John has the property of having voted for *himself*, and this is certainly *not* the property that John and Mary are supposed to have in common.

So Reinhart's Coreference Rule accounts for standard Condition B effects on coreference, and also for the exceptional coreference patterns found in focus constructions and in Evans' examples.

2.3 Heim's Exceptional Codetermination Rule

Irene Heim (1998) observed that at least three aspects of Reinhart's account need some further consideration. First, the theory does not explicitly state what it means for one LF to be *semantically indistinguishable* from another. One possibility that comes to mind immediately is that two LFs should be regarded as semantically indistinguishable if and only if they express the same proposition. But this would not work: (2.21a) and (2.21b) express the same proposition, but intuitively, at least in the context of (2.21), there is a significant semantic difference between them. So the question of when two LFs should be regarded as semantically indistinguishable is not trivial and should be addressed with care.

[6]For reasons of readability, F-marking has been suppressed here and will often be suppressed below.

Second, the Coreference Rule is not compatible with the VP Identity condition on VP ellipsis. To see this, consider the basic case of VP ellipsis in (2.22):

(2.22) Max called his mother and Bob did too.

We saw before that the pronoun in the source clause can either be bound, as in (2.23a), or referential, as in (2.23b). (2.23a) gives rise to the sloppy reading of the target clause (*Bob called Bob's mother*) while (2.23b) gives rise to the strict reading of the target clause (*Bob called Max's mother*).

(2.23) a. $[\text{Max}]^1$ $[\text{t}_1$ called his_1 mother$]$
b. $[\text{Max}]^1$ $[\text{t}_1$ called his mother$]$ his = Max

But the Coreference Rule predicts that a speaker will never use (2.23b), because it is semantically indistinguishable from its binding alternative (2.23a). This means that, as long as the VP Identity condition is assumed, the Coreference Rule wrongly predicts that the target clause in (2.22) does not have a strict reading.

The third issue with the Coreference Rule is that it is only concerned with coreference. Other kinds of anaphora, such as cobinding, are just as unacceptable as coreference in typical Condition B environments, and we would like to have a rule that embodies these restrictions all in one go, rather than separate rules for coreference, cobinding, and possibly yet other kinds of anaphora. Examples (2.24) and (2.25) illustrate this point.

(2.24) [John voted for him] him = John

(2.25) $[\text{Every man}]^1$ $[\text{t}_1$ said that he_1 voted for $\text{him}_1]$

The Coreference Rule correctly rules out coreference between [John] and [him] in (2.24), but fails to rule out cobinding of $[\text{he}_1]$ and $[\text{him}_1]$ in (2.25). Intuitively, these cases are analogous and should be ruled out by one and the same mechanism.

Heim's contribution, then, is twofold: first, she refines the notion of semantic indistinguishability. Second, she proposes a new constraint which preserves all the empirical virtues of Reinhart's Coreference Rule, is compatible with the VP Identity condition, and applies not only to coreference but also to cobinding and other kinds of anaphora. Let me first discuss the new constraint.

2.3.1 The Exceptional Codetermination Rule

Heim's theory is stated in terms of *codetermination*, a notion which embraces that of binding, cobinding, and coreference (and yet other anaphoric relations as well).

2.8. Definition. [Codetermination]
Let C be a context, let LF be a logical form, and let A and B be two NPs in LF. We say that A and B are codetermined in LF/C iff:

- A binds B in LF, or
- A and B corefer in C, or
- There is a third NP which is codetermined with A and B in LF/C.

The first ingredient of Heim's theory is a revised version of Condition B, which prohibits codetermination, rather than binding.

2.9. Definition. [Heim's Condition B]
Pronouns cannot be codetermined with their coarguments.

The second ingredient of the theory is a rule which states that codetermination is sometimes exceptionally allowed.

2.10. Definition. [Exceptional Codetermination Rule]
Let LF be a logical form in which a pronoun is codetermined with, but not bound by one of its coarguments. Then, LF is (marginally) allowed, in violation of Condition B, if it is semantically distinguishable from its binding alternative in the given context.

The reader is invited to check that Heim's Condition B and her Exceptional Codetermination Rule account for the standard Condition B effects, not only involving coreference, but also involving cobinding and other kinds of codetermination. The proposal also accounts for the exceptional cases in which codetermination *is* allowed in Condition B environments. Finally, it is compatible with VP Identity: (2.23b) is no longer ruled out.

2.3.2 Semantic Indistinguishability

When should two logical forms be regarded as semantically indistinguishable? For one thing, they should express the same proposition. But Heim notes that Evans' examples, repeated in (2.26) and (2.27), show that there is more to it.

(2.26) I know what John and Mary have in common.
Mary voted for John and John voted for him too.

(2.27) If everyone voted for Oscar, then certainly Oscar voted for him.

Heim suggests that these are typical cases in which *structured meaning* matters. In (2.26), there is a certain property P that Mary and John are supposed to have in common, namely, the property $[\lambda x.\ x$ voted for John$]$ of having voted for John. If such a particular property is under discussion, then an LF which says that John has the property P is to be distinguished from an LF which says that John has the property $[\lambda x.\ x$ voted for $x]$ (even though these two LFs as a whole denote the same proposition). The same reasoning can be applied to the example in (2.27), where it is the property of having voted for Oscar that is under discussion. This intuition can be implemented by defining semantic indistinguishability not only in terms of propositional content, but also in terms of *focus* values:[7]

2.11. Definition. [Semantic Indistinguishability]
Two logical forms LF and LF′ are semantically indistinguishable iff:

1. LF and LF′ express the same proposition, and

2. LF and LF′ have the same focus value.

Let me illustrate how this works for one of Evans' examples. Consider the final clause in (2.27), given in (2.28):

(2.28) ...then certainly Oscar voted for him.

Let us assume that [Oscar] is F-marked (notice that it must be accented in the given context). Now consider the logical form in (2.29a), where [him] corefers with [Oscar], and its binding alternative in (2.29b):

(2.29) a. $[\text{Oscar}]^1_F$ [t_1 voted for him] him = Oscar
b. $[\text{Oscar}]^1_F$ [t_1 voted for himself$_1$]

(2.29a) and (2.29b) express exactly the same proposition. However, their focus values differ. For example, (2.30) is a focus alternative of (2.29a) but not of (2.29b). Thus, (2.29a) is semantically distinguishable from its binding alternative.

(2.30) $[\text{Fred}]^1$ [t_1 voted for him] him = Oscar

Interestingly, the notion of meaning explored by Groenendijk (2007) comprises both the information an expression provides (≈ its propositional content) and the alternatives it gives rise to or presupposes (≈ its focus value). So in Groenendijk's system, "being semantically indistinguishable" in the sense defined here really comes down to "having the same meaning".

The three issues that Heim raised concerning Reinhart's original proposal are now resolved. It must be noted, however, that the explanatory aspect of Reinhart's theory has been lost: Heim's theory cannot be derived from the assumption that speakers generally seek to avoid risks of being misinterpreted. Furthermore, the theory does not yet account for Strong Crossover effects and Dahl's puzzle.

[7]Heim herself does not provide a concrete implementation, but I am quite convinced that the implementation given here is in line with what she has in mind (see also Heim, 2007).

2.4 Fox's Locality Constraint

Consider Dahl's puzzle again:

(2.17) Max said that he called his mother. Bob did too.

a. . . . Bob too said that Bob called Bob's mother.
b. . . . Bob too said that Max called Max's mother.
c. . . . Bob too said that Bob called Max's mother.
d. #. . . Bob too said that Max called Bob's mother.

The possible logical forms of the source clause are repeated in (2.18):

(2.18) Max said that he called his mother.

a.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}_1]^2$ $[t_2$ called his_2 mother]]		(2.17a)
b.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}_1]^2$ $[t_2$ called his_1 mother]]		(2.17a)
c.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}]^2$ $[t_2$ called his mother]]	he=his=Max	(2.17b)
d.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}]^2$ $[t_2$ called his_2 mother]]	he=Max	(2.17b)
e.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}_1]^2$ $[t_2$ called his mother]]	his=Max	(2.17c)
f.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}]^2$ $[t_2$ called his_1 mother]]	he=Max	(2.17d)

Heim's account does not rule out any of the logical forms in (2.18). In particular, it wrongly permits (2.18f), and thus (2.17d) as a possible reading of the target clause in (2.17). Reinhart's account does rule out (2.18f), but it also rules out (2.18c), (2.18d), and (2.18e) (all the logical forms that involve coreference), and thus wrongly predicts that (2.17a) is the only reading for the target clause in (2.17). Thus, an additional constraint is required to account for Dahl's puzzle.

Danny Fox (1998) proposed that the puzzle may be explained in terms of so-called *economy conditions*. The idea is that such economy conditions prevent the grammar from generating logical forms whose interpretation is identical to that of simpler alternative logical forms. After all, why would the grammar generate complicated LFs if the meaning they represent can just as well be represented by simpler alternatives?

If this is indeed how the grammar works, then the next question to ask is what the criteria for simplicity are. Why should one LF be considered simpler than another? Fox suggests that one natural criterion involves the length of binding dependencies: one LF is simpler than another if its binding dependencies are shorter. This idea can be implemented as follows:

2.12. Definition. [Locality]
A logical form is ruled out if it is semantically indistinguishable from one of its locality alternatives.

2.13. Definition. [Locality Alternatives]
Let LF be a logical form, and let A, B and P be three nodes in LF such that A

c-commands B and B c-commands P, and such that A, but not B, binds P. Then the structure obtained from LF by:

- Quantifier raising B in case it has not been raised yet;
- Adjusting P's binding index so that it's bound by B instead of A.

is a locality alternative of LF.

Locality rules out (2.18b) and (2.18f), and thus accounts for Dahl's puzzle ((2.18b) is ruled out because it is semantically indistinguishable from its locality alternative in (2.18a) and (2.18f) is ruled out because it is semantically indistinguishable from its locality alternative in (2.18d)).

To further support his theory, Fox (1999a, p.131) discusses two striking cases in which Dahl's puzzle is obviated. First, if we slightly change the source clause, as in (2.31), then the fourth reading of the target clause suddenly *is* available.

(2.31) Max said that his mother called him. Bob did too.

a. …Bob too said that Bob's mother called Bob.
b. …Bob too said that Max's mother called Max.
c. …Bob too said that Bob's mother called Max.
d. …Bob too said that Max's mother called Bob.

This is accounted for by Locality. In (2.17), the LF in which [he] corefers with [Max] and [his] is bound by [Max] is illegitimate, because it is semantically indistinguishable from its locality alternative, in which [his] is bound by [he] instead of [Max]. However, this argument does *not* carry over to (2.31): the LF of (2.31) in which [his] corefers with [Max] and [him] is bound by [Max] does not have a locality alternative, because [his] does not c-command [him] in (2.31).

A second case in which Dahl's puzzle is obviated, is obtained by modifying the embedded clause in the source clause with *only*:

(2.32) Max said that only he had called his mother. Bob did too.

a. …Bob too said that only Bob had called Bob's mother.
b. …Bob too said that only Max had called Max's mother.
c. …Bob too said that only Bob had called Max's mother.
d. …Bob too said that only Max had called Bob's mother.

According to Fox, (2.32d) is an available reading of (2.32).[8] This is accounted for by Locality: the crucial LF, in which [he] corefers with [Max] and [his] is bound by [Max], conveys that Max said that he, Max, was the only one with

[8] I must remark here that the judgments of my informants do not really confirm those of Fox. There was quite some variability in the judgments, but many informants reacted to (2.32) just as they reacted to Dahl's original example. I will leave this data-issue aside here. Further empirical research is needed to straighten out the facts.

the property [$\lambda x.\ x$ called Max's mother] (the others didn't call Max's mother). The locality alternative of this LF, in which [his] is bound by [he] instead of [Max], conveys that Max said that he, Max, was the only one with the property [$\lambda x.\ x$ called x's mother] (the others didn't call their own respective mothers). These two LFs express different propositions and are therefore *not* semantically indistinguishable. As a consequence, the reading in (2.32d) is not ruled out.

A third case in which Dahl's puzzle is obviated, which was not discussed by Fox, is when Dahl's sentence is uttered in the context of a particular question, such as:

(2.33)
a. Did Max call everyone's mother?
b. Well, I don't know...
c. MAX said he called his mother, and BOB did too.
d. But I haven't heard from SUE and MARY yet.

Notice that (2.33c) is identical to Dahl's original sentence. But in the given context, the target clause can easily be taken to mean that Bob said that Max called Bob's mother. Again, Locality accounts for this fact: in the context of (2.33), the crucial LF differs in focus value from its locality alternative.

Thus, Locality accounts not only for Dahl's original puzzle, but also for some cases in which the puzzle is obviated. The third case, in particular, has two important implications. First, to account for the fact that Dahl's puzzle is obviated in (2.33) it is crucial that semantic indistinguishability is not formalized in terms of propositional content only. Rather, as was already argued on independent grounds in section 2.3.2, the notion should also make reference to *focus values*. This was not made quite explicit in Fox's original proposal.

Second, cases like (2.33) are problematic for alternative accounts of Dahl's puzzle such as Fiengo and May (1994), Kehler (1993), and Schlenker (2005).

Fox (1999a, p.132) notes that Locality does not only account for Dahl's puzzle, but also for Strong Crossover effects. To see this, consider the following example:

(2.34) Who did he say we should invite?
a. $[\text{who}]^1\ [[\text{he}_1]^2\ [\text{t}_2 \text{ said we should invite } \text{t}_1]]$
b. $[\text{who}]^1\ [[\text{he}_1]^2\ [\text{t}_2 \text{ said we should invite } \text{t}_2]]$

The logical form in (2.34a) should be ruled out, because (2.34) cannot be used to ask who have the property [$\lambda x.x$ said we should invite x]. This is a Strong Crossover effect: the pronoun in (2.34a) is cobound with a trace that it c-commands. (2.34a) is successfully ruled out by Locality, because it is semantically indistinguishable from its locality alternative in (2.34b).

Thus, Locality accounts for Dahl's puzzle, for some striking cases in which Dahl's puzzle is obviated, and also for Strong Crossover effects.

The question that arises next is whether Locality can be combined, or even unified either with Reinhart's Coreference Rule or with Heim's Exceptional Codetermination Rule in order to obtain a theory that accounts for all the data we have seen so far.[9] This question has been addressed by Daniel Büring.

2.5 Büring: Have Local Binding!

We have seen that Heim's account resolved several problems with Reinhart's original account. So, at first sight, the most logical step to take would be to combine Fox's Locality constraint with Heim's Exceptional Codetermination Rule. But Büring (2005b) is more ambitious. He observes that Heim's reinterpretation of Reinhart's approach brought about certain complications. In particular, Condition B had to be reformulated so as to apply to *codetermination* rather than binding.[10] One of the issues that motivated this complication was that Reinhart's Coreference Rule was only concerned with coreference, and not with cobinding and other kinds of anaphora. For example, it failed to rule out cobinding in:

(2.35) [Every man]^1 [t$_1$ said that he$_1$ voted for him$_1$]

Büring observes that, once Fox's Locality constraint is adopted, cobinding in (2.35) *is* ruled out. After all, (2.35) is semantically indistinguishable from its locality alternative (2.36).

(2.36) [Every man]^1 [t$_1$ said that [he$_1$]^2 [t$_2$ voted for him$_2$]]

Thus, Büring argues, the complications proposed by Heim become superfluous. He therefore proposes to return to the simple formulation of Condition B (as in definition 2.5) and adopt Reinhart's original Coreference Rule alongside Fox's Locality constraint. Finally, he observes that the latter two constraints can actually be collapsed into one:

2.14. Definition. [Have Local Binding!]
A logical form is ruled out if it is semantically indistinguishable from one of its HLB alternatives.

2.15. Definition. [HLB Alternative]
Let LF be a logical form, let A and B be two noun phrases in LF such that A c-commands B and neither A nor any node c-commanded by A binds B. Then the logical form obtained from LF by:

- Quantifier raising A in case it has not been raised yet;

[9] That this is indeed a natural question to ask has been obscured a bit by the fact that Fox considered his Locality constraint to be a notational variant of Heim's Exceptional Codetermination Rule. But, as also noticed by Büring (2005b), the two constraints are not equivalent.

[10] I personally think that the main drawback of Heim's proposal was that it did not preserve the explanatory aspect of Reinhart's account. But this does not seem to be what motivated Büring; his own proposal does not preserve the explanatory aspect of Reinhart's account either.

- Replacing B with a pronoun bound by A.

is an HLB alternative of LF.

The reader is invited to check that this constraint (HLB) does indeed account for Condition B and Strong Crossover effects, and also for the exceptional cases in which coreference is not subject to Condition B effects.

But HLB is *not* compatible with VP Identity. That is, if VP Identity is assumed, HLB yields wrong predictions both for the simple cases of VP ellipsis that were already problematic for Reinhart's account from the start, and for the more intricate pattern found in Dahl's puzzle. To see this, first consider the simple case of VP ellipsis in (2.22), repeated here as (2.37):

(2.37) Max called his mother and Bob did too.

HLB predicts that coreference is impossible in the source clause. Thus, as long as VP Identity is assumed, HLB wrongly predicts that the target clause does not have a strict reading (Bob called Max's mother too).

Next, recall Dahl's puzzle:

(2.17) Max said that he called his mother. Bob did too.

a. . . . Bob too said that Bob called Bob's mother.
b. . . . Bob too said that Max called Max's mother.
c. . . . Bob too said that Bob called Max's mother.
d. #. . . Bob too said that Max called Bob's mother.

(2.18) Max said that he called his mother.

a.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}_1]^2$ $[t_2$ called his_2 mother]]		(2.17a)
b.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}_1]^2$ $[t_2$ called his_1 mother]]		(2.17a)
c.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}]^2$ $[t_2$ called his mother]]	he=his=Max	(2.17b)
d.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}]^2$ $[t_2$ called his_2 mother]]	he=Max	(2.17b)
e.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}_1]^2$ $[t_2$ called his mother]]	his=Max	(2.17c)
f.	$[\text{Max}]^1$ $[t_1$ said $[\text{he}]^2$ $[t_2$ called his_1 mother]]	he=Max	(2.17d)

HLB allows (2.18a) but rules out all the other LFs in (2.18). This means that, as long as VP Identity is assumed, HLB wrongly predicts that (2.17a) is the only possible reading of the target clause in (2.17).

So Büring is forced to depart from VP Identity. Instead, he adopts the following constraint on VP ellipsis, which is originally due to Fox (1999a):[11]

2.16. DEFINITION. [NP Parallelism]
Corresponding noun phrases in the antecedent and elided VPs must either:

[11] This formulation of NP Parallelism is taken from (Büring, 2005a, p.132). A slightly different formulation is given in (Büring, 2005b, p.267). I assume that these two formulations are intended to be equivalent, and use the above because it is slightly more explicit.

- have the same referential value, or
- be bound in parallel in their respective conjuncts.

2.17. Definition. [Referential Value]
The referential value of a noun phrase A is:

- the individual to which A refers, or
- the referential value of the NP that binds A.

Büring does not say explicitly what it means for two noun phrases to be *bound in parallel*. If the constraint were to be evaluated seriously, this would first have to be made more precise of course. But even so, we could ask whether the generalization that seems to be embodied by NP Parallelism is empirically correct. I think it is both too weak and too strong. Let me first consider a case of *cascaded ellipsis* which shows that (what seems to be) the intended generalization is too weak to rule out certain illicit readings of elided VPs.

(2.38) Bob called his mother, and Max did too. But Tom didn't.

NP Parallelism wrongly predicts that (2.38) has a so-called *mixed reading* which can be paraphrased as follows:

(2.39) Bob called Bob's mother, and Max called Max's mother.
But Tom didn't call Max's mother.

Next, consider the following well-known example from Rooth (1992a), which shows that NP Parallelism is not only too weak, but also too strong.

(2.40) First John told Mary that I was bad-mouthing her.
Then Sue heard I was.

NP Parallelism erroneously rules out the sloppy reading of (2.40), which says that Sue heard that I was bad-mouthing her, Sue. This is because such a sloppy reading would involve "non-parallel" binding of the pronouns in the elided VP and its antecedent.

It *might* be possible, of course, to adjust the NP Parallelism constraint in such a way that it becomes strong enough to rule out the undesired readings in (2.38), while still being weak enough to allow the strict reading in (2.37) and the three possible readings in Dahl's puzzle, and even somewhat weaker than it presently is so as to allow for the sloppy reading in (2.40).

The point is that Büring's account of binding and coreference is at best compatible with an ad hoc, still to be worked out, non-standard theory of VP ellipsis.

It would be preferable to have a theory of binding and coreference that is compatible with mainstream theories of ellipsis such as those based on VP Identity.[12] This was, I think, one of the main reasons for Tanya Reinhart to eventually depart from her 1983 theory and develop a new account in the mid-1990's. This account will be discussed next.

2.6 Reinhart's Interface Rule

Reinhart's 1983 theory was based on two assumptions. First, that binding relations are encoded in syntactic structure, while coreference is not; and second, that speakers generally try to avoid risks of being misinterpreted. It follows from these two assumptions that speakers generally prefer to use bound pronouns, which explicitly encode the intended anaphoric relations, rather than referential pronouns, which may well not be resolved as intended.

In her later work, Reinhart concludes that these assumptions, as plausible as they may seem at first, eventually yield the wrong predictions and should be reconsidered. More specifically, she proposes to leave the first assumption intact (binding is encoded by syntactic structure, coreference is not) but replace the second assumption, which is about *speakers*, with an alternative assumption about *hearers*. The general assumption is that hearers *minimize interpretive options*. In the specific case of anaphora, this means that if a certain interpretation is ruled out by grammatical restrictions on binding, then a hearer will recognize that this interpretation was not intended, even if it could in principle be derived via other anaphoric mechanisms. In other words, interpretations which are ruled out by restrictions on binding cannot be *sneaked in* via other anaphoric mechanisms. Reinhart points out that the existence of such a mechanism would be extremely useful. For communication to proceed efficiently, it is crucial for a hearer to keep interpretive options to a minimum at all times.

Reinhart (2006) formalizes this idea in terms of a notion called *covaluation*.

---

[12] It should be remarked that Fox (1999a, p.117) presents a particular case of cascaded ellipsis in support of NP Parallelism:

(i) Smithers thinks that his job sucks. Homer does, too. However, Marge doesn't.

Notice that (i) is structurally analogous to (2.38). However, (i) *does* have a mixed reading, at least for people who recognize that it is about a popular American sitcom, *The Simpsons*, in which Marge is Homer's wife and does not have a job of her own. The availability of such a mixed reading would be in accordance with NP Parallelism, and not with VP Identity. However, I don't think that this really is an argument in favor of NP Parallelism. First, the availability of a mixed reading in (i) is *exceptional*. Typically, mixed readings are not available for cases of cascaded ellipsis, and this is left unexplained by NP Parallelism. Second, the fact that a mixed reading *is* exceptionally available in (i) can, I think, be explained without doing away with the essence of VP Identity. Such an explanation will be discussed in section 5.2.

2.18. Definition. [Covaluation] Let C be a context, let LF be a logical form, and let A and B be two NPs in LF. Then A and B are covalued in LF/C iff:

- A does not bind B and B does not bind A in LF, and
- A and B are cobound in LF or A and B corefer in C.

Notice that covaluation is essentially a generic term for coreference and cobinding. As such, it is more general than coreference alone, but less general than Heim's notion of codetermination, which covered other kinds of anaphora as well. Reinhart proposes the following constraint on covaluation:

2.19. Definition. [Interface Rule]
A logical form LF is ruled out if one of its binding alternatives LF′ is such that:

a. LF and LF′ are semantically indistinguishable, and

b. The transition from LF to LF′ is illicit, because:

  - LF′ is ruled out by restrictions on binding (Condition B), or
  - The existing binding relations in LF are not preserved in LF′, or
  - LF′ is ruled out by another application of the Interface Rule.

The notion of a *binding alternative* has to be revised slightly. The only difference between definition 2.7 and definition 2.20 is that in the former A and B are supposed to *corefer*, while in the latter A and B are supposed to be *covalued.*

2.20. Definition. [Binding Alternatives]
Let C be a context, let LF be a logical form, and let A and B be two noun phrases such that A c-commands B in LF and such that A and B are covalued in LF/C. Then the structure obtained from LF by:

- Quantifier raising A in case it has not been raised yet;
- Replacing B with a pronoun or trace bound by A.

is called a binding alternative of LF.

Let us see whether the Interface Rule accounts for the data accumulated so far. Doing so will sometimes require quite some effort, because, as we will see, the workings of the Interface Rule are sometimes rather complex. Always keep in mind though, that the intuition behind it is very simple: interpretations which are ruled out by restrictions on binding cannot be sneaked in via other anaphoric mechanisms.

Let us first consider (2.19), repeated here as (2.41), which exhibits a basic Condition B effect on coreference. The Interface Rule correctly rules out (2.41a) because it is semantically indistinguishable from its binding alternative in (2.41b), and (2.41b) violates Condition B.

(2.41) Max washed him.

a. $[\text{Max}]^1$ $[t_1$ washed him] him = Max

b. $[\text{Max}]^1$ $[t_1$ washed $\text{him}_1]$

Now, let us see whether the Interface Rule can deal with Condition B environments in which coreference is exceptionally permitted. Consider (2.20), repeated here as (2.42):

(2.42) Only Max himself voted for him.

a. [only] [[Max himself]^1 [t_1 voted for him]] him = Max himself

b. [only] [[Max himself]^1 [t_1 voted for him_1]]

The Interface Rule does *not* rule out coreference here, because (2.42a) is not semantically indistinguishable from its binding alternative in (2.42b). Intuitively speaking, coreference is not sneaking in an interpretation that is ruled out by restrictions on binding, but rather gives rise to an interpretation that is different from what would be obtained from binding. Evans' examples are dealt with in a similar way. So the Interface Rule accounts for standard Condition B effects on coreference, and also for the cases in which coreference is exceptionally allowed in Condition B environments.

The issues which the Coreference Rule was facing and which were addressed by Heim's Exceptional Codetermination Rule are also satisfactorily dealt with by the Interface Rule. In particular, the Interface Rule does not only account for cases of illicit coreference, but also for cases of illicit cobinding, and it allows strict identity readings in VP ellipsis. Let me illustrate this with some examples. First consider a case of illicit cobinding:

(2.43) Every man said that he washed him.

a. [Every man]^1 [t_1 said that [he_1]^2 [t_2 washed him_1]]

b. [Every man]^1 [t_1 said that [he_1]^2 [t_2 washed him_2]]

The logical form in (2.43a), in which [he_1] and [him_1] are cobound, should be ruled out. The Interface Rule accounts for this: (2.43a) is ruled out because it is semantically indistinguishable from its binding alternative (2.43b), and (2.43b) violates Condition B.

Next, consider the simple case of VP ellipsis in (2.22), repeated here as (2.44):

(2.44) Max called his mother and Bob did too.

a. ...Bob called his own mother too. [sloppy]

b. ...Bob called Max's mother too. [strict]

The Coreference Rule prohibited coreference in the source clause and thus ruled out the strict reading in (2.44b). The Interface Rule, on the other hand, does *not* rule out coreference in the source clause (coreference is only ruled out if binding is, too, and binding is certainly possible here). Thus, the Interface Rule correctly admits the strict reading in (2.44b).

But the Interface Rule is not just an alternative for Heim's Exceptional Codetermination Rule. It does *more*. In particular, it accounts, at once, for Strong Crossover effects. To see this, consider example (2.34), repeated here as (2.45).

(2.45) Who did he say we should invite?

a. [who]^1 [[he$_1$]^2 [t$_2$ said we should invite t$_1$]]
b. [who]^1 [[he$_1$]^2 [t$_2$ said we should invite t$_2$]]

The logical form in (2.45a) is ruled out by the Interface Rule, because it is semantically indistinguishable from its binding alternative (2.45b), and (2.45b) does not preserve the existing binding relations in (2.45a). In particular, the trace that was bound by the wh-element in (2.45a) is bound by the pronoun in (2.45b).

Finally, Reinhart claims that the Interface Rule also accounts for Dahl's puzzle. To see whether this is indeed the case, let me briefly resume the puzzle:

(2.17) Max said that he called his mother. Bob did too.

a. ...Bob too said that Bob called Bob's mother.
b. ...Bob too said that Max called Max's mother.
c. ...Bob too said that Bob called Max's mother.
d. #...Bob too said that Max called Bob's mother.

(2.18) Max said that he called his mother.

a.	[Max]^1 [t$_1$ said [he$_1$]^2 [t$_2$ called his$_2$ mother]]		(2.17a)
b.	[Max]^1 [t$_1$ said [he$_1$]^2 [t$_2$ called his$_1$ mother]]		(2.17a)
c.	[Max]^1 [t$_1$ said [he]^2 [t$_2$ called his mother]]	he=his=Max	(2.17b)
d.	[Max]^1 [t$_1$ said [he]^2 [t$_2$ called his$_2$ mother]]	he=Max	(2.17b)
e.	[Max]^1 [t$_1$ said [he$_1$]^2 [t$_2$ called his mother]]	his=Max	(2.17c)
f.	[Max]^1 [t$_1$ said [he]^2 [t$_2$ called his$_1$ mother]]	he=Max	(2.17d)

The Interface Rule is supposed to do two things: first, it is supposed to rule out (2.18f) as a logical form of the source clause, and thus (2.17d) as a possible reading of the target clause. Second, it is supposed to allow enough logical forms of the source clause to derive each of the legitimate readings of the target clause. In particular, it should not rule out both (2.18a) and (2.18b), or both (2.18c) and (2.18d), or (2.18e).

Let us see whether this is indeed established. First consider (2.18f). This LF is indeed ruled out. To see this, we have to consider the binding alternative of (2.18f), which is (2.18b). First observe that (2.18f) and (2.18b) are semantically indistinguishable. Next, observe that (2.18b) is ruled out *by another application of the Interface Rule*: (2.18b) is semantically indistinguishable from its binding alternative, (2.18a), and (2.18a) does not leave the existing binding relations in (2.18b) intact: [his] is no longer bound by [Max] in (2.18a). So (2.18b) is ruled out by the Interface Rule, and this means that (2.18f) itself is prohibited as well.

So far so good. But the problem is that exactly the same line of reasoning yields that the Interface Rule rules out (2.18e) as well: (2.18e)'s binding alternative is (2.18b), just like that of (2.18f). (2.18e) and (2.18b) are semantically indistinguishable, and we have already seen that (2.18b) is ruled out by the Interface Rule. Thus, (2.18e) must be ruled out by the Interface Rule as well. As a consequence, (2.17c) is wrongly excluded as a possible reading of the target clause.

To sum up, the Interface Rule improves on the Coreference Rule in two ways: it accounts for cases of illicit cobinding, and it avoids the strict identity problem in simple cases of ellipsis. It also improves on Heim's Exceptional Codetermination Rule: it accounts, at once, for Strong Crossover effects. However, it still does not account for the more complex case of VP ellipsis in Dahl's puzzle.

Apart from this remaining empirical problem, there are two other aspects of Reinhart's theory that call for further attention. First, there is a striking discrepancy between (the simplicity of) the intuition behind the Interface Rule and (the complexity of) its actual formulation. Recall the basic intuition: interpretations which are ruled out by restrictions on binding cannot be sneaked in via other anaphoric mechanisms. We should expect, then, that the formal statement of the rule should say something like: a logical form LF is ruled out if it is semantically indistinguishable from one of its binding alternatives LF′, and LF′ is ruled out by constraints on binding. The actual formulation of the Interface Rule is much more complicated. In particular, it additionally requires that the existing binding relations in LF are preserved in LF′ and that LF′ is not ruled out by recursive applications of the Interface Rule.

The second issue that needs further attention is that the Interface Rule is not only undesirably complex in its formulation, but also in its workings. The analysis of sentence (2.46) illustrates this:

(2.46) Max said that he washed him.

This sentence cannot be taken to mean that Max said that he washed *himself*. That is, the co-arguments of *washed* cannot be anaphorically related. A similar intuition applies to simpler examples such as:

(2.47) Max washed him.

In these simpler examples, the intuition is straightforwardly accounted for: Condition B prohibits binding and (as a consequence) the Interface Rule prohibits covaluation. We would like the Interface Rule to deal with the more complex example in (2.46) in a similar way. But this turns out not to be the case. To see this, consider the logical form in (2.48).

(2.48) $[\text{Max}]^1$ $[\text{t}_1$ said that he washed $\text{him}_1]$ he = Max

The reading represented by (2.48) is not an available reading for (2.46), so the LF should be ruled out. To see if it is, we should consider its binding alternative:

(2.49) [Max]^1 [t$_1$ said that he$_1$ washed him$_1$]

Is the transition from (2.48) to (2.49) illegitimate? Only if (2.49) is ruled out by another application of the Interface Rule. To see if it is, we must consider the binding alternative of (2.49):

(2.50) [Max]^1 [t$_1$ said that [he$_1$]^2 [t$_2$ washed him$_2$]]

The fact that [he] binds [him] in (2.50) is in conflict with Condition B. Now we can start to calculate backwards to the original LF: (2.50) is ruled out by Condition B; therefore, the Interface Rule rules out (2.49); and therefore, another application of the Interface Rule rules out (2.48). So the Interface Rule does account for the illegitimacy of (2.48), but in a roundabout way. And even more complex examples can easily be constructed of course.

In conclusion, Reinhart's Interface Rule successfully accounts for Condition B and Strong Crossover effects. Moreover, it allows for strict identity readings in VP ellipsis. But it does *not* account for Dahl's puzzle, its actual formulation is more complex than its underlying intuition, and its workings are (sometimes) undesirably complicated. This concludes my discussion of Reinhart's Interface Rule. Let me now summarize what has been established in this chapter.

2.7 Summary

Reinhart's (1983) proposal was based on two assumptions. First, binding is encoded in the syntax, while coreference is not. Second, speakers try to avoid ambiguity. It follows from these two assumptions that speakers will always prefer to use bound pronouns, which unambiguously encode the intended anaphoric relations, rather than referential pronouns, which could well be misinterpreted. This idea motivated Reinhart's Coreference Rule, which accounts for Condition B effects on coreference (given a syntactic Condition B constraint on binding) and also for cases in which pronouns are exceptionally allowed to corefer with one of their coarguments.

Heim (1998) noted that three aspects of Reinhart's proposal needed further attention. First, the Coreference Rule accounts for Condition B effects on coreference, but not for Condition B effects on cobinding and other kinds of codetermination. For example, cobinding is not ruled out in:

(2.51) Every man said that he voted for him.

Second, it rules out strict readings in cases of VP ellipsis such as:

(2.52) Max called his mother and Bob did too.

Third, it does not make precise what semantic indistinguishability means exactly.

In response to these issues, Heim does two things. First, she observes that semantic indistinguishability may be sensitive to "properties under discussion". This observation has been formalized here in terms of focus values. Second, she proposes a new rule, the Exceptional Codetermination Rule (ECR). This rule accounts for Condition B effects on cobinding and other kinds of codetermination, and it also allows for strict readings in VP ellipsis. But it does not preserve the *explanatory* aspect of Reinhart's approach, that is, it cannot be derived from the general assumptions that binding, but not coreference, is encoded in syntax, and that speakers try to avoid misinterpretation. Furthermore, it does not account for Strong Crossover effects and for Dahl's puzzle.

Fox (1999a) suggested that Strong Crossover effects and Dahl's puzzle may be derived from the general idea that syntactic derivations are subject to certain *economy* principles. The idea is that a complicated structure is not derived if there is a simpler structure with exactly the same interpretation. In particular, a structure may not be derived if there is an alternative structure in which binding relations are more local. This idea motivated Fox's Locality constraint, which accounts both for Strong Crossover effects and for Dahl's puzzle.

To enhance the empirical coverage of Fox's Locality constraint, Büring (2005b) combined it with Reinhart's Coreference Rule. In fact, he proposed a new constraint, Have Local Binding (HLB), which incorporates the effects of both the Coreference Rule and Locality. There are several problems with this proposal, however. First, it is not clear what the intuition is that underlies HLB. Reinhart's Coreference Rule and Fox's Locality constraint are derived from general ideas about the workings of grammar and the behavior of speakers in discourse. But these ideas really seem to be independent of one another. There does not seem to be an more general idea that underlies all of them. Therefore, it is unclear why the Coreference Rule and Locality should be unified, and not just be considered as two separate mechanisms working in tandem. The second problem with Büring's proposal is that it rules out strict readings in VP ellipsis, just as Reinhart's Coreference Rule did. This also has consequences for Dahl's puzzle. To make the right predictions, HLB must be combined with some ad-hoc, still to be worked out theory of VP ellipsis.

In the late 1990's, Reinhart herself departed from her 1983 theory because of the persistent problem with strict readings in VP ellipsis. The alternative view she developed, which took its final shape in (Reinhart, 2006), leaves the first assumption of her 1983 theory intact (binding is encoded in syntactic structure, coreference is not) but replaces the second assumption (speakers avoid ambiguity) with an alternative (hearers minimize interpretive options). In particular, Reinhart assumes that interpretations which are ruled out by constraints on binding are not sneaked in via other anaphoric mechanisms such as coreference and cobinding. This idea underlies the Interface Rule. We have seen that this rule accounts for Condition B effects and Strong Crossover effects, and has no trouble with strict readings in VP ellipsis. However, it does not account for Dahl's puzzle, and there is an undesirable mismatch between (the complexity of) its formulation and its workings on the one hand, and (the simplicity of) the underlying intuition on the other. These remaining issues will be addressed in the next chapter.