MAXIMALITY IN THE SEMANTICS OF WH-CONSTRUCTIONS

A Dissertation Presented

by

HOTZE RULLMANN

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1995

Department of Linguistics
MAXIMALITY IN THE SEMANTICS
OF WH-CONSTRUCTIONS

A Dissertation Presented

by

HOTZE RULLMANN

Approved as to style and content by:

Barbara H. Partee, Chair

Angelika Kratzer, Member

Margaret J. Speas, Member

James E. Cathey, Member

John J. McCarthy, Department Head
Department of Linguistics
ACKNOWLEDGMENTS

Writing this dissertation has been a challenging experience, but it would have been much more difficult without the support of a great many good people: teachers, fellow students, and friends (and fortunately I have experienced that the intersection of these three sets is by no means empty). I would like to use this opportunity to thank all those without whose help and friendship my career as a graduate student would never have come to a good end. After all, there aren’t that many moments in our lives when we can officially and publicly express our personal gratitude to the people who deserve it.

First of all I would like to thank the members of my dissertation committee, who kept faith in me despite slow progress. I owe a great debt to the chairperson of my committee, Barbara Partee. Despite her busy schedule as a department head she always found time to read and comment on my work. I always came out of those meetings with increased confidence and with a better understanding of the problem I was working on. Angelika Kratzer has also been an ideal committee member. She gave detailed comments on everything I wrote, and was sometimes critical, but always encouraging. Together, Barbara and Angelika have taught me so many things about semantics, both in the classroom and in private meetings. Peggy Speas gave many useful suggestions
from a syntactic point of view, and kept asking me tough questions. Jim Cathey provided a valuable outsider’s perspective on my work. All committee members have been unstinting with their time and commitment, willing to read long drafts in a short period of time and showing great flexibility in arranging meetings.

The linguistics department of the University of Massachusetts at Amherst is one of the most friendly and fun places to be a graduate student. It is also a very stimulating environment for learning and doing linguistics. I thank the faculty (past and present) for the numerous classes in which they shared their knowledge and showed how exciting linguistics can be. Many thanks to Emmon Bach, Lyn Frazier, Roger Higgins, Angelika Kratzer, John McCarthy, Barbara Partee, Tom Roeper, Lisa Selkirk, Peggy Speas, Edwin Williams, and Ede Zimmermann. I don’t think I could have gotten a better training in linguistics anywhere else in the world.

Earlier on in Holland I was also fortunate to have inspiring teachers. Johan van Benthem raised my interest in natural language by teaching generalized quantifier theory in his introductory logic course. I learned the first principles of syntax from Eric Reuland, whose course on the little Dutch word _er_ showed me that detailed analysis of a seemingly small problem may pay off. Jan Koster taught me to look at the big picture. I will always be grateful to him for suggesting I apply to UMass.
When I came to Amherst I was lucky to become part of a great group of classmates. I won’t forget the many lunches we had together at Earthfoods in our first year. Because of our shared experience of going through grad school I have always felt we somehow were a unit, despite our diverse characters and interests. I hope to stay in touch with you guys: Réjean Canac-Marquis, Veena Dwivedi, Kai von Fintel, Bill Philip, Bernhard Rohrbacher, and Sam Rosenthall.

I will always remember South College and its inhabitants with great fondness. I have enjoyed the company and friendship of many fellow students, faculty, and visitors. Two good friends I want to mention are Paul Portner and Jeff Runner. I have also much appreciated the friendship of (non-linguists) Mecke Nagel and Greg Leng.

There are various people with whom I had discussions about this work in its earlier stages of development. The issues I discussed with some of them unfortunately never made it to the final product, but I’m sure I will one day come back to them. Thanks for those conversations to Lee Baker, Ginny Brennan, Réjean Canac-Marquis, Diana Cresti, Veena Dwivedi, Kai von Fintel, Helen de Hoop, Jan Koster, Donna-Jo Napoli, Bill Philip, Paul Portner, Michael Rochemont, Jeff Runner, Henriëtte de Swart, Lisa Selkirk, Anna Szabolcsi, and Hubert Truckenbrodt.
Finally, I would like to express my gratitude to my parents for their financial and emotional support over all those years. Very special thanks go to Geertje Boschma for always being there even when she was far away.
This dissertation provides evidence that the notion of maximality plays a central role in the semantics of wh-constructions, in particular wh-questions, comparatives, and free relatives. It is argued that the semantics of each of these constructions involves reference to the maximal element of a certain set. The way maximality manifests itself depends on the algebraic structure of that set. Degrees are ordered linearly and the maximal element of a set of degrees is therefore the highest degree in that set. In a set in which the elements are only ordered in a join semi-lattice the maximal element is the sum of all the elements in the set.

In chapter 2, I show how the maximality account of comparatives can explain their semantic properties, in particular with respect to the distribution of negative polarity items and the interpretation of disjunction. Chapter 3 extends the maximality analysis to questions and free relatives. In free relatives maximality has a
‘universalizing’ effect, whereas in questions it gives rise to the exhaustiveness. Chapter 4 discusses a particular kind of questions, namely how many-questions, focusing on the scope interactions between the how many-phrase and other elements in the sentence.

A common thread running through the dissertation is the interaction between wh-movement and negation. ‘Negative’ elements can block wh-movement in certain cases, a phenomenon which is known in the literature as the negative island effect. In chapter 1 the negative island effect is introduced and the accounts of it found in the literature are briefly discussed. In chapters 2, 3, and 4 it is shown how maximality can account for the negative island effect. In chapter 5 certain remaining issues are discussed including the question whether the maximality account of negative island effects can be extended to other kinds of ‘weak’ or ‘selective’ islands, in particular wh-islands. A central issue in this matter is the proper division of labor between syntax and semantics. The chapter also discusses the role played by pragmatics in the explanation of the negative island effect and a maximality effect induced by focus.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ............................................................... iv

ABSTRACT .................................................................................. viii

CHAPTER

1. INTRODUCTION ....................................................................... 1
   1.1 Outline ................................................................................ 1
   1.2 Negative Islands ................................................................. 3
      1.2.1 The Basic Facts ......................................................... 3
      1.2.2 Negative Islands as Selective Islands ......................... 5
      1.2.3 Wh-Phrases Which Are Sensitive to Negative Islands ....... 9
      1.2.4 Elements Which Induce Negative Islands .................... 11
      1.2.5 Constructions in Which Negative Islands Occur ............ 16
   1.3 Existing Accounts of Negative Islands ................................. 17
      1.3.1 Rizzi (1990): Relativized Minimality .......................... 18
      1.3.2 Cinque (1990): D-Linking ........................................... 24
      1.3.3 Szabolcsi and Zwarts (1991): Monotonicity .................. 26
      1.3.4 De Swart (1992): Scope ............................................. 30
      1.3.5 Szabolcsi and Zwarts (1993): Algebraic Structure ......... 33

2. MAXIMALITY IN COMPARATIVES .......................................... 38
   2.1 Introduction ......................................................................... 38
      2.1.1 The Negative Island Effect in Comparatives ................. 38
      2.1.2 Outline of Chapter 2 .................................................... 41
      2.1.3 Limitations of the Analysis ......................................... 43
      2.1.4 Some Background Assumptions ................................. 47

   x
2.2 Comparatives and Maximality .............................................. 51
  2.2.1 Comparative Clauses Denote Maximal Degrees ............... 51
  2.2.2 The Interpretation of More Than-Comparatives ............ 56

2.3 Explaining the Negative Island Effect in Comparatives .......... 60

2.4 Inferential Properties of the Comparative ........................... 64
  2.4.1 Downward Entailment ....................................... 64
  2.4.2 Anti-Additivity ............................................. 69

2.5 Less Than-Comparatives ................................................. 75
  2.5.1 Maximality in Less Than-Comparatives ....................... 75
  2.5.2 Less Than-Comparatives and Monotonicity ................. 78
  2.5.3 The Ambiguity of Less Than-Comparatives ................. 82
  2.5.4 Why Aren’t Less Than-Comparatives Always Ambiguous? .... 88
  2.5.5 Towards a More Principled Account of the Ambiguity .......... 92

2.6 Equatives and Differential Comparatives ............................. 97
  2.6.1 Equatives .................................................. 97
  2.6.2 Differential Comparatives ................................... 103

2.7 Universal Quantification vs. Maximality ............................. 107

2.8 Further Issues in the Semantics of Comparatives .................. 112
  2.8.1 Unexpected Wide Scope Readings ............................ 112
  2.8.2 The Russell Ambiguity ..................................... 116

2.9 Conclusion .......................................................... 122

2.10 Degrees as Equivalence Classes (Appendix) ......................... 124
  2.10.1 Introduction ............................................... 124
  2.10.2 Constructing Degrees ...................................... 125
  2.10.3 Measuring Degrees ......................................... 127
  2.10.4 Some Complications ....................................... 131

3. MAXIMALITY IN QUESTIONS AND FREE RELATIVES ................. 135

3.1 Introduction ....................................................... 135

3.2 Negative Islands and Maximality in Questions .................... 136
5.4 Maximalind Effects Induced by Focus .......................... 221
  5.4.1 The Structured Meaning Approach to Focus .......... 221
  5.4.2 Negation and Focus .................................. 223
  5.4.3 Negation and Degree Expressions ...................... 226

BIBLIOGRAPHY .................................................... 229
CHAPTER 1
INTRODUCTION

1.1 Outline

This dissertation is concerned with the notion of maximality. It provides evidence that maximality plays a central role in the interpretation of \textit{wh}-constructions, in particular \textit{wh}-questions, comparatives, and free relatives. It will be argued that the semantics of each of these constructions involves reference to the maximal element of a certain set. The way maximality manifests itself depends on the algebraic structure of that set. Degrees for instance are ordered linearly and the maximal element of a set of degrees is therefore the highest degree in that set. This form of maximality is found in comparatives and degree questions. In a set in which the elements are only ordered in a join semi-lattice, however, the maximal element is the sum of all the elements in the set. This kind of maximality occurs in free relative clauses and in questions (other than degree questions).
In chapters 2, 3, and 4 of this dissertation I will investigate the semantics of the constructions mentioned above and show how maximality is involved in each instance. In chapter 2, I show how the maximality account of comparatives (originally due to von Stechow (1984a)) can account for their semantic properties, in particular with respect to the distribution of negative polarity items and the interpretation of disjunction. The discussion includes comparatives formed with less (rather than with more or -er) and equatives, arguing that the maximality account is better able to deal with the semantics of these constructions than a competing account involving universal quantification. Chapter 3 extends the maximality analysis to questions and free relatives. In free relatives maximality has a ‘universalizing’ effect (Jacobson (1990)), whereas in questions it gives rise to the phenomenon which in the literature has been discussed under the rubric of strong exhaustiveness (Groenendijk and Stokhof (1982)). In chapter 4, I discuss a particular kind of questions, namely how many-questions, focusing on the scope interactions between the how many-phrase and other elements in the sentence. In accounting for the scope ambiguities found in these questions (and the lack thereof in certain cases) the interpretation of wh-traces plays a crucial role.

A common thread running through the dissertation is the interaction between wh-movement and negation. ‘Negative’ elements can block wh-movement in certain cases, a phenomenon which is known in the literature as the negative island effect. In the rest of chapter 1 the negative island effect is introduced and the accounts of it found in the literature are briefly discussed. In chapters 2, 3, and 4 it will be shown how
maximality can account for the negative island effect. In the final chapter of the dissertation, chapter 5, certain remaining issues are discussed including the question whether the maximality account of negative island effects can be extended to other kinds of ‘weak’ or ‘selective’ islands, in particular \textit{wh}-islands. A central issue in this matter is the proper division of labor between syntax and semantics. The chapter also discusses the role played by pragmatics in the explanation of the negative island effect and a maximality effect induced by focus.

1.2 Negative Islands

1.2.1 The Basic Facts

One of the central topics of this dissertation is the fact that ‘negative’ elements such as negation and downward entailing quantifiers can interfere with \textit{wh}-movement. Two examples of this phenomenon are given in (1) and (2):

(1) a. I wonder how tall Marcus is.
    b. * I wonder how tall Marcus isn’t.

(2) a. I wonder how tall every basketball player is.
    b. * I wonder how tall no basketball player is.
While the (a) sentences are unexceptional, there is something very odd about the (b) sentences. Although nothing seems to be wrong with their syntax, (1b) and (2b) somehow don’t make any sense. However, it is hard to put a finger on what exactly is wrong with these sentences.

In Ross (1984), the earliest systematic discussion of the phenomenon, sentences such as (1b) and (2b) were assumed to instantiate a special kind of syntactic island, and were referred to as ‘inner islands’. Inner islands have also become known in the literature as ‘negative islands’, the term which I will use in this dissertation. The term ‘island’ may seem to prejudge the issue of what the exact nature of the phenomenon is, because it suggests that it has something in common with the well-known island effects first studied in Ross (1967). Since these island effects are generally assumed to be caused by certain locality constraints incorporated in the syntax, a similar kind of analysis in terms of syntactic locality would suggest itself for negative islands. In fact, such a syntactic account of negative islands has been argued for by Rizzi (1990), among others. This syntactic account has by no means gone unchallenged. Others have argued for a semantic explanation of negative islands (e.g. Szabolcsi and Zwarts (1991; 1993)). In this dissertation, I will use the term ‘negative island’ in a theory-neutral way to refer to whatever it is that is responsible for the contrast between (1a) and (2a) on the one hand and (1b) and (2b) on the other. I will call examples like (1b) and (2b) ‘ungrammatical’, using this term in a pre-theoretical sense without any prior commitment to what causes this ungrammaticality.
One of the goals of this dissertation is to explain the negative island effect. I will argue that the effect stems from a semantic maximality requirement associated with questions and other \textit{wh}-constructions such as comparatives and free relative clauses which show the same effect. My explanation can be summarized as follows: The embedded question in (1a) asks for the \textit{maximal} degree of tallness \(d\) such that Marcus is tall to degree \(d\) (or \(d\)-tall, for short). Analogously, (2b) asks for the maximal degree \(d\) such that Marcus is \textbf{not} \(d\)-tall. However, the set of degrees \(d\) such that Marcus is \textbf{not} \(d\)-tall does not contain a maximal member, which is why the sentence is ungrammatical.

\textbf{1.2.2 Negative Islands as Selective Islands}

Negative islands by no means block the extraction of all \textit{wh}-phrases. Extraction of a \textit{which}-phrase, such as \textit{which opponent}, is unproblematic. Compare (1) and (2) to (3) and (4), for instance:

(3)  
\begin{enumerate}
  \item a. I wonder which opponent Marcus can beat.
  \item b. I wonder which opponent Marcus can’t beat.
\end{enumerate}

(4)  
\begin{enumerate}
  \item a. I wonder which opponent every basketball player can beat.
  \item b. I wonder which opponent no basketball player can beat.
\end{enumerate}
These examples show that negative islands are *selective* islands in the sense of Postal (1992), that is, islands which only block the extraction of certain *wh*-phrases, but not others. By contrast, *non-selective islands* block the extraction of all phrases, including *which*-phrases. A good example are complex NP islands:

(5)  

*Complex NP island*

a. * I wonder how tall Marcus can beat the opponent who is.

b. * I wonder which coach Marcus can beat the man who knew.

Selective islands are more commonly known as *weak* islands, and non-selective islands as *strong* islands (Cinque 1990). However, as both Postal and Cinque himself (1990, p.161, footnote 2) point out, the weak/strong terminology is somewhat misleading. What distinguishes negative islands from complex NP islands is not that the former are ‘weaker’ than the latter, but that complex NP islands block all extractees, whereas negative islands only block some. I will therefore adopt Postal’s terminology, rather than Cinque’s, despite the fact that the latter appears to be rather well entrenched in the literature by now.

Negative islands are not the only class of selective islands. The selective islands that have been discussed most extensively in the literature are *wh*-islands, an example of which is given in (6):
(6)  *Wh-island*
  a. * I wonder how tall Mary asked whether Marcus was.
  b. ? I wonder which opponent Mary wonders whether Marcus can beat.

The *that*-complements of certain matrix predicates also behave like selective islands.
Cinque (1990) claims that the complements of factive verbs such as *regret* and extraposed subject clauses are selective islands:

(7)  *Factive island*
  a. * I wonder how tall Mary regrets that Marcus is.
  b. ? I wonder which opponent Mary regrets that Marcus can beat.

(8)  *Extraposition island*
  a. * I wonder how tall it matters that Marcus is.
  b. ? I wonder which opponent it matters that Marcus can beat.

Although the (b) examples of (6)-(8) are somewhat degraded, they are much better than the corresponding (a) examples, and certainly not as bad as (5b).

It is not quite clear what the exact extension of the class of selective islands is, nor whether selective islands form a homogeneous class that should be accounted for in a uniform way (a point which is also made by Szabolcsi and Zwarts (1993)). Hegarty
(1992) has challenged the claim that the class of matrix verbs inducing selective islands is coextensive with the class of factive verbs. *Agree*, for instance, is not factive but does seem to exhibit the selective island effect:

(9)  
\[\begin{align*}
\text{a. } & \quad * \text{ How tall do you agree that Marcus is?} \\
\text{b. } & \quad ? \text{ Which opponent do you agree that Marcus can beat?}
\end{align*}\]

It should also be noted that there is at least one difference between the negative islands and the other selective islands. As (3) and (4) show, extraction of a *which*-phrase does not lead to any decrease in grammaticality in the case of negative islands, but (6b)-(8b) demonstrate that extraction of a *which*-phrase out of any of the other selective islands does result in at least a slight degradation. (This degraded status of *which*-extractions is probably the reason why the term ‘weak island’ arose.)

Most of the literature on weak islands has focused on *wh*-islands and negative islands. In this dissertation I will mostly be concerned with negative islands. In the final chapter I will explore the question whether my account in terms of maximality can be extended to other selective islands. I won’t have anything to say in this dissertation about the status of non-selective (‘strong’) islands. I will assume that their status as island follows from purely syntactic locality restrictions, whatever the nature of these principles may be.
1.2.3 Wh-Phrases Which Are Sensitive to Negative Islands

The next question that arises is what kinds of wh-phrases are blocked by negative islands. I won’t try to give a complete inventory here, but merely mention some of the expressions that are sensitive to negative islands. The following examples show that in general all wh-phrases of the form how $A$ are blocked, where $A$ is a (gradable) adjective or adverb:

(10)  a. I wonder how heavy this piece of equipment is (*n’t).
    b. I wonder how crucial the coach considers each/*no game.
    c. I wonder how fast Lou can(*’t) run.
    d. I wonder how far Lou will/*won’t throw the ball.

The same is true for NP’s of the form how much of a(n) $N$:

(11)  a. I wonder how much of a problem this is(*n’t) going to be.
    b. I wonder how much of a help each/*no player will be.

Other NP’s denoting measures such as how many pounds, how many miles, etc., are also sensitive to negative islands:
What all these examples have in common is that they involve *wh*-phrases that refer to degrees. The sensitivity of degree expressions to negative islands will be discussed in chapters 2 and 3 of this dissertation. In chapter 4, I will turn to a different kind of *wh*-expression that also shows sensitivity to negative islands, namely *wh*-phrases of the form *how many N*. At first sight, these do not appear to be affected by negative islands:

(13)  
  a. I wonder how many opponents Marcus can/can’t beat.  
  b. I wonder how many books every student/no student read.

However, as will be discussed extensively in chapter 4, negative islands do have an effect on the scope of the *wh*-phrase in such questions (as was pointed out originally by Kroch (1989)).

Another set of expressions that are sensitive to negative islands are certain adjuncts such as *why* and *how*:

(14)  
  a. I wonder how the coach said/*didn’t say Marcus played *t.  
  b. I wonder why the coach said/*didn’t say Marcus left *t.
Note that these sentences are only ungrammatical on the reading in which how/why is extracted out of the embedded clause, that is, on the reading on which the negation has scope over the trace of the *wh*-phrase, indicated by *t*. Examples like these contrast sharply with cases in which the *wh*-phrase is an NP object of the embedded clause:

(15) a. I wonder which opponent the coach said/didn’t say Marcus played *t*.
    b. I wonder who the coach said/didn’t say Marcus played *t*.

In this dissertation, I will take examples like (10)-(12) in which the *wh*-phrase is some kind of measure expression as the central cases. In chapter 5, I will try to extend the analysis to cases like (14) involving *wh*-phrases such as how and why, and show why negative islands do not block the extraction of NP’s such as *which opponent* and *who* in (15).

### 1.2.4 Elements Which Induce Negative Islands

Negative islands are not only induced by the negation *not* and negative quantifiers like *no basketball player*, but by all expressions which denote *downward entailing* functions, as was pointed out by Rizzi (1990). (See also Szabolcsi and Zwarts (1991) for more elaborate discussion.) A downward entailing (or *monotone decreasing*) function is a function which reverses inclusion relations. Formally, a downward
entailing function can be defined as follows (cf. Ladusaw (1979); Barwise and Cooper (1981); Zwarts (1981, 1986)):

(16) **Definition of a Downward Entailing Function:**

A function \( f \) is **downward entailing** (or **monotone decreasing**) iff

for all \( X, Y \) in the domain of \( f \): if \( X \subset Y \), then \( f(Y) \subset f(X) \).

It is easy to show that the negation \( \neg \) is a downward entailing expression. Let’s assume for the sake of argument that \( \neg \) is a VP-operator and therefore denotes a function from sets to sets, namely the function that maps every set onto its complement. Now take two arbitrary sets \( A \) and \( B \), such that \( A \subset B \). For instance, let \( A \) be the denotation of the VP _are dancing_ (that is, the set of people who are dancing) and \( B \) the denotation of the VP _are moving_ (the set of people who are moving). Then it will be true that \( \neg -B \subset \neg -A \); the set of people who are not moving (= the denotation of the VP _are not moving_) is a subset of the set of people who are not dancing (= the denotation of the VP _are not dancing_).

The NP _no basketball player_ can also be shown to be downward entailing. Let’s assume, essentially following Barwise and Cooper (1981), that NP’s denote sets of sets of individuals, or in other words, functions from sets of individuals to truth values. The NP _no basketball player_ denotes the function \( f \) from sets of individuals to truth values such that \( f(X) = 1 \) iff \( X \) does not contain any basketball players. We furthermore make
the assumption that in the domain of truth-values (i.e. the set \{1,0\}) ‘\(\subset\)’ is interpreted as the material implication ‘\(\rightarrow\)’. Thus, we have \(0 \subset 0, 0 \subset 1, 1 \subset 1\), but \(1 \not\subset 0\). Now again, assume that \(A\) is the set of people who are dancing and \(B\) is the set of people who are moving. Then \(f(B) \subset f(A)\), because if \(No\ basketball\ player\ is\ moving\) is true, then certainly \(No\ basketball\ player\ is\ dancing\) must be true, too.

In general, in order to decide whether an NP is downward entailing we only have to check whether the implication in (17) holds:

\[
(17) \quad \text{NP is moving} \rightarrow \text{NP is dancing}.
\]

In (18) I show some downward entailing NP’s for which this implication does hold:

\[
(18) \quad \begin{align*}
\text{a.} & \quad \text{No player is moving} \rightarrow \\
& \quad \text{No player is dancing.} \\
\text{b.} & \quad \text{Fewer than ten players are moving} \rightarrow \\
& \quad \text{Fewer than ten players are dancing.} \\
\text{c.} & \quad \text{At most ten players are moving} \rightarrow \\
& \quad \text{At most ten players are dancing.} \\
\text{d.} & \quad \text{Few players are moving} \rightarrow \\
& \quad \text{Few players are dancing.}
\end{align*}
\]
(19) shows that all these NP’s do indeed block extraction of a *wh*-phrase like *how tall*:

(19)  
   a. * I wonder how tall *no player is.*
   b. * I wonder how tall *fewer than ten players* are.
   c. * I wonder how tall *at most ten players* are.
   d. * I wonder how tall *few players* are.

On the other hand, the italicized NP’s in (20) are not downward entailing, as shown by the invalidity of the following implications:

(20)  
   a. Marcus is moving $\not\Rightarrow$ Marcus is dancing.
   b. This player is moving $\not\Rightarrow$ This player is dancing.
   c. Every player is moving $\not\Rightarrow$ Every player is dancing.
   d. Most players are moving $\not\Rightarrow$ Most players are dancing.
   e. Some players are moving $\not\Rightarrow$ Some players are dancing.

Now let’s see whether these NP’s induce negative islands:
See de Swart (1991) for an extensive discussion of the monotonicity properties of adverbs.

\begin{enumerate}
  \item I wonder how tall \textit{Marcus} is.
  \item I wonder how tall \textit{this player} is.
  \item I wonder how tall \textit{every player} is.
  \item I wonder how tall \textit{most players} are.
  \item I wonder how tall \textit{some players} are.
\end{enumerate}

There are other downward entailing expressions besides negation and monotone decreasing NP’s. Other examples include adverbs like \textit{never} and \textit{seldom} and verbs such as \textit{fail} and \textit{deny}. That these expressions are downward entailing is demonstrated in (22) and (23),\footnote{See de Swart (1991) for an extensive discussion of the monotonicity properties of adverbs.} and that they induce negative islands is shown in (24) and (25):

\begin{enumerate}
  \item Jim never moves $\Rightarrow$ Jim never dances.
  \item Jim seldom moves $\Rightarrow$ Jim seldom dances.
\end{enumerate}

\begin{enumerate}
  \item Jim fails to move $\Rightarrow$ Jim fails to dance.
  \item Jim denies that Marcus is moving $\Rightarrow$ Jim denies that Marcus is dancing.
\end{enumerate}

\begin{enumerate}
  \item * I wonder how eager Marcus \textit{never} is.
  \item * I wonder how eager Marcus \textit{seldom} is.
\end{enumerate}
(25)  a.  * I wonder how eager John fails to be.
    b.  * I wonder how eager Jim denies that Marcus is.

1.2.5 Constructions in Which Negative Islands Occur

All the examples of negative islands I discussed so far involved (embedded) questions. However, negative islands are by no means restricted to questions. They can be observed in other *wh*-constructions as well, such as comparatives, free relatives, and pseudo-clefts. Some examples are given below:

**Comparative**

(26)  a.  These people weigh more than Bill does(*n’t).
    b.  Lou runs faster than Marcus can(*not) swim.

**Free relative clause**

(27)  a.  I don’t weigh what these people (*don’t) weigh.
    b.  Lou can run however fast Marcus can(*not) run.

**Pseudo-cleft**

(28)  What these people (*don’t) weigh is at least 300 pounds.
Most of the literature on negative islands has focused on questions, and there hasn’t been much overt discussion of negative island effects in other \textit{wh}-movement constructions, such as free relative clauses and comparatives, although most authors probably assume implicitly that their analyses carry over to such constructions. Interestingly, though, in the literature on comparatives we can find some discussion of cases such as (26) (see von Stechow (1984)), although this was not recognized as part of a larger pattern shared by other \textit{wh}-movement constructions. Von Stechow argues that comparative clauses can be analyzed as expressions that refer to \textit{maximal} degrees and that sentences like (26) are bad because in downward entailing contexts maximal degrees are undefined. In this dissertation, I will take von Stechow’s analysis of the comparative case as my starting point for a general analysis of negative island effects across \textit{wh}-movement constructions, arguing that all \textit{wh}-constructions sensitive to negative islands involve maximality in some way.

1.3 Existing Accounts of Negative Islands

In this section, I give a summary of various previous accounts of the negative island effect (and selective islands in general) found in the literature. Although I will criticize these earlier analyses for their shortcomings, I want to emphasize that my own account incorporates certain insights from each and is therefore indebted to all of them. This discussion of previous treatments will also set the stage for the presentation of my
own analysis, starting in chapter 2. I will present these analyses in roughly chronological order, but it should be remembered that they were developed more or less in the same time frame, with authors often reacting to early unpublished drafts of work that appeared in print only much later. Moreover, the reader should be keep in mind that the different authors do not all deal with the same range of facts. Some of them focus on the negative island effect, whereas others deal primarily with other selective islands, such as \textit{wh}-islands.

1.3.1 Rizzi (1990): Relativized Minimality

The most purely syntactic account of negative islands is that of Rizzi (1990), whose strategy is to assimilate negative islands as much as possible to \textit{wh}-islands. Rizzi’s point of departure is Huang’s (1982) influential observation that the extraction out of a \textit{wh}-island is much worse for an adjunct such as \textit{how} than it is for an argument like \textit{which car}:

\begin{equation}
\begin{align*}
\text{(29) a. } & \text{ ? Which car do you wonder how to fix } t \text{?} \\
\text{b. } & \text{ * How do you wonder which car to fix } t \text{?}
\end{align*}
\end{equation}

Rizzi points out that the characteristic responsible for this contrast can in fact not be the traditional argument/non-argument distinction, that is, the distinction between
obligatory elements that are governed by the verb and optional elements that are not. While it is true that in (29b) how is not an argument of the embedded verb fix, other verbs, such as behave, do take manner adverbials as arguments, witness the fact that they cannot be omitted (see (30a)). Nevertheless, extraction out of a wh-island is still bad, as shown in (b):

    b. * How do you wonder whether to behave t?

Rizzi argues that adverbials which are arguments of a verb are theta-marked. The ungrammaticality of adverbials under extraction out of a wh-island can therefore not be ascribed to a lack of theta-marking. Rather, Rizzi argues, the contrast between how and which car lies in the kind of theta-role these phrases receive. Theta-roles such as agent, theme, patient, experiencer, and goal are referential according to Rizzi, whereas other theta-roles like manner and measure are non-referential. The semantic difference between referential and non-referential theta-roles is that constituents that bear referential theta-roles refer to participants in an event, but the bearers of non-referential theta-roles merely qualify the event.

Additional support for Rizzi’s claim that the distinction responsible for the contrast between (29a) and (b) is not that between arguments and non-arguments but lies in the kind of theta-role the wh-phrase receives, comes from the behavior of
measure phrases. A measure phrase can be the argument of a verb, as shown by the following example:

(31) John weighs *(200 pounds).

Despite the argument status of a measure phrase, extraction out of a wh-island is ungrammatical:

(32) * How much do you wonder who weighs?

On Rizzi’s account, the ungrammaticality of (32) is due to the fact that measure phrases do not refer to participants in an event and therefore bear a non-referential theta-role.

Rizzi implements the referential/non-referential distinction in his theory of locality by stipulating that only constituents that bear a referential theta-role have an index. The locality conditions for indexed constituents are less strict than those for constituents without indices. In principle, according to Rizzi, there are two mechanisms for chain formation: binding and (antecedent) government. Binding (c-command by a coindexed constituent) can take place over arbitrary distances, but is only available for constituents that have an index. Constituents without a referential theta-role have to resort to antecedent government, a much more local relation. I will not go into Rizzi’s definition of government. For our purposes it is sufficient to know that under Rizzi’s
principle of Relativized Minimality, antecedent government in an A’-chain (A’-government) is blocked by any intervening A’-specifier. Thus, in the configuration (33), α cannot A’-govern γ if β is an A’-specifier:

(33) .... α .... β .... γ ....  

where α c-commands β and β c-commands γ.

Now let’s go back to (29), repeated here as (34):

(34)  
a. ? Which car do you wonder how to fix t?  
b. * How do you wonder which car to fix t?  

In (b) the intermediate wh-phrase which car is in an A’-specifier position (the Spec of CP) and therefore blocks government between how and its trace in the embedded clause. In (a) antecedent government between which car and its trace is likewise impossible because of the intervening A’-specifier how, but which car has the alternative option of binding its trace because it bears a referential theta-role (theme) and consequently has an index. Whatever residual unacceptability (a) has is due to the relatively weak locality condition of Subjacency.
So far I have sketched Rizzi’s account of why \( wh \)-complements are selective islands (weak islands in his terms). He tries to account for negative islands by making them look similar to \( wh \)-islands. In order to do this, he has to stipulate that negation, and downward entailing elements generally, are A’-specifiers and as such block the formation of A’-government chains. Rizzi assumes that sentential negation (\( not \) in English, \( pas \) in French) is an A’-specifier of the inflectional head T\(^0\) (Tense). Downward entailing quantifiers such as \( no \) \( job \) and \( few \) \( jobs \) have to be in an A’-specifier position at LF, according to Rizzi. His motivation for this claim comes from Klima’s (1964) observation that ‘affective’ elements (that is, elements that license negative polarity items) trigger subject/aux inversion in case they are preposed:

(35) a. With no job/few jobs would Bill be happy.
    b. * With some job/a few jobs would Bill be happy.

Rizzi takes these examples to show that, when preposed, downward entailing quantifiers occupy the Spec of CP triggering movement of the auxiliary to C, whereas quantifiers that are not downward entailing are adjoined to IP. What happens at S-structure in (35), must happen at LF in sentences without overt preposing. Downward entailing quantifiers obligatorily move to Spec of CP (an A’-specifier position) at LF, but non-downward entailing quantifiers are adjoined to IP. Because of their status as A’-specifiers, downward entailing quantifiers block antecedent government relations in A’-chains. Hence the negative island effect.
Rizzi acknowledges an important problem with his claim that downward entailing quantifiers are A’-specifiers at LF. In questions, the Spec of CP is already filled. Consider for instance (36):

(36) * How many pounds did nobody weigh?

_Nobody_ can’t move to the Spec of CP at LF because this position is already filled by the _wh_-phrase _how many pounds_, but _nobody_ does induce a negative island effect in this sentence. Rizzi suggests that this problem can be solved if we assume that the Spec of IP (the position of the subject) can optionally count as an A’-specifier at LF. However, it is clear that assumption won’t work if the downward entailing quantifier is an (indirect) object, as in (37):

(37) * How many pounds did you tell nobody you weighed?

To account for this sentence, Rizzi would have to claim that the indirect object position also can optionally count as an A’-specifier position at LF, but note that this would make the claim that all downward entailing quantifiers are A’-specifiers at LF effectively vacuous. If any NP-position can turn into an A’-specifier position, then the claim that downward entailing quantifiers are always in A’-specifier position at LF is a mere stipulation. Moreover, Rizzi’s only empirical argument for his contention that at LF downward entailing quantifiers are A’-specifiers whereas other quantifiers are
adjoined to IP, is the English inversion facts. These do not carry over to other languages. In Verb-Second languages like German or Dutch, any preposed constituent is in the Spec of CP and triggers inversion, but only downward entailing elements cause the negative island effect. So there is no cross-linguistic support for the claim that the negative island effect correlates with A’-specifier status. A further problem for Rizzi’s account is pointed out by Szabolcsi and Zwarts (1991). The verb deny induces negative island effects, as shown in (38), but since it is a head it cannot count as an A’-specifier, not even at LF:

(38) * How many pounds did John deny that he weighed?

We may conclude that Rizzi’s claim that downward entailing quantifiers behave like A’-specifiers at LF is essentially not more than a stipulation, and that his attempt to assimilate negative islands to wh-islands is therefore unsuccessful.

1.3.2 Cinque (1990): D-Linking

In Rizzi’s theory the difference between wh-phrases that can be extracted out of selective islands and those that can’t is located in theta-theory. A wh-phrase can be extracted out of a selective island only if it bears a referential theta-role. Cinque (1990) modifies Rizzi’s theory by introducing another notion of referentiality belonging to the
domain of pragmatics. According to Cinque, *wh*-phrases are referential in this pragmatic sense if they are *Discourse-linked* (or *D-linked* for short) in the sense of Pesetsky (1987).² A *wh*-phrase is D-linked if it quantifies over the elements of a set that is preestablished in the discourse context. *Wh*-phrases of the form *which N* are always D-linked. Thus, (39a) is only felicitous in a context where there is a preestablished set of students, where speaker and hearer both have a set of students ‘in mind’, so to speak. (39b) can also be used in such a context, but it can also be felicitous without a preestablished set to quantify over:

(39) a. Which students did you invite for the party?
   b. Who did you invite for the party?

Cinque points out (following Comorovski (1989a,b)) that D-linking greatly facilitates extraction out of a *wh*-island. In English, this can be illustrated by the contrast between (40a) and (b) (taken from Kroch (1989); Cinque’s own data are mostly from Romance):

² It is not very clear what the exact relationship is supposed to be between D-linking and referentiality in Cinque’s analysis. Sometimes he seems to equate referentiality with D-linking, but at another point in the text he says that “the notion of referentiality subsumes the notion of D-linking” (Cinque (1990), p. 16), adding in a footnote that “D-linking is only one way in which a phrase can become referential. Reference to specific members in the mind of the speaker is another.” (footnote 17, p. 164).
(40)  a. ??What were you wondering how to fix?
    b. Which car were you wondering how to fix?

*Which car* is D-linked and can be extracted out of a *wh*-island. D-linking of *what* is only marginally possible.

Cinque’s observation that D-linking facilitates the extraction of *wh*-phrases out of selective islands raises the question what the relationship is between Cinque’s pragmatic notion of referentiality and Rizzi’s notion of referential theta-roles. Cinque assumes that the latter concept is still needed. He implements his proposal in Rizzi’s theory by stipulating that a *wh*-phrase can only have an index (and be extracted out of a selective island) if two conditions are satisfied: (i) it must bear a referential theta-role (i.e. be referential in Rizzi’s sense); and (ii) it must be D-linked (i.e. be referential in Cinque’s sense). Cinque’s analysis therefore is of a somewhat dual nature, and it inherits the problems of Rizzi’s theory. His observation of the role played by D-linking is an important one, however, to which I will return in chapter 5.

1.3.3 Szabolcsi and Zwarts (1991): Monotonicity

As we have seen Rizzi starts out with a syntactic theory about *wh*-islands as selective islands and then tries to extend it to negative islands by treating all downward
entailing elements as A’-specifiers. Szabolcsi and Zwarts (1991) follow the opposite strategy. Their point of departure is the observation that downward entailing elements cause selective island effects, which they then try to extend to a general theory in which all selective islands are explained in terms of monotonicity.

The framework in which Szabolcsi and Zwarts formulate their theory is that of categorial grammar. In categorial grammar constituents can be combined not only by function-argument application, but also by function composition. The possibility of function composition is exploited for handling unbounded dependencies by composing the string of words between the wh-phrase and the corresponding ‘gap’ into one function. In the derivation of the sentence Who do you believe that Mary likes, for instance, the string do you believe that Mary likes is composed into one function which is then combined with the wh-phrase who. Szabolcsi and Zwarts argue that it is the monotonicity behavior of this composite function, which they call the ‘path’, that determines selective island-hood. Extraction of a wh-phrase which is sensitive to selective island effects, such as how or how tall, is only possible if the path is an upward *entailing* function. We have seen above that a downward entailing function is a function which reverses the inclusion relation $\subset$. An upward entailing (or *monotone increasing*) function, by contrast, is a function which preserves the inclusion relation:

\[ f \subseteq g \text{ for all } f \subseteq g \]

\[ f \subseteq g \text{ for all } f \subseteq g \]

---

3 An earlier and shorter version of this paper appeared as Szabolcsi and Zwarts (1990).
(41) Definition of an Upward Entailing Function:

A function $f$ is upward entailing (or monotone increasing) iff

for all $X$, $Y$ in the domain of $f$: if $X \subseteq Y$, then $f(X) \subseteq f(Y)$.

The monotonicity properties of the path depend on the monotonicity properties of the functions of which it is the composition. This can be illustrated with an example. In the sentence How tall do you think that everybody is? the path is the composite function do you think that everybody is. The individual functions that make up this path are all monotone increasing functions, and therefore the path itself also is monotone increasing. In *How tall do you think that nobody is?, on the other hand, one of the functions making up the path (nobody) is monotone decreasing, and as a result the path itself is monotone decreasing, and therefore constitutes a selective island. Szabolcsi and Zwarts present a simple calculus for determining the monotonicity behavior of a composite function on the basis of the monotonicity behavior of the functions that go into composing it, but the details of this need not concern us here.

Szabolcsi and Zwarts’s approach not only differs from that of Rizzi and Cinque in that it tries to account for all selective island effects in terms of monotonicity. It also introduces a new notion of referentiality. While for Rizzi referentiality is a notion of theta-theory and for Cinque it belongs to pragmatics, Szabolcsi and Zwarts turn it into a purely semantic notion. For them, the distinction between referential and non-referential wh-phrases (i.e. those that are not sensitive to negative islands and those that are) is
primarily one of semantic type: referential *wh*-phrases range over objects of type *e* and non-referential *wh*-phrases range over objects of higher types.⁴ According to Szabolcsi and Zwarts, the difference between *wh*-phrases that are sensitive to selective islands, such as *how* or *how tall*, and *wh*-phrases that aren’t, like *which student*, is that the latter range over individuals (objects of type *e*), whereas the former range over objects of higher types. Entities of type *e* form an unstructured domain; they are discrete individuals, which do not overlap or stand in inclusion relations to one another. Hence, we do not expect expressions of type *e* to ‘care’ about the preservation or non-preservation of inclusion relations, i.e. monotonicity. Inclusion relations do play a role in higher order domains, however, because such domains have algebraic structure and the elements of such domains are partially ordered. Therefore, expressions of higher types can be expected to be sensitive about the preservation or non-preservation of inclusion. As Szabolcsi and Zwarts themselves point out, this reasoning amounts only to a partial explanation of the selective island effects. It explains why expressions of type *e* cannot be sensitive to monotonicity and why expressions of higher types *can* be; it does not explain why expressions of higher types *are* in fact sensitive to the monotonicity of the path.

⁴ Independently, Frampton (1990; 1991) also reaches the conclusion that referentiality is a matter of semantic type. I will postpone discussion of his proposal until chapter 4.
1.3.4 De Swart (1992): Scope

In a reaction to Szabolcsi and Zwarts’s work, de Swart (1992) argues that what determines whether a given element causes selective island effects is not monotonicity, but scope. She is primarily concerned with two specific constructions that show sensitivity to selective islands, namely extraction of *combien* in French and so-called *wat voor*-split in Dutch. In both of these constructions a *wh*-specifier is extracted stranding the head noun. Discussing these constructions falls outside the scope of this thesis, but de Swart’s analysis seems to extend to selective island effects in general, although it is not quite clear from her paper whether she would want to commit herself to such a generalization. Here I will discuss de Swart’s analysis as if it pertained to selective island effects in general.

De Swart’s main claim is that all quantifiers cause selective island effects, regardless of their monotonicity properties. Apparent counterexamples to this, such as (42) arise because monotone increasing quantifiers like *every player* can ‘get out of the way’ by taking wide scope over the *wh*-phrase:

(42) How much does every player weigh?

That is, according to de Swart, (42) is only grammatical under a reading in which it gets the interpretation paraphrased in (43a). On this kind of reading the question requires a
pair list-answer (see Groenendijk and Stokhof (1984), May (1985)), such as for example (43b).\(^5\)

(43)  a. For every player, how much does he weigh?

For downward entailing quantifiers the option of getting wide scope in this way is not available, because for independent reasons the pair-list reading is not possible for downward entailing quantifiers (Groenendijk and Stokhof (1984)). Thus, while (44a) can be interpreted as in (44b), (45a) cannot have the reading paraphrased in (45b).

(44)  a. Who did every player invite?
       b. For every player, who did he invite?

(45)  a. Who did no player invite?
       b. * For no player, who did he invite?

De Swart concludes that the fact that downward entailing quantifiers cause selective island effects is due to their inability to take scope over a c-commanding wh-phrase.

\(^5\) For our present purposes it is not important whether the pair-list reading comes about by quantifier raising (as in May (1985)), by quantifying in (as in Groenendijk and Stokhof (1984)), or as a special case of the functional reading of questions (as proposed by Engdahl (1986) and Chierchia (1993)).
I agree with de Swart that (43a) is the most salient reading of (42), but I do not believe that the reading in which the quantifier has narrow scope with respect to the wh-phrase is excluded. This can be seen when we consider an example in which the quantifier can be forced to have narrow scope. Take for instance (46).

(46) How much does the coach want every player on the team to weigh?

Although this sentence can be interpreted by giving the quantifier wide scope over the wh-phrase how much (‘for every player, how much does the coach want him to weigh’), it clearly also has a reading in which every player on the team has narrow scope with respect to the verb want and hence also with respect to the wh-phrase. On the former, wide scope reading, the coach must have a de re wish about each of the players, whereas on the latter, narrow scope reading, this is not the case. The coach may simply have a general wish about the weight of his players, without having any particular wishes about the individual players. The possibility of a narrow scope reading can be demonstrated even more clearly if the quantifier includes a bound pronoun:

(47) How much do most coaches, want every player on their, team to weigh?

This sentence allows a reading in which most coaches has narrow scope with respect to the wh-phrase, and because most coaches binds the pronoun their the universal quantifier every player must also have narrow scope. We may conclude that it is
possible for upward entailing quantifiers to have narrow scope with respect to \textit{wh}-phrases like \textit{how much}. So even in cases where the quantifier cannot ‘get out of the way’ by taking wide scope, it does not cause a selective island effect. The contrast between upward entailing quantifiers like \textit{every player} or \textit{most players} and downward entailing quantifiers like \textit{no player} can therefore not be reduced to the difference in scope taking behavior with respect to \textit{wh}-phrases.

\section*{1.3.5 Szabolcsi and Zwarts (1993): Algebraic Structure}

In a more recent paper, Szabolcsi and Zwarts have proposed a drastic revision of their original theory.\footnote{An earlier version of the new theory appeared as Szabolcsi (1992).} Recall that in their previous paper only a partial explanation was given for the selective island effects. It was explained why expressions of type \textit{e can’t} be sensitive to monotonicity and why expressions of higher types \textit{can} be, but it was not explained why expressions of higher types \textit{are} in fact sensitive to monotonicity. In their 1993 paper, Szabolcsi and Zwarts try to remedy this defect of their earlier account by taking a somewhat more general and abstract perspective on ‘intervention phenomena’, as they call it. Rather than tying selective island effects to monotonicity, they try to explain how a ‘scopal element’ can interfere with the movement of a \textit{wh}-phrase in terms of the algebraic structure of the denotation domain of the \textit{wh}-phrase and the semantics of the scopal element (by a ‘scopal element’ they mean any item that
can participate in scope ambiguities, e.g. quantifiers, negation, conjunctions, intensional verbs, etc.).

According to Szabolcsi and Zwarts, each scopal elements is associated with certain algebraic operations. Negation, for instance, is associated with taking complements, universal quantification corresponds to meet (intersection), existential quantification corresponds to join (union) etc. Referential domains, on the other hand, form algebraic structures, such as Boolean algebras, lattices, and semi-lattices, on which such operations may or may not be defined. A Boolean algebra is an algebraic structures in which the operations of complement, meet, and join, are all defined. In lattices meet and join are defined, but complement isn’t. In a join semi-lattice is join is defined, but not meet and complement.

Szabolcsi and Zwarts’s central claim is that \textit{wh}-movement is blocked by a given scopal element if the algebraic operation associated with it is not defined for the referential domain that the \textit{wh}-phrase ranges over. To take an example, according to Szabolcsi and Zwarts the \textit{wh}-word \textit{how} ranges over the domain of manners, which they claim forms a (free) join semi-lattice. Since complement is not defined in join semi-lattices, it is predicted that negation (\textit{not}) should block extraction of \textit{how}, which is indeed the case:
To reiterate, in Szabolcsi and Zwarts’s more recent account selective island effects depend on the interaction between the algebraic operations associated with scopal elements on the one hand and the denotation domains that \(wh\)-phrases range over, on the other. A selective island effect arises whenever a \(wh\)-phrase is extracted over a scopal element that requires an algebraic operation that is not defined in the denotation domain of the \(wh\)-phrase.

Why are \(wh\)-phrases that range over individuals, such as *which student*, not sensitive to selective island effects? The reason is, according to Szabolcsi and Zwarts, that individuals (entities of type \(e\)) are collected into unordered sets, and the operations of complement, meet, and join are all three defined for sets. Therefore \(wh\)-phrases ranging over individuals can be extracted over all types of scopal elements.

Szabolcsi and Zwarts’s (1993) paper is quite programmatic in nature. As the authors note themselves, they do not address \(wh\)-islands and only make some speculative suggestions about ‘factive’ islands. The paper also leaves open a number of questions. For instance, it is unclear exactly what the evidence is for claiming that a certain denotation domain has a certain algebraic structure. Manners are said to form a free join semilattice, a structure in which join is defined, but complement and meet are not. The only evidence for this claim appears to be the fact that, according to Szabolcsi
and Zwarts, extraction of *how* is blocked both by negation and by universal quantifiers. Now note that this kind of reasoning is in danger of being circular. Extraction of *how* is blocked by negation and universal quantifiers because manners form a join semi-lattice, but the evidence for the latter claim is precisely the fact that negation and universal quantifiers block the extraction of *how*.  

Another question left unanswered by Szabolcsi and Zwarts (1993) pertains to their explanation of the fact that *wh*-phrases which range over individuals are not sensitive to selective islands. According to Szabolcsi and Zwarts, individuals form an unstructured domain. (Note by the way that this is in contrast to a theory like that of Link (1983) in which the domain of individuals is not an unordered set but forms a complete join semi-lattice.) Now obviously algebraic operations like complementation are not defined on domains without any algebraic structure. We cannot take the complement of the individual John and form an individual not-John. On Szabolcsi and Zwarts’s assumptions, this would lead us to expect that (contrary to fact) *wh*-phrases ranging over individuals should be sensitive to negative islands, because complementation is not defined on individuals. Szabolcsi and Zwarts get around this by assuming that the algebraic operations are not performed on the individuals themselves.

---

7 Similarly, Szabolcsi and Zwarts make a distinction between what they call ‘amounts’ and ‘numbers’. Intersection is supposed to be defined for numbers, but not for amounts. But again, the evidence given in support of this difference in algebraic structure is based on data involving ‘number-interpretations’ and ‘amount-interpretations’ of *wh*-phrases of the form *how many* N with respect to extraction across universal quantifiers, data which the account in terms of algebraic structure is supposed to account for in the first place.
but rather on sets of individuals. The power set of the set of all individuals forms a Boolean algebra, and Boolean algebras are closed under the operations of complement, meet (intersection), and join (union). But this immediately raises the question why similarly we can’t form sets of manners and take the complement of those, rather than of the manners themselves. Szabolcsi and Zwarts assume that individuals can be collected into sets so that algebraic operations can be performed on those sets, but that this option is not available for objects of higher types, like manners. As far as I can tell, the question why there should be this difference is not addressed in their account.

Despite these open questions, Szabolcsi and Zwarts’s paper raises the important issue of the role of the algebraic structure of referential domains in an explanation of selective island effects. In fact, algebraic structure will also play a role in my own account of the negative island effect, albeit in quite a different way.
CHAPTER 2
MAXIMALITY IN COMPARATIVES

2.1 Introduction

2.1.1 The Negative Island Effect in Comparatives

In the previous chapter I have summarized various accounts of the negative island effect (and other selective island effects) found in the literature, which for the most part concentrate on questions. In this chapter, I will discuss one other work, which examines the negative island effect in comparatives, namely von Stechow (1984a).1 His explanation of the effect is embedded in a general discussion of the

---

1 The ungrammaticality of negation in a comparative clause was already pointed out by Lees (1961) in what was probably the first discussion of comparatives in generative grammar. Since then, the same observation has been made by a number of authors (among others Huddleston (1967), Ross (1969), Green (1970), Cresswell (1976)). Ross (1980) is the first more extended discussion. Interestingly, Huddleston (1967) already stressed the parallelisms between comparatives and questions with respect to their sensitivity to negation, citing examples such as the following (p. 97):
   (i) * How talkative isn’t he?
   (ii) * Mary’s more tactful than Jill isn’t talkative.
semantics of comparison; he does not relate the pattern of data he finds with comparatives to parallel patterns found in other *wh*-movement construction such as questions.

Von Stechow discusses examples like the following:

(1)  
a. John weighs more than Bill weighs.  
b. John weighs more than every one of us weighs.  
c. * John weighs more than no one of us weighs.

With the benefit of hindsight it seems plausible to suggest that the ungrammaticality of (1c) is an instance of the negative island effect. It is not just *no one* which causes the comparative to be ungrammatical, but any downward entailing element:

(2)  
a. * John weighs more than Bill *doesn’t* weigh.  
b. * John weighs more than *nobody* weighs.  
c. * John weighs more than *few people* weigh.  
d. * John weighs more than *fewer than five people* weigh.  
e. * John weighs more than *at most five people* weigh.  
f. * John weighs more than Bill *never* weighed.  
g. * John weighs more than Bill *seldom* weighed.
Contrast these examples with the following sentences in which the downward entailing element is replaced by an upward entailing one:

(3)  

a. John weighs more than Bill weighs.  
b. John weighs more than everybody else weighs.  
c. John weighs more than most people weigh.  
d. John weighs more than many people weigh.  
e. John weighs more than at least five people weigh.  
f. John weighs more than Bill always weighed.  
g. John weighs more than Bill often weighed.  

Assuming that the negative island effects in questions and comparatives are indeed one and the same phenomenon, we could try to account for the ungrammaticality of sentences like (2a-g) by extending one of the existing theories of negative islands to comparatives. I will pursue the opposite strategy and argue that von Stechow’s treatment of negative islands in comparatives can be extended to other wh-movement constructions to give an insightful analysis of negative island effects in general.

Von Stechow (1984a) gives an overview and comparison of various theories of the semantics of comparatives\(^2\) and then offers his own synthesis. His main conclusion

is that comparative clauses are expressions which denote maximal degrees. This allows him to explain the negative island effect, because, as will be discussed below, in downward entailing contexts maximal degrees are not defined. Von Stechow gives an extensive motivation of his account of the semantics of comparatives by showing that, in addition to the negative island effect, it explains a number of interesting semantic properties of the comparatives, in particular the fact that comparatives are downward entailing and anti-additive (in the sense of Zwarts (1981; 1986) and Hoeksema (1983)), and can license negative polarity items. However, von Stechow only discusses comparatives which are formed with the affix -er or with more, leaving out of the discussion comparatives with less and equatives (as... as...). In this chapter, I will show how von Stechow’s maximality analysis can be extended to account for the interpretation of these types of comparatives as well, bringing to light certain interesting ambiguities that are crucial for determining the monotonicity properties of these constructions, but which are not found in the types of comparatives discussed by von Stechow.

2.1.2 Outline of Chapter 2

The organization of this chapter is as follows. I will first outline von Stechow’s analysis (section 2.2), and show how it accounts for the negative island effect (section 2.3) and other properties of comparatives (section 2.4). Then, in section 2.5, I will give
additional support for the semantic analysis of comparatives in terms of maximality by showing that it can not only account for the types of comparatives that von Stechow discusses, but also for comparatives with *less*. Interestingly, these comparatives exhibit an ambiguity which affects their behavior in terms of monotonicity and negative polarity licensing. On one reading, comparatives with *less* are downward entailing and license negative polarity items, while on the other reading, they are upward entailing and do not license negative polarity items. I will show that von Stechow’s maximality analysis can be extended in such a way that it predicts the monotonicity and polarity licensing behavior of both readings. In section 2.6, I turn to equatives and so-called differential comparatives (like *John is two inches taller than Bill is*), showing that these also can be interpreted in two ways. Although this is an ambiguity of a different kind than the one exhibited by comparatives with *less*, it is also one which affects monotonicity and negative polarity licensing. Section 2.7 compares the maximality account of comparatives to an alternative account in terms of universal quantification over degrees. It is concluded that the maximality account is more successful than the universal quantification account in dealing with comparatives with *less* and with equatives and differential comparatives. In section 2.8, I discuss two further issues in the semantics of comparatives. The first, discussed in section 2.8.1, involves the surprising fact that NP’s in comparative clauses can often get wide scope outside the comparative clause. This is a problem for any theory of the comparative that I know of, including the maximality account advocated here. The second issue, which is the topic of section 2.8.2, is the well-known ambiguity of sentences like *I thought your yacht was larger than it is*. Von
Stechow’s own account for this ambiguity involving quantifier raising (QR) is unsatisfactory, as shown by Hoeksema (1984) and Heim (1985), and it therefore does not offer any support for the maximality account. However, since the correct solution, which involves quantification over possible worlds, is perfectly general, it works equally well in the maximality account as in any other account of the semantics of comparatives. Section 2.9 sums up the main conclusions of this chapter. Section 2.10 is an appendix to this chapter in which it is outlined how the degrees can be defined as equivalence classes of entities.

In the rest of this section, I will first mention some aspects of the comparative construction that fall outside the scope of this study, and then I will sketch some background assumptions.

2.1.3 Limitations of the Analysis

The comparative construction is well known for its complexity, presenting many puzzles to both syntactic and semantic theory. Although much research has been done on comparatives, no consensus has been reached on many, if not most, aspects of the construction. It will therefore not come as a surprise that I will not try to come up with an overall theory of comparatives in this study. In particular, I don’t have anything to say about the syntax of comparatives. I won’t take a stance, for instance, on the hotly-
debated issue of the nature of the syntactic relation the head of the comparative and the
gap inside the comparative clause.\footnote{This relation has been argued to involve unbounded deletion (Bresnan 1976; 1977), successive cyclic wh-movement (Chomsky 1977), percolation of ‘slash’ categories (Gazdar 1981), binding (Pinkham 1982), or an extraction operator in the framework of the Lambek calculus (Hendriks 1992a and forthcoming).} Except for some remarks in section 2.5.5, I will also remain non-committal as to the relationship between comparative deletion sentences like (4a) and subdeletion sentences like (4b):

\begin{align*}
(4) & \quad \text{a. The table is longer than the desk is.} \\
& \quad \text{b. The table is longer than the desk is wide.}
\end{align*}

For the sake of concreteness I will assume that at the syntactic level that is the input to semantic interpretation (‘logical form’) (4a) is converted into something like (5) by copying the head of the comparative (the adjective \textit{long}) into the gap in the comparative clause:

\begin{align*}
(5) & \quad \text{The table is longer than the desk is long.}
\end{align*}

Another restriction that I will impose on myself is that I only discuss the semantics of clausal comparatives, that is, comparatives in which \textit{than} is followed by a complete clause, such as the examples in (6), and no phrasal comparatives like those in (7):
Von Stechow (1984), following Bresnan (1973), assumes that phrasal comparatives are derived by ellipsis from clausal comparatives. In the literature, various proposals can be found for generating (some or all) phrasal comparatives directly rather than by ellipsis from a clausal comparative (Hankamer 1973; Gazdar 1981; Napoli 1983; Hendriks (forthcoming)). The evidence offered in support of such analyses is, for the most part, of a syntactic nature. However, Hoeksema (1983; 1984) argues that (certain) phrasal comparatives also differ semantically from clausal comparatives, in particular with respect to monotonicity and negative polarity licensing. Counter arguments to Hoeksema’s claims are offered by Heim (ms. 1985) and Hendriks (1993 and forthcoming).

(6) a. John is taller than Bill is.
   b. Mary ran faster than I expected her to run.

(7) a. John is taller than Bill.
   b. Mary ran faster than a tornado.
   c. Mary ran faster than yesterday.

I will ignore the question whether and how phrasal comparatives are related to clausal ones by either syntactic or semantic means, as well as how various forms of ellipsis in comparative clauses are to be accounted for.

Another type of comparative construction I won’t discuss are so-called multihead comparatives like (8) (von Stechow (1984a), Hendriks (1992b)):

(8) More silly lectures have been given by more silly professors than I expected.

---

4 Von Stechow (1984), following Bresnan (1973), assumes that phrasal comparatives are derived by ellipsis from clausal comparatives. In the literature, various proposals can be found for generating (some or all) phrasal comparatives directly rather than by ellipsis from a clausal comparative (Hankamer 1973; Gazdar 1981; Napoli 1983; Hendriks (forthcoming)). The evidence offered in support of such analyses is, for the most part, of a syntactic nature. However, Hoeksema (1983; 1984) argues that (certain) phrasal comparatives also differ semantically from clausal comparatives, in particular with respect to monotonicity and negative polarity licensing. Counter arguments to Hoeksema’s claims are offered by Heim (ms. 1985) and Hendriks (1993 and forthcoming).
Finally, there is one issue which is ignored in this dissertation which is potentially relevant for our understanding of the negative island effect. This is the possibility of the occurrence of a ‘pleonastic’ negation in comparative clauses in certain languages like French, Spanish, and Italian (Jespersen (1917), Rivero (1970), Napoli and Nespor (1975)). The following examples are from Italian (= Napoli and Nespor’s ex. (4) and (5)):

(9) a. Maria è più intelligente di quanto non sia Carlo.
   Maria is more intelligent than not is (subjunctive) Carlo.
   ‘Maria is more intelligent than Carlo is.’

b. Maria è più intelligente di quanto tu non creda.
   Maria is more intelligent than you not believe (subjunctive).
   ‘Maria is more intelligent than you believe.’

The negation occurs in a position in which it would cause ungrammaticality in the corresponding sentences in other languages, like English or Dutch, but surprisingly, the sentence is equivalent in meaning to its English counterpart without the negation.\(^5\) The cross-linguistic tendency to insert a negation in the comparative clause is an intriguing...

\(^5\) This is not quite accurate. Napoli and Nespor argue that the pleonastic negation in comparatives in Italian does have an effect on the meaning, in that it occurs “when the speaker presupposes that his statement contradicts someone else’s or his own previously held belief.” (Napoli and Nespor 1976, p. 811) They also point out that pleonastic negation requires the use of the subjunctive in the comparative clause, whereas the indicative is used in comparatives without negation.

46
Van der Wouden (1993) suggests that the occurrence of what he calls ‘paratactic’
negation in comparative clauses (and other environments) is due to the downward
entailing character of the construction. I suspect that it is no coincidence that in some
languages the occurrence of negation in a comparative clause leads to
ungrammaticality, while in others negation may occur in that environment, but when it
does, it appears to be spurious semantically. I will leave this interesting issue for further
research.

2.1.4 Some Background Assumptions

Many, if not most, analyses of the semantics of comparatives, starting with
Russell (1905) and including Cresswell (1976), Hoeksema (1983), and von Stechow

6 Seuren (1973; 1984) uses data like these to support his claim that comparative
clauses in English contain a negation at underlying structure. He cites further historical
and dialectal data from English as support for his analysis. Modern English than is
claimed to be historically derived from pon-ne (‘by which not’). In certain English
dialects nor can be used instead of standard English than (data attributed to Joly
(1967), which I have not seen) (Seuren (1973) ex. (43)):
(i) He is richer nor you’ll ever be.
According to Seuren, Cockney English also allows the use of overt negatives in
comparative clauses (Seuren (1973) ex. (48)):
(ii) She did a better job than what I never thought she would.
Although such data are indeed intriguing, von Stechow (1984b) is quite right in
pointing out that an analysis of the comparative in modern standard English ultimately
has to be supported by internal data alone. Seuren’s principal argument is the behavior
of negative and positive polarity items, which will be discussed in detail below.
(1984a), involve reference to degrees. The basic idea is that in a sentence like (10) we do not directly compare Mary to Sue, but rather what is compared is the degree to which Mary is tall and the degree to which Sue is tall:

(10) Mary is taller than Sue is.

If we analyze comparatives in terms of degrees, then of course the question arises what kind of things degrees are. A number of authors (Cresswell (1973), Klein (1980; 1991), and Hoeksema (1983)) have shown that degrees can be construed as equivalence classes of ordinary objects. Since my concern in this chapter is with the logic of comparative sentences and not with the ontology of degrees, I will leave discussion of this idea and of some of its complications for section 2.10, which functions as an appendix to this chapter. For my present purposes it is sufficient to note that degrees are ordered by a smaller-than relation ≤, which is reflexive, transitive, anti-symmetric, and connected (that is, a linear order).

A gradable predicate like tall or weigh can be analyzed as a relation between entities and degrees. The proposition that Mary is tall to degree d can then be expressed by an atomic formula like ‘tall(mary,d)’. In informal representations, I will often write ‘Mary is d-tall’, instead.
Measure phrases, such as *six feet* or *two inches*, can be used both in combination with the positive form of the adjective and with the comparative:

(11) a. Mary is six feet tall.
    b. Mary is taller than six feet.
    c. Mary is two inches taller than Sue is.

Following most of the literature, I will assume that such measure phrases are names for degrees. Thus, for instance, *six feet* will denote the equivalence class of all objects that are six feet tall. This analysis has the advantage that an example like (11a) can simply be analyzed as expressing the proposition that Mary stands in the two-place relation tall to the degree denoted by *six feet*:

(12) \text{tall(mary,6ft)}

I will assume that (12) is true iff Mary is *exactly* six feet tall. It has sometimes been argued in the literature (e.g. Horn (1972; 1989), Gazdar (1979)) that, as far as truth conditions are concerned, *Mary is six feet tall* only means that Mary is *at least* six feet tall, and that the conclusion that Mary is not more than six feet tall is a matter of a Gricean scalar implicature. (But see also Horn (1992) and references cited there for arguments showing that the Gricean line is much harder to defend for explicit measure phrases and numerals than it is for other cases.) In this dissertation I will adopt the
‘exactly’ interpretation, but it should be stressed that I am mainly making this assumption in order to keep the discussion as simple as possible. The essentials of my analysis are equally compatible with the ‘at least’ interpretation. In some footnotes I will briefly discuss how my analysis would be affected if I assumed the ‘at least’ interpretation rather than the ‘exactly’ interpretation that I will stick to in the main text.

The positive use of the adjective without an accompanying measure phrase (as in *Mary is tall*) can be interpreted as saying that Mary stands in the tall-relation a degree $d$ such that $d$ is greater than a contextually determined standard of tallness $s$: $\exists d [\text{tall}(m, d) \land d > s]$ (cf. the use of an operator $pos$ in Cresswell (1976) and von Stechow (1984a)).

To conclude this section, let me say something about the logical translation language I will be using in this chapter, as well as in the rest of the dissertation. In principle, I am assuming an intensional language with quantification over possible worlds in the object language, as in Gallin’s (1975) Ty2. (For an introduction to Ty2, see Gamut (1991).) The basic types are $e$ (for entities), $t$ (for truth values), and $s$ (for possible worlds). $x$ and $y$ will be used as variables of type $e$, and $w$ as a variable of type $s$, with subscripts and primes added when necessary. The type of degrees will be called $d$. Since degrees can be construed as equivalence sets of entities (see section 2.10), $d$ can be taken as a abbreviation of the type $<e,t>$. In logical representations, I will use $d$ (again with subscripts or primes when necessary) as a variable ranging over degrees,
but no confusion should arise. Most of this chapter deals with extensional phenomena, in which case I will omit the possible world variables for the sake of simplicity. Often informal representations will be used, mixing ordinary English with logical expressions.

2.2 Comparatives and Maximality

2.2.1 Comparative Clauses Denote Maximal Degrees

Central to von Stechow (1984a) analysis of comparatives is the idea, ultimately going back to Russell (1905), that a comparative clause is a denoting expression referring to a (unique) degree. For instance, the complement clause of than in a sentence like (13) denotes the degree to which Bill is tall:

(13) John is taller than Bill is.

This sentence is true iff John is tall to a degree which exceeds the unique degree to which Bill is tall. If Bill is six feet tall, then the comparative clause in (13) will denote the degree six feet. (13) will therefore be true iff (14) is true:
The categorical status of than is not very clear. It is sometimes assumed that than is a complementizer. Evidence against this comes from Dutch which has a comparative construction containing an overt complementizer dat ('that') following dan ('than') (see den Besten 1978):

(i) I ga liever naar Praag dan dat ik naar Berlijn ga.
    I go rather to Prague than that I to Berlin go.
    ‘I’d rather go to Prague than to Berlin’

Den Besten suggests that than is a preposition. Others have argued that (at least in some cases) than acts as a conjunctor, because the possibilities for ellipsis in the

(14) John is taller than six feet.

One advantage of analyzing comparative clauses as expressions that denote degrees is that it assigns (13) and (14) the same logical structure.

Implementing the idea that the comparative clause is a degree-denoting expression is rather straightforward. Bresnan (1973) has shown that comparative clauses contain a gap corresponding to a degree denoting expression:

(15) a. This desk is wider than it is _ high.
    b. * This desk is wider than it is four feet high.

For the sake of concreteness I will assume, following Chomsky (1977), that this gap is created by an instance of wh-movement, but the essentials of the analysis are equally compatible with other syntactic assumptions. The gap in (15a) is translated as a variable over degrees $d$, which is bound by an implicit operator $Op$, in the Spec-of-CP of the than-clause:

---

7 The categorical status of than is not very clear. It is sometimes assumed that than is a complementizer. Evidence against this comes from Dutch which has a comparative construction containing an overt complementizer dat ('that') following dan ('than') (see den Besten 1978):

(i) I ga liever naar Praag dan dat ik naar Berlijn ga.
    I go rather to Prague than that I to Berlin go.
    ‘I’d rather go to Prague than to Berlin’
As mentioned earlier, I will assume that in a ‘comparative deletion’ sentence like (13) the adjective tall is copied into the comparative clause, resulting in a structure like (17):

(17) John is taller than \([_{CP} Op, [_{IP} Bill is d_{tall}]]\).

Given these basic syntactic assumptions, the most straightforward way of making the comparative clause into a degree-denoting expression is by translating \(Op\) as the iota-operator:

(18) John is taller than \(\epsilon[Bill is d\text{-tall}].\)

This way, the comparative clause is treated as a definite description of degrees, an analysis that goes back to Russell (1905). Von Stechow, however, gives data that show that use of the iota-operator will actually give the wrong result. Consider the following examples:

(19) a. John is richer than a linguist can be.

b. John is richer than \(\epsilon[\text{a linguist can be } d\text{-rich}].\)

comparative clause appear to be the same as those in conjunctions (cf. Napoli (1983) and Hendriks (1991 and forthcoming)).
(20)  a. John swam faster than Bill could run.
    b. John swam faster than \( \iota d \) [Bill could run \( d \)-fast].

Take (19a) first. According to its ‘translation’ in (19b), that sentence says that John is richer than the unique degree \( d \) such that a linguist can be \( d \)-rich. But of course there is no such unique degree, because some linguists happen to be richer than others. Similarly for (20a), which according to (20b), will be true iff John swam at a speed which exceeds the unique speed at which Bill could run. But obviously there is no such unique speed. If Bill could run, say, 5 miles per hour, then he could also run 4, or 3, or 2 miles per hour. Thus, the use of the iota-operator predicts that in these examples the referent of the comparative clause is undefined, and that they should therefore lack a truth value. This is clearly the wrong result.

Von Stechow proposes that instead of referring to the unique speed at which Bill can run, the comparative complement in (20) refers to the maximal speed at which he can run. This proposal can be implemented by means of an operator \( \text{max} \) which maps a set of degrees onto the degree in the set that is greater than or equal to all other degrees in the set:
(21) **Definition of the Maximality Operator max:**

Let DEG be a set of degrees ordered by the relation $\leq$, then

$$\text{max}(\text{DEG}) = \{d \mid d \in \text{DEG} \land \forall d' \in \text{DEG} [d' \leq d]\}.$$ 

Using $\text{max}$, we can now spell out the interpretation of (19a) and (20a) as (22) and (23), respectively:

(22) John is richer than $\text{max}(\lambda d[\text{a linguist can be } d-\text{rich}]).$

(23) John swam faster than $\text{max}(\lambda d[\text{Bill could run } d-\text{fast}]).$

These representations predict the truth conditions of the sentences correctly: John is rich to a degree which is greater than the maximal degree to which a linguist can be rich, and John swam at a speed exceeding the maximal speed at which Bill could run.

(22) and (23) are still very informal and incomplete representations of the meanings of those sentences. I will give a more detailed account of the semantics of comparatives in the next subsection. In section 2.3 I will then show how maximality can explain the negative island effect in comparatives. In the rest of the chapter is devoted to further empirical arguments in support of the maximality account of comparatives.
2.2.2 The Interpretation of More Than-Comparatives

In discussing the semantics of comparative construction, I have so far concentrated on the denotation of the than-clause. In this section I will complete the picture by showing how the meaning of the than-clause fits into the semantics of the comparative construction as a whole, and sketch how the meaning of comparative sentences is built up compositionally. As already indicated, I will only give an explicit semantics for adjectival comparatives (and only those adjectival comparatives in which the adjective is used predicatively). For the moment I will moreover limit myself to what I will call more than-comparatives, that is, comparatives involving either the comparative form of the adjective (as in (24a)) or the adjective preceded by the modifier more (as in (24b)). Less than-comparatives (like (24c)) and equatives (such as (24d)) will be discussed in later sections. The same applies to what von Stechow has called differential comparatives, comparatives which have an overt measure phrase that indicates the difference between the two degrees that are being compared (cf. (24e)).

(24) a. John is taller than Bill is.
    b. John is more beautiful than Bill is.
    c. John is less tall than Bill is.
    d. John is as tall as Bill is.
    e. John is two inches taller than Bill is.
The discussion in this subsection again draws heavily on von Stechow’s account. To start, consider the following basic example:

(25)  a. John is six feet tall.
     b. tall(j,6ft)

The predicate *tall* is treated here as an expression of type \(<d,<e,t>>\), a two-place relation between entities and degrees. The measure phrase is an expression of type \(d\), denoting a degree. The interpretation of (25a) can then be represented by the formula in (25b). Now consider the comparative (26a). This sentence is true iff John is tall to a degree which is greater than six feet. These truth conditions are captured in (26b):

(26)  a. John is taller than six feet.
     b. \(\exists d [\text{tall}(j,d) \land d > 6\text{ft}]\)

The comparative form of the adjective *taller* has the same type as the positive form *tall*, namely \(<d,<e,t>>\). Both *tall* and *taller* can take a measure phrase like *six feet* as their argument, but they differ in the syntactic realization of this argument. *Tall* takes the measure phrase on its left, while *taller* takes it on its right. The affix *-er* takes an expression of type \(<d,<e,t>>\) (tall) and turns it into another expression of the same type (taller). *-er* itself must therefore be of type \(<<d,<e,t>>,d,<e,t>>\). The modifier *more* will be of the same type.
Now let’s return to sentential comparatives, such as (27a). Because, as argued above, the comparative clause (than) Bill is denotes a (maximal) degree and is therefore of type d just like the measure phrase six feet, (27a) has the same logical structure as (26a), and can be given the translation in (27b):

(27) a. John is taller than Bill is.
   b. \( \exists d [\text{tall}(j,d) \land d > \text{max}(\lambda d'[\text{tall}(b,d')])] \)

*John is taller than Bill is* is true iff John stands in the tall-relation to a degree that is greater than the maximal degree to which Bill is tall.

Summarizing, the assignment of types is as follows:

(28) a. *tall*: type \( <d,<e,t>> \)
   b. *six feet*: type \( d \)
   c. *six feet tall*: type \( <e,t> \)
   d. *taller*: type \( <d,<e,t>> \)
   e. *taller than six feet*: type \( <e,t> \)
   f. *than Bill is*: type \( d \)
   g. *taller than Bill is*: type \( <e,t> \)
(29) shows how the interpretation of the AP taller than Bill is (tall), whose syntactic structure is given in (a), is built up compositionally (the arrow $\rightarrow$ stands for ‘translates as’):

(29) a. $[_{\text{AP}} [_{\lambda} \text{tall} [_{\lambda} -\text{er}]] \text{ than } [_{\text{CP}} Op, [_{\text{IP}} \text{Bill is } d_{r}\text{-tall}}]]$

b. $[_{\text{CP}} Op, [_{\text{IP}} \text{Bill is } d_{r}\text{-tall}}] \rightarrow \max(\lambda d'[\text{tall}(b,d')])$

c. $\text{tall} \rightarrow \lambda d\lambda x[\text{tall}(x,d)]$

d. $-\text{er} \rightarrow \lambda R\lambda d'\lambda x[\exists d[R(d)(x) \land d > d']]$

e. $\text{taller} \rightarrow \lambda d'\lambda x[\exists d[\text{tall}(x,d) \land d > d']]$

f. $\text{taller than Bill is} \rightarrow \lambda x\exists d[\text{tall}(x,d) \land d > \max(\lambda d'[\text{tall}(b,d'))]]$

Here $R$ is a variable of type $<d,<e,t>>$, the type of degree adjectives. Note that the existential quantifier over degrees is built into the meaning of the affix -er. This is important because we don’t want the existential quantifier to have scope over the subject. (30a) has the interpretation given in (b) and not that in (c) (the latter reading would imply that every girl has the same height, which is not what we want):

(30) a. Every girl is taller than Bill is.

b. $\forall x[\text{girl}(x) \rightarrow \exists d[\text{tall}(x,d) \land d > \max(\lambda d'[\text{tall}(b,d'))]]$

c. $\exists d\forall x[[\text{girl}(x) \rightarrow \text{tall}(x,d)] \land d > \max(\lambda d'[\text{tall}(b,d'))]]$

I will now turn to the explanation of the negative island effect in comparatives.
2.3 Explaining the Negative Island Effect in Comparatives

Von Stechow’s maximality requirement offers a nice explanation for the negative island effect in comparative clauses. The relevant examples are repeated here from section 2.1:

**Upward Entailing Contexts**

(31) a. John weighs more than Bill weighs.
    b. John weighs more than everybody else weighs.
    c. John weighs more than most people weigh.
    d. John weighs more than many people weigh.
    e. John weighs more than at least five people weigh.
    f. John weighs more than Bill always/often weighed.

**Downward Entailing Contexts**

(32) a. * John weighs more than Bill doesn’t weigh.
    b. * John weighs more than nobody weighs.
    c. * John weighs more than few people weigh.
    d. * John weighs more than fewer than five people weigh.
    e. * John weighs more than at most five people weigh.
    f. * John weighs more than Bill never/seldom weighed.

60
According to von Stechow, the reason why sentences like the ones in (32) are ungrammatical is that \textit{than}-clauses refer to \textit{maximal} degrees, and that maximal degrees are not defined in downward entailing contexts. Let’s first concentrate on (32a), repeated as (34a), and contrast it with the grammatical (33). These sentences have the interpretations indicated in (33b) and (34b):

\begin{align*}
\text{(33a)} & \quad \text{John weighs more than Bill weighs.} \\
\text{(33b)} & \quad \text{John weighs more than } \max(\lambda d[\text{Bill weighs } d\text{-much}]).
\end{align*}

\begin{align*}
\text{(34a)} & \quad * \text{ John weighs more than Bill doesn’t weigh.} \\
\text{(34b)} & \quad \text{John weighs more than } \max(\lambda d[\text{Bill doesn’t weigh } d\text{-much}]).
\end{align*}

Recall that we are assuming the ‘exactly’ interpretation for gradable adjectives, under which \textit{Bill weighs 150 pounds} is true iff Bill weighs \textit{exactly} 150 pounds. Suppose that Bill does in fact weigh exactly 150 pounds. Then the set denoted by \(\lambda d[\text{Bill weighs } d\text{-much}]\) will be the singleton set containing 150 pounds as its only member. The \textit{max}-operator then trivially picks out this single weight. Note that for this particular example the iota-operator would have achieved exactly the same result.\footnote{If we were assuming that ‘weigh \(d\text{-much}\)’ means ‘weigh at least \(d\text{-much}\)’, we would also get the right result for (33). Suppose that Bill weighs exactly 150 pounds. Then on the at least interpretation, \(\lambda d[\text{Bill weighs } d\text{-much}]\) would denote the set of weights up to and including 150 pounds. The maximal member of that set is 150 pounds. (33) would therefore be interpreted as saying that John weighs more than 150 pounds, which is the desired result.}
In (34) there is a problem however. Again on the assumption that Bill weighs exactly 150 pounds, \( \lambda d[\text{Bill doesn’t weigh } d\text{-much}] \) denotes the set of all degrees of weight except 150 pounds, and it is clear that this set does not have a maximal member. Thus, \( \max(\lambda d[\text{Bill doesn’t weigh } d\text{-much}]) \) is undefined. This explains the ungrammaticality of (34b).⁹

Von Stechow’s explanation carries over to all cases in which the comparative clause contains a downward entailing element. The reason is basically that the effect of a downward entailing operator is to take the complement of a given set, as is discussed at length by Szabolcsi and Zwarts (1993). For instance, the set denoted by (35a) is the complement of the set denoted by (35b), and similarly for the pairs in (36)-(38):

(35) a. \( \lambda d[\text{Bill doesn’t weigh } d\text{-much}] \)  
   b. \( \lambda d[\text{Bill weighs } d\text{-much}] \)

(36) a. \( \lambda d[\text{Nobody weighs } d\text{-much}] \)  
   b. \( \lambda d[\text{Somebody weighs } d\text{-much}] \)

⁹ The explanation would also work if we had assumed the ‘at least’ interpretation of ‘weigh \( d\text{-much} \)’ rather than the ‘exactly’ interpretation. Under that interpretation \( \lambda d[\text{Bill doesn’t weigh } d\text{-much}] \) would denote the set of all weights greater than 150 pounds. This set also lacks a maximum.
(37)  a. $\lambda d[\text{Fewer than ten people weigh } d\text{-much}]$
    b. $\lambda d[\text{At least ten people weigh } d\text{-much}]$

(38)  a. $\lambda d[\text{Bill never weighed } d\text{-much}]$
    b. $\lambda d[\text{At some point Bill weighed } d\text{-much}]$

Now since the sets in (b) are all finite and contain a maximum, their complements in (a) will be infinite and do not have a maximum. Take for instance the set denoted by (36a). This set contains all the degrees of weight $d$ such that nobody weighs $d$-much and this set obviously does not have a maximum member.

We conclude then that in downward entailing contexts maximal degrees are not defined, while in upward entailing contexts they are. This explains why comparative clauses can contain upward entailing elements, but not downward entailing ones. Of course, this explanation of the negative island effect is only valid to the extent that von Stechow’s claim that comparative clauses denote maximal degrees can be shown to provide an adequate analysis of the semantics of comparatives. In the rest of this chapter I will show that this is indeed the case, first by reviewing von Stechow’s own arguments pertaining to comparatives like taller than and then by showing how the maximality account can be extended to comparatives such as less tall than and as tall as, resulting in insightful analyses of the ambiguity of these constructions. For all types of comparatives, monotonicity properties will provide crucial evidence.
2.4 Inferential Properties of the Comparative

2.4.1 Downward Entailment

It is a well-known fact that comparatives can license negative polarity items (Ross (1969), Seuren (1973), Hoeksema (1983), von Stechow (1984a)):

(39) a. Seymour makes more money than any other student does.
    b. Sarah was taller than anybody had expected.
    c. Belinda is much richer than I will ever be.
    d. He told me more jokes than I cared to write down.
    e. Floyd broke more glasses than most people could stand.
    f. He said the sky would sooner fall than he would budge an inch.

As Ladusaw has shown, negative polarity items are only licensed by expressions that are downward entailing. An important criterion any adequate semantics for the comparative must therefore meet is that it explains the downward entailing character of the construction. In this section I will show (following von Stechow) that the analysis of comparatives in terms of maximality does indeed predict downward entailment.
Recall the definition of a downward entailing function:

\[(40)\quad \textbf{Definition of a Downward Entailing Function:}\]

A function \(f\) is downward entailing (or monotone decreasing) iff for all \(X, Y\) in the domain of \(f\): if \(X \subseteq Y\), then \(f(Y) \subseteq f(X)\).

This definition gives us a straightforward empirical test for determining whether a given linguistic environment \(V\ldots W\) is downward entailing. We just check whether inferences of the following form are valid:

\[(41)\quad V A W \Rightarrow V B W\]

where \(A\) and \(B\) are two expressions such that \([B] \subseteq [A]\). The following valid inferences show that the comparative is indeed downward entailing:

\[(42)\quad \begin{align*}
\text{a.} & \quad \text{Seymour is richer than a student can be} \Rightarrow \\
& \quad \text{Seymour is richer than a foreign student can be} \\
\text{b.} & \quad \text{George runs more often than Bill swims} \Rightarrow \\
& \quad \text{George runs more often than Bill swims in the ocean}
\end{align*}\]

There is another characterization of downward entailing functions which is equivalent to (40) (Zwarts 1986):
Below it will be shown that these implications can actually be strengthened to equivalences.

Alternative Definition of a Downward Entailing Function:

A function $f$ is downward entailing (or monotone decreasing) iff for all $X, Y$ in the domain of $f$:

$$f(X \cup Y) \subseteq f(X) \cap f(Y).$$

This means that we can show that a linguistic context $V...W$ is downward entailing by showing that inferences of the following form are valid:

$$V (A \text{ or } B) W \Rightarrow V A W \text{ and } V B W$$

The following examples demonstrate that such inferences are indeed supported in comparatives:

(45)  

a. Carla is smarter than Julia or Susan is $\Rightarrow$
    Carla is smarter than Julia is and Carla is smarter than Susan is.

b. Belinda runs more often than Sandra swims or cycles $\Rightarrow$
    Belinda runs more often than Sandra runs and Belinda runs more often than Sandra cycles.

c. Andrea is richer than Bob knows or wants to admit $\Rightarrow$
    Andrea is richer than Bob knows and Andrea is richer than Bob wants to admit.

---

10 Below it will be shown that these implications can actually be strengthened to equivalences.
The fact that more than-comparatives are downward entailing is correctly predicted under the semantics for the comparative outlined in the preceding sections. To show why, let’s first look at a concrete example, such as (42a). The logical structure of Seymour is richer than a student can be is given informally in (46a) and that of Seymour is richer than a foreign student can be in (47a); the translations of the two sentences are (46b) and (47b):

\[(46) \begin{align*}
\text{a. Seymour is richer than } & \text{max}(\lambda d[\text{a student can be } d\text{-rich}]). \\
\text{b. } & \exists d[\text{rich}(w)(s,d) \land d>\text{max}(\lambda d'[\exists w' \exists x[\text{student}(w')(x) \land \text{rich}(w')(x,d')]])]
\end{align*} \]

\[(47) \begin{align*}
\text{a. Seymour is richer than } & \text{max}(\lambda d[\text{a foreign student can be } d\text{-rich}]). \\
\text{b. } & \exists d[\text{rich}(w)(s,d) \land d>\text{max}(\lambda d'[\exists w' \exists x[\text{student}(w')(x) \land \text{foreign}(w')(x) \\
\hspace{1cm} \land \text{rich}(w')(x,d')]])]
\end{align*} \]

Note that here I am switching to an intensional language with explicit quantification over possible world variables. The modal can is rendered as an existential quantifier over possible worlds, ignoring the role played by the accessibility relation between worlds.

Because the set of foreign students is a subset of the set of students, the set of degrees denoted by $\lambda d[\text{a foreign student can be } d\text{-rich}]$ will be a subset of the set denoted by $\lambda d[\text{a student can be } d\text{-rich}]$. Therefore, $\text{max}(\lambda d[\text{a foreign student can be } d\text{-rich}])$
rich]) will be smaller than or equal to \( \text{max}(\lambda d[\text{a student can be } d\text{-rich}]) \), which in turn implies that if Seymour is richer than \( \text{max}(\lambda d[\text{a student can be } d\text{-rich}]) \), he will also be richer than \( \text{max}(\lambda d[\text{a foreign student can be } d\text{-rich}]) \).

Shifting the attention to the general case, consider an adjective phrase of the form (48a), where \( \text{Adj} \) is a gradable adjective and \( CP \) is a comparative clause. (48a) denotes the set given in (48b), where \( D \) is the set of degrees denoted by \( CP \).

(48) a. \([_{\text{AP}} \text{ more } \text{Adj} \text{ than } CP]\)

b. \(\{x | x \text{ is more } \text{Adj} \text{ than } \text{max}(D)\}\)

Now it is easy to see that if \( D \) is expanded, then the maximum degree \( \text{max}(D) \) will go up or stay the same, and consequently, the set of objects that are more \( \text{Adj} \) than \( \text{max}(D) \) will either shrink or stay the same. In other words, the following implication holds:

(49) If \( D \subseteq D' \), then
\[
\{x | x \text{ is more } \text{Adj} \text{ than } \text{max}(D')\} \subseteq \{x | x \text{ is more } \text{Adj} \text{ than } \text{max}(D)\}.
\]

This means that \textit{more than}-comparatives are downward entailing.
2.4.2 Anti-Additivity

Hoeksema (1983) has shown that (clausal) comparatives have an even stronger property than that of being downward entailing, namely that of being anti-additive in the sense of Zwarts (1981; 1986).\textsuperscript{11} It can be demonstrated that von Stechow’s semantics also correctly predicts this property.

Anti-additive functions are defined in (50):

\begin{equation}
\text{(50) Definition of an Anti-Additive Function:}
\end{equation}

A function \( f \) is anti-additive iff

\[
\text{for all } X, Y \text{ in the domain of } f: f(X \cup Y) = f(X) \cap f(Y).
\]

The test an expression for anti-additivity we have to determine whether equivalences of the following form hold:

\begin{equation}
V A \text{ or } B W \iff V A W \text{ and } V B W
\end{equation}

By comparing (50) and (51) to (43) and (44), it can be seen that anti-additivity amounts to a strengthening of downward entailment. Whereas a downward entailing expression

\textsuperscript{11} Anti-additivity is also discussed by Cresswell (1976) and von Stechow (1984), although they didn’t use that term.
only supports an implication in one direction (\(V A or B W \Rightarrow V A W \text{ and } V B W\)), with an anti-additive expression the implication is bidirectional (\(V A or B W \Leftarrow V A W \text{ and } V B W\)). Hence, the downward entailing expressions are a proper subset of the anti-additive expressions. Not and nobody, for instance, are anti-additive, and therefore also downward entailing; not everybody and at most five people are downward entailing, but not anti-additive:

\[
\begin{align*}
(52) \quad & a. \quad \text{Abe did not sing or dance } \Leftarrow \text{Abe did not sing and Abe did not dance.} \\
& b. \quad \text{Nobody sang or danced } \Leftarrow \text{Nobody sang and nobody danced.} \\
& c. \quad \text{Not everybody sang or danced } \Rightarrow \\
& \quad \text{Not everybody sang and not everybody danced.} \\
& c'. \quad \text{Not everybody sang and not everybody danced } \not\Rightarrow \\
& \quad \text{Not everybody sang or danced.} \\
& d. \quad \text{At most five people sang or danced } \Rightarrow \\
& \quad \text{At most five people sang and at most five people danced.} \\
& d'. \quad \text{At most five people sang and at most five people danced } \not\Rightarrow \\
& \quad \text{At most five people sang or danced.}
\end{align*}
\]

Using patterns like this as a test, we see that the comparative construction is clearly anti-additive:\(^\text{12}\)

---

\(^\text{12}\) These sentences also have another reading which gives the disjunction wide scope. Under this reading, Carla is smarter than Julia or Susan is is equivalent to Carla is smarter than Julia is or Carla is smarter than Susan is. One way of deriving this
reading is by giving the NP Julia or Susan wide scope by means of quantifier raising or quantifying in. (For more discussion of wide scope readings of NP’s in comparative clauses, see section 2.8.1.) However, I believe there is reason to prefer an alternative analysis under which the disjunction is treated as the ‘wide scope or’ of Rooth and Partee (1982). Note that the intended reading of the sentence can be paraphrased as Carla is smarter than Julia or Susan is, but I don’t know which, one of the tests discussed by Rooth and Partee. Further evidence for the ‘wide scope or’ analysis and against the quantifier raising analysis comes from the fact that a wide scope reading of the disjunction is possible for a sentence like John weighs more than a welterweight or a bantamweight (but I don’t know which), without giving the indefinite noun phrases a welterweight and a bantamweight wide scope.

The anti-additivity of comparatives is predicted under von Stechow’s account. Consider (53a), for instance. The logical structure of Carla is smarter than Julia or Susan is is given in (54a) and that Carla is smarter than Julia and Carla is smarter than Susan in (54b):

(53) a. Carla is smarter than Julia or Susan is ⇔
    Carla is smarter than Julia is and Carla is smarter than Susan is.

b. Belinda runs more often than Sandra swims or cycles ⇔
    Belinda runs more often than Sandra runs and Belinda runs more often than Sandra cycles.

c. Andrea is richer than Bob knows or wants to admit ⇔
    Andrea is richer than Bob knows and Andrea is richer than Bob wants to admit.

(54a) Carla is smarter than Julia and Carla is smarter than Susan

(54b) Andrea is richer than Bob knows and Andrea is richer than Bob wants to admit.
(54)  

a. Carla is smarter than \(\text{max}(\lambda d[\text{Julia or Susan is } d\text{-smart}])\).

b. Carla is smarter than \(\text{max}(\lambda d[\text{Julia is } d\text{-smart}])\) and Carla is smarter than \(\text{max}(\lambda d[\text{Susan is } d\text{-smart}])\).

If the degree to which Carla is smart is greater than the maximal degree \(d\) such that Julia or Susan is \(d\)-smart, then Carla’s degree of smartness must exceed both Julia’s degree of smartness and Susan’s degree of smartness. Hence, (54a) implies (54b).

Conversely, if the degree to which Carla is smart is greater than the degree to which Julia is smart and greater than the degree to which Susan is smart, then Carla’s degree of smartness must be greater than the maximal degree \(d\) such that Julia or Susan is \(d\)-smart. Therefore, (54b) implies (54a). The equivalence can also be explained in somewhat plainer English: Suppose that Julia is smarter than Susan. Then saying that Carla is smarter than both Julia and Susan is truth-conditionally equivalent to saying that Carla is smarter than Julia. If, on the other hand, Julia is less smart than Susan, then the claim that Carla is smarter than both Julia and Susan will be true iff Carla is smarter than Susan. In other words, the proposition that Carla is smarter than both Julia and Susan is equivalent to the proposition that Carla is smarter than whoever is the smartest of the two other women.
More generally, the following holds:

\[(55) \quad \{x \mid x \text{ is more } Adj \text{ than } \max(D \cup D')\} = \{x \mid x \text{ is more } Adj \text{ than } \max(D)\} \cap \{x \mid x \text{ is more } Adj \text{ than } \max(D')\}\]

Hoeksema (1983) gives an interesting additional empirical argument for the anti-additivity of the comparative construction. Zwarts (1981; 1986) has shown that there are negative polarity items for which being in the scope of a downward entailing function is not enough and which need to be licensed by an anti-additive function. An example is the Dutch negative polarity item *ook maar* (hard to translate, but roughly equivalent to English ‘at all’), which can be licensed by anti-additive expressions like *niet* (‘not’) and *niemand* (‘nobody’), but not by negative polarity items which are not anti-additive, such as *niet iedereen* (‘not everybody’) and *ten hoogste vijf mensen* (‘at most five people’):

\[(56) \quad \begin{align*}
a. & \quad \text{Ik denk niet dat hij ook maar iets gedaan heeft.} \\
& \quad \text{I think not that he at-all anything done has.} \\
& \quad \text{‘I don’t think he did anything at all.’} \\

b. & \quad \text{Niemand heeft ook maar iets gedaan.} \\
& \quad \text{Nobody has at-all anything done.} \\
& \quad \text{‘Nobody did anything at all.’}
\end{align*}\]
(56)  c.  Niet iedereen heeft ook maar iets gedaan.
   Not everybody has at-all anything done.
   ‘Not everybody has done anything at all.’

d.  * Ten hoogste vijf mensen hebben ook maar iets gedaan.
   At most five people have at-all anything done.
   ‘At most five people have done anything at all.’

The distribution of ook maar can be used as a test for anti-additivity. Hoeksema shows that ook maar does indeed occur in comparatives, as in the following examples (adapted from Hoeksema’s (13a-c)):

(57)  a.  Het feest duurde langer dan ook maar iemand verwacht had.
   The party lasted longer than at-all anybody expected had.
   ‘The party lasted longer than anybody at all had expected.’
b.  Ik wacht liever een uur dan ook maar een minuut te laat te zijn.
   I wait rather an hour than at-all a minute too late to be.
   ‘I would rather wait for an hour than be even a minute late.’
c.  Wim was vervelender dan ook maar iemand voor hem was geweest.
   Wim was more-obnoxious than at-all anybody before him had been.
   ‘Wim was more obnoxious than anybody at all before him had been.’
2.5 Less Than-Comparatives

So far I have only discussed one kind of comparative clauses, namely more than comparatives. In this section, I will show how von Stechow’s analysis can be extended to less than-comparatives. Interestingly, this is not just a routine exercise in semantics, but it brings to light an unexpected ambiguity which is lacking in more than-comparatives. As with more than-comparatives, inferential properties like upward and downward monotonicity and anti-additivity, as well as the distribution of negative polarity items, will take center stage as diagnostics.

2.5.1 Maximality in Less Than-Comparatives

The arguments for maximality that were given in sections 2.2 and 2.3 concerning more than-comparatives carry over to comparatives with less like (58):

(58) John is less tall than Bill is.

Maximality is needed in this construction just as much as it is in the taller than construction. Consider for instance the following example:

(59) John is less strong than he could be.
Here John’s actual degree of strength is compared to the maximal degree of strength that he could attain. The sentence is true iff John’s actual strength is smaller than his maximal potential strength.

Less than-comparatives also show the same negative island effects that the more than-comparatives do. This is demonstrated by the following set of data:

**Upward Entailing Contexts**

(60) a. John weighs less than Bill weighs.
    b. John weighs less than everybody else weighs.
    c. John weighs less than most people weigh.
    d. John weighs less than many people weigh.
    e. John weighs less than at least five people weigh.
    f. John weighs less than Bill always/often weighed.

**Downward Entailing Contexts**

(61) a. * John weighs less than Bill doesn’t weigh.
    b. * John weighs less than nobody weighs.
    c. * John weighs less than few people weigh.
    d. * John weighs less than less than five people weigh.
    e. * John weighs less than at most five people weigh.
    f. * John weighs less than Bill never/often weighed.
The sentences in (61) are out for exactly the same reason that the corresponding more than-comparatives in (32) are, namely a failure of the maximality requirement that is built into the semantics of the comparative clause.

Given these data, the most straightforward way of spelling out the meaning of (59) is as follows:

(62) John is less strong than max(\lambda d[he could be d-strong]).

On this analysis the logical translation assigned to less than-comparatives differs only minimally from the translation that was assigned to more than-comparatives in section 2.2, the only difference being that the >-sign is changed to a <-sign. (63a), for instance, will get the translation given in (63b).\(^{13}\) (64) shows how this translation can be derived compositionally:

(63) a. John is less tall than Bill is.

b. \(\exists d[tall(j,d) \land d < max(\lambda d'[tall(b,d')])]

\(^{13}\) It is important to note that this analysis will only work on the ‘exactly’ interpretation of tall(j,d). The ‘at least’ interpretation would predict that John is less tall than Bill is would be true iff there is a degree d such that Bill is taller than d and John is at least as tall as d. This would mean that the sentence could even be true if John is in fact taller than Bill, an entailment which is clearly unwanted. This is one of the main reasons why I chose to adopt the ‘exactly’ interpretation rather than the ‘at least’ interpretation (cf. section 2.1.4).
(64) a. \([A_P [\text{less tall} \text{ than } [C_P Op, [\text{IP Bill is } d-\text{tall}]])]\) 

b. \([C_P Op, [\text{IP Bill is } d-\text{tall}] \rightarrow max(\lambda d'[\text{tall}(b,d')])]\) 

c. tall \rightarrow \lambda d\lambda x[\text{tall}(x,d)] 

d. less \rightarrow \lambda R\lambda d'\lambda x[\exists d[R(d)(x) \land d < d']]) 

e. less tall \rightarrow \lambda d'\lambda x[\exists d[\text{tall}(x,d) \land d < d']]) 

f. less tall than Bill is \rightarrow \lambda x\exists d[\text{tall}(x,d) \land d < max(\lambda d'[\text{tall}(b,d')])] 

2.5.2 Less Than-Comparatives and Monotonicity

On the interpretation proposed for less than-comparatives in the preceding section we predict that they are not downward entailing (let alone anti-additive). This can be illustrated with the following (invalid) inferences:

(65) a. Seymour is less rich than \(max(\lambda d[\text{a student can be } d-\text{rich}])\) \(\not\rightarrow\) Seymour is less rich than \(max(\lambda d[\text{a foreign student can be}])\). 

b. George runs less often than \(max(\lambda d[\text{Bill swims } d-\text{often}])\) \(\not\rightarrow\) George runs less often than \(max(\lambda d[\text{Bill swims in the ocean } d-\text{often}])\).

These examples lead to the conclusion that less than-comparatives are not downward entailing, and that therefore they are not anti-additive either (recall that the anti-additive functions are a proper subset of the downward entailing functions). This is as expected
In the literature, very little attention has been paid to the monotonicity properties of less than-comparatives, as opposed to those of more than-comparatives. Hendriks (1993 and forthcoming) also reaches the conclusion that less than-comparatives are upward entailing on the basis of the (in)validity of the following inferences:

(i) Fewer students danced than teachers sang a ballad ⇒ Fewer students danced than teachers sang.
(ii) Fewer students danced than teachers sang ≠ Fewer students danced than teachers sang a ballad.

under the interpretation for less than-comparatives proposed above. In fact, under this interpretation less than-comparatives are upward entailing. To see why this is the case, consider (66a) and its denotation (66b) under the proposed interpretation:

(66)  
   a. \([\text{AP} \text{ less Adj than CP}]\)  
   b. \(\{x | x \text{ is less Adj than } \max(D)\}\)

If the set of degrees D is expanded, then \(\max(D)\) increases or stays the same, therefore the set of entities that are less Adj than \(\max(D)\) becomes larger or stays the same. Thus we have the following implication:

(67)  
   If \(D \subseteq D'\), then \(\{x | x \text{ is less Adj than } \max(D)\} \subseteq \{x | x \text{ is less Adj than } \max(D')\}\).

Because, under the proposed interpretation, less than-comparatives are not downward entailing, we would expect them not to license negative polarity items. However, the following examples show that this expectation is not borne out:

\[\text{In the literature, very little attention has been paid to the monotonicity properties of less than-comparatives, as opposed to those of more than-comparatives. Hendriks (1993 and forthcoming) also reaches the conclusion that less than-comparatives are upward entailing on the basis of the (in)validity of the following inferences:}\
\[\begin{align*}
    & \text{(i)} \quad \text{Fewer students danced than teachers sang a ballad } \Rightarrow \text{ Fewer students danced than teachers sang.} \\
    & \text{(ii)} \quad \text{Fewer students danced than teachers sang } \not\Rightarrow \text{ Fewer students danced than teachers sang a ballad.}
\end{align*}\]
Almost no discussion can be found in the literature concerning the occurrence of negative polarity items in less than-comparatives. An interesting observation is made by Mittwoch (1974), however, who gives data showing that while ever can be licensed by less than-comparatives, modal need (also a negative polarity item) can’t:

(i) a. John was more helpful than Mary ever was.
   b. John was less helpful than Mary ever was.

(ii) a. John was more helpful than he need have been.
   b. * John was less helpful than he need have been.

In Dutch, we find exactly the same contrast between ook maar iemand (‘anybody at all’) and hoeven (‘need’):

(iii) a. Jan was meer behulpzaam dan ook maar iemand voor hem geweest was.
   b. Jan was minder behulpzaam dan ook maar iemand voor hem geweest was.

(iv) a. Jan was meer behulpzaam dan hij hoefde te zijn.
   b. * Jan was minder behulpzaam dan hij hoefde te zijn.

In light of the foregoing discussion, I suggest that the ungrammaticality of (iib) and (ivb) is due to the unavailability of the less-than-minimum interpretation in these examples. That this reading is indeed not available can be shown by replacing need with the polarity neutral should (and hoeven with moeten ‘must’):

(v) John was less helpful than he should have been.

(vi) Jan was minder behulpzaam dan hij moest zijn.

These sentences clearly only have the less-than-maximum reading, which does not license negative polarity items. Exactly why the less-than-minimum reading is not available is not clear to me; it may have something to do with the fact that need and should (as well as hoeven and moeten) are universal quantifiers (over possible worlds), whereas ever and ook maar iemand are existentials.

Surprisingly, these sentences are good. We are faced with a paradoxical situation. We have good arguments to assign less than-comparatives an interpretation which predicts that they are not downward entailing and should therefore not be able to license negative polarity items. Nevertheless, we find examples in which negative polarity items do in fact occur in less than-comparatives.¹⁵

¹⁵ Almost no discussion can be found in the literature concerning the occurrence of negative polarity items in less than-comparatives. An interesting observation is made by Mittwoch (1974), however, who gives data showing that while ever can be licensed by less than-comparatives, modal need (also a negative polarity item) can’t:

(i) a. John was more helpful than Mary ever was.
   b. John was less helpful than Mary ever was.

(ii) a. John was more helpful than he need have been.
   b. * John was less helpful than he need have been.

In Dutch, we find exactly the same contrast between ook maar iemand (‘anybody at all’) and hoeven (‘need’):

(iii) a. Jan was meer behulpzaam dan ook maar iemand voor hem geweest was.
   b. Jan was minder behulpzaam dan ook maar iemand voor hem geweest was.

(iv) a. Jan was meer behulpzaam dan hij hoefde te zijn.
   b. * Jan was minder behulpzaam dan hij hoefde te zijn.

In light of the foregoing discussion, I suggest that the ungrammaticality of (iib) and (ivb) is due to the unavailability of the less-than-minimum interpretation in these examples. That this reading is indeed not available can be shown by replacing need with the polarity neutral should (and hoeven with moeten ‘must’):

(v) John was less helpful than he should have been.

(vi) Jan was minder behulpzaam dan hij moest zijn.

These sentences clearly only have the less-than-maximum reading, which does not license negative polarity items. Exactly why the less-than-minimum reading is not available is not clear to me; it may have something to do with the fact that need and should (as well as hoeven and moeten) are universal quantifiers (over possible worlds), whereas ever and ook maar iemand are existentials.
Closely related to the problem of negative polarity items in these examples is another problem concerning the disjunction or. Above I showed that under the proposed interpretation less than-comparatives are not downward entailing. Now recall that the downward entailing contexts are a superset of the anti-additive contexts, which leads us to expect that less than-comparatives will not be anti-additive either. However, equivalences like the following do seem to be intuitively valid:

(69)  a. Carla is less smart than Julia or Susan is ⇔
      Carla is less smart than Julia is and Carla is less smart than Susan is.

     b. Belinda runs less often than Sandra swims or cycles ⇔
        Belinda runs less often than Sandra swims and Belinda runs less often
        than Sandra cycles.

     c. Andrea is less rich than Bob knows or wants to admit ⇔
        Andrea is less rich than Bob knows and Andrea is less rich than Bob
        wants to admit.

It is also worth noting that the Dutch negative polarity item ook maar, which only occurs in anti-additive contexts, can show up in less than-comparatives:

(70)  a. Het feest duurde minder lang dan ook maar iemand verwacht had.
      The party lasted less long than at-all anybody expected had.
      ‘The party lasted less long than anybody at all had expected.’
(70)  b. Wim was minder vervelend dan ook maar iemand voor hem was geweest.

Wim was less obnoxious than at-all anybody before him had been.

‘Wim was less obnoxious than anybody at all before him had been.’

So again we have a paradox here. On the basis of the interpretation of less than-comparatives that was argued for, we would expect that they are not anti-additive, but actual entailment patterns and the distribution of polarity items both seem to show that they are. In the next section, I will argue that the paradox can be resolved once we recognize that less than-comparatives are systematically ambiguous, although in many instances only one of the two readings is actually available.

2.5.3 The Ambiguity of Less Than-Comparatives

The problem with the examples in (68) is not only that negative polarity items occur in a context where they are not supposed to, but also that these sentences do not mean what we would expect them to mean based on the interpretation proposed in section 2.5.1. Semantically, negative polarity items like any and ever are existential quantifiers. We would therefore expect (68a) to mean that Seymour makes less money than the maximal amount of money \( d \) such that some other student makes \( d \)-much. These truth conditions are very weak, of course. According to this interpretation the
sentence would be true if Seymour does not have the highest income of all the students. But the actual truth conditions of the sentence are much stronger. What (68) says is that Seymour is the student with the lowest income. Similar remarks apply to the other examples. So the problem with (68) is not simply one of the distribution of negative polarity items. Rather, it has to do with what these sentences mean.

I want to argue that *less than*-comparatives actually can have two readings. In addition to the reading argued for in section 2.5.1, which I will refer to as the ‘less-than-maximum’ reading, there is a second reading which can be called the ‘less-than-minimum’ reading. In many cases only one of the two readings is actually available, which is why the ambiguity is not immediately detected. There are instances of *less than*-comparatives in which the ambiguity does show up. The following sentence has both the less-than-maximum and the less-than-minimum interpretation:

(71) The helicopter was flying less high than a plane can fly.

On the less-than-maximum reading the sentence is true iff the altitude at which the helicopter was flying was below the maximal altitude at which a plane can fly. This particular reading is brought out in the following context:

---

16 Shortly before submitting this dissertation I discovered that this ambiguity was already briefly discussed by Seuren (1979).
Because the helicopter was flying less high than a plane can fly, the jet fighter could easily fire at it from above.

According to the less-than-minimum interpretation of (71), on the other hand, the helicopter was flying at an altitude below the minimal altitude at which a plane can fly. This reading is made salient in the context (73):

The jet fighter was trying to chase the helicopter, but because the helicopter was flying less high than a plane can fly, the helicopter crashed into a building.

Another example exhibiting the same ambiguity is the following:

In their first semester, the students have to take at least three and at most six classes for credit. Harry took fewer classes than was allowed.

This sentence can mean either that Harry took fewer than six classes (less-than-maximum), or that he took fewer than three (less-than-minimum).

The less-than-maximum reading of (71) is captured in (75a). Provisionally, we can represent the less-than-minimum reading as in (75b), using a ‘minimality’ operator min analogous to the maximality operator max:
(75)  a. The helicopter was flying less high than $\max(\lambda d[\text{a plane can fly } d\text{-high}])$.
   b. The helicopter was flying less high than $\min(\lambda d[\text{a plane can fly } d\text{-high}])$.

An obvious objection against the use of $\min$ is that it is completely ad hoc. (75b) captures the truth conditions of one of the two readings of (71), but it does not explain why this sentence is ambiguous, and it also fails to explain why the corresponding *more than*-comparative (*The helicopter was flying higher than a plane can fly*) is not ambiguous in the same way. The latter sentence cannot mean that the helicopter was flying higher than the minimal altitude at which a plane can fly. In the next subsection, I will try to give a more principled account of the ambiguity of *less than*-comparatives, but for the moment I will stick to (75) in order to explore the monotonicity properties of the two readings of *less than*-comparatives.

As we have seen, the less-than-maximum interpretation is not downward entailing. If the helicopter was flying at an altitude below the maximal altitude at which a plane can fly, then it wasn’t necessarily flying below the maximal altitude at which a propeller plane could fly (perhaps the maximal altitude of propeller planes is lower than that of jet planes). The less-than-minimum reading, however, is downward entailing. Because the set of propeller planes is a subset of the set of (all) planes, the minimal altitude at which a propeller plane can fly is greater than or equal to the minimal altitude at which an arbitrary plane can fly. Thus, if the helicopter was flying at an altitude below the minimal altitude at which a plane can fly, then it was also flying below the
minimal altitude at which a propeller plane can fly. Thus while the inference in (76) is not valid, the inference in (77) is:

(76) The helicopter was flying less high than $\text{max}(\lambda d[\text{a plane can fly } d\text{-high}]) \not\Rightarrow$
The helicopter was flying less high than $\text{max}(\lambda d[\text{a propeller plane can fly } d\text{-high}])$.

(77) The helicopter was flying less high than $\text{min}(\lambda d[\text{a plane can fly } d\text{-high}]) \rightarrow$
The helicopter was flying less high than $\text{min}(\lambda d[\text{a propeller plane can fly } d\text{-high}])$.

To see why in general the less-than-minimum interpretation of less than-comparatives is downward entailing, consider the set given in (78):

(78) $\{x | x \text{ is less } \text{Adj} \text{ than } \text{min}(D)\}$

If the set of degrees D is expanded, then its smallest element $\text{min}(D)$ will go down or stay the same, and thus the set $\{x | x \text{ is less } \text{Adj} \text{ than } \text{min}(D)\}$ will shrink or stay the same. In other words:
If \( D \subseteq D' \), then 
\[
\{ x \mid x \text{ is less } \text{Adj} \text{ than } \text{min}(D') \} \subseteq \{ x \mid x \text{ is less } \text{Adj} \text{ than } \text{min}(D) \}.
\]

Therefore, on the less-than-minimum interpretation, *less than*-comparatives are downward entailing.

Because *less than*-comparatives are only downward entailing on the less-than-minimum interpretation, we predict that negative polarity items will only be licensed on the less-than-minimum reading, and not on the less-than-maximum reading. That this prediction is borne out can be seen if we replace the indefinite article in (71) by *any*:

\[
\text{(80)} \quad \text{The helicopter was flying less high than any plane could fly.}
\]

In contrast to (71), this sentence can only have the less-than-minimum interpretation. The same is true for the examples we saw earlier of *less than*-comparatives in which a negative polarity item appears (see (68)). For instance, if Seymour makes less money than any other student does (cf. (68a)), then Seymour’s income is smaller than the minimum income of any student.

On the less-than-minimum reading, *less than*-comparatives are also anti-additive, as can be seen from the fact that (81a) is true iff (81b) is true:
a. The helicopter is flying less high than $\min(\lambda d[\text{a jet plane or a propeller plane can fly } d\text{-high}])$.

b. The helicopter is flying less high than $\min(\lambda d[\text{a jet plane can fly } d\text{-high}])$ and the helicopter is flying less high than $\min(\lambda d[\text{a propeller plane can fly } d\text{-high}])$.

More generally, the following holds:

\[
\{x \mid x \text{ is less } Adj \text{ than } \min(D \cup D')\} = \{x \mid x \text{ is less } Adj \text{ than } \min(D)\} \cap \{x \mid x \text{ is less } Adj \text{ than } \min(D')\}
\]

The anti-additivity of the less-than-minimum reading explains why the inferences in (69) are valid, and it also accounts for the fact that the Dutch negative polarity item ook maar iemand can appear in less than-comparatives (cf. (70)).

2.5.4 Why Aren’t Less Than-Comparatives Always Ambiguous?

If, as I claim, less than-comparatives can have two distinct readings, then why does this ambiguity not show up all the time, but only in rather special cases like (71) and (74)? The answer is that often one of the two potential readings is not available for independent reasons, while in other cases the two readings fall together. An example of
the latter possibility is the sentence *John is less tall than Bill is*. Since there is only one degree \( d \) such that Bill is \( d \)-tall, the less-than-maximum and the less-than-minimum reading fall together for this sentence. Cases in which one of the two readings is not available are more interesting. We have already seen one reason why this can happen. If a negative polarity item appears in the comparative clause, only the less-than-minimum interpretation is available because the less-than-maximum reading is not downward entailing and therefore does not license negative polarity items. But there is another reason why one of the two readings may not be available. In certain cases either a maximum or a minimum does not exist. Consider for instance the following two examples:

(83)  
\( \begin{align*} 
\text{a.} & \quad \text{Students spend less than a professor can spend.} \\
\text{b.} & \quad \text{Students live on less than a professor can live on.}
\end{align*} \)

Both sentences are unambiguous, but they have different interpretations. (83a) only has the less-than-maximum interpretation (‘Students spend less than the maximum amount a professor can spend’), whereas (83b) only has the less-than-minimum interpretation (‘Students can live on less than the minimum amount a professor can live on’). Why this contrast? I believe the explanation is really quite straightforward. The amount a professor can spend has a maximum, but no minimum (short of zero), but the amount a professor can live on has a minimum, but no maximum. It is no surprise therefore that (83a) cannot have the less-than-minimum reading and that (83b) lacks the less-than-
maximum reading. The sentence *The helicopter was flying less high than a plane can fly* is ambiguous because the altitude at which a plane can fly has both a maximum and a minimum; thus both readings are available.

To elaborate somewhat more on the difference between the three kinds of sentences (those that only have the less-than-maximum reading, those that only have the less-than-minimum reading, and those that are ambiguous), consider the following three open propositions:

(84)  
   a. a professor can spend *d*-much  
   b. a professor can live on *d*-much  
   c. a plane can fly *d*-high

Each of these open propositions determines a pragmatic scale in the sense of Fauconnier (1975a; 1975b; 1979). Importantly, the direction of the inferences they allow is different in each case. The scale determined by (84a) allows inferences from a higher point on the scale to a lower point, but not vice versa, as can be seen from the validity of the inference in (85a) and the invalidity of that in (85b):

(85)  
   a. A professor can spend $3000 ⇒ A professor can spend $2000.  
   b. A professor can spend $2000 ≠ A professor can spend $3000.
(84b) however only allows inferences in the other direction, that is, from a lower point on the scale to a higher one:

(86)  
   a.  A professor can live on $3000 ≠ A professor can live on $2000.
   b.  A professor can live on $2000 ⇒ A professor can live on $3000.

(84c) finally allows no entailments in either upward or downward direction, as can be seen from the fact that both of the following inferences are invalid:

(87)  
   a.  A plane can fly 1,000 feet high ≠ A plane can fly 500 feet high.
   b.  A plane can fly 30,000 feet high ≠ A plane can fly 40,000 feet high.

Summarizing, less than-comparatives are only ambiguous between a less-than-maximum and a less-than-minimum interpretation if they involve a pragmatic scale which has both a maximum and a minimum. Such a scale does not allow inferences in either upward or downward direction. If a scale has a maximum but no minimum (only downward inferences allowed) then only the less-than-maximum reading is available. A scale which has a minimum but no maximum (only upward inferences allowed) only has the less-than-minimum interpretation.
2.5.5 Towards a More Principled Account of the Ambiguity

So far I have represented the less-than-minimum reading in an ad hoc fashion with the help of the \textit{min} operator. Because of its stipulative character this kind of approach is unsatisfactory, if only because it leaves us without an explanation for why we don’t find the same ambiguity with \textit{more than}-comparatives. In this section I will try to provide a more principled explanation of the ambiguity of \textit{less than}-comparatives. I will first sketch an account based on a syntactic copying operation,\footnote{This account is based on important suggestions made to me by Barbara Partee.} but then I will raise some important questions about this proposal in the light of further data.

In discussing the semantics of comparatives, I have been assuming a process by which comparative deletion like (88a) is related to its subdeletion counterpart (88b) by means of a copying operation:

\begin{align*}
\text{(88)} & \quad \text{a. The helicopter was flying higher than a plane can fly.} \\
& \quad \text{b. The helicopter was flying higher than a plane can fly high.}
\end{align*}

In order to be more explicit about how this comes about, let’s suppose that \textit{higher} is decomposed as \textit{-er high} (where \textit{-er} is the comparative morpheme) and that \textit{high} is then copied into the comparative clause:
(89) The helicopter was flying -er high than a plane can fly _ high.

It is a traditional observation that less is the comparative form of little (see for instance Bresnan (1973)). Thus less high can be decomposed as something like -er little high, where -er is the comparative morpheme. The sentence The helicopter was flying less high than a plane can fly will then be analyzed as in (90):

(90) The helicopter was flying -er little high than a plane can fly.

Now note that when we copy the head the comparative into the comparative clause we have a choice between copying either high or little high. As a consequence, the sentence is ambiguous between the structures (91a) and (b), where the part that is copied is underlined:

(91) a. The helicopter was flying -er little high than a plane can fly _ high.
    b. The helicopter was flying -er little high than a plane can fly _ little high.

I will assume that little is an operator which turns an adjective into its antonym; in other words, little high means the same thing as low. This implies that (91a) and (b) can be paraphrased as in (92a) and (b):
(92)  
\( a. \) The helicopter was flying less high than how high a plane can fly.  
\( b. \) The helicopter was flying less high than how low a plane can fly.

As far as semantics is concerned, (91a) and (b) correspond to the representations in (93a) and (b), respectively:

(93)  
\( a. \) The helicopter was flying less high than \( max(\lambda d [\text{a plane can fly } d \text{-high}]) \).  
\( b. \) The helicopter was flying less high than \( max(\lambda d [\text{a plane can fly } d \text{-little high}]) \).

(93) captures the ambiguity that we are after. (a) represents the less-than-maximum interpretation of the sentence and (b) the less-than-minimum reading, assuming that the maximal degree \( d \) such that a plane can fly \( d \)-little high is identical to the minimal degree \( d \) such that a plane can fly \( d \)-high.

This account of the ambiguity of less than comparatives does not hinge on the use of two distinct operators \( max \) and \( min \), but rather it assumes that less than-comparatives always involve maximality. The ambiguity is not a lexical one; it depends on what part of the head of the comparative gets copied into the comparative clause. This way it is explained why more than comparatives are not ambiguous. If the head of the comparative is \(-er \text{ tall} \) then the only option is to copy the adjective \text{tall}. Of course, the account also raises some important question, in particular, with respect to the
nature of the supposed copying process. The account relies on purely syntactic means like decomposition and copying to solve a problem that seems more semantic than syntactic. It would be nice if a semantic analogue of the syntactic copying process could be found, but I have no suggestion on how this could be done.

There is also a more specific reason why I have some doubt about this syntactic copying approach. The same ambiguity between a less-than-minimum and a less-than-maximum interpretation also shows up with certain more than comparatives, namely those involving adjectives like short or low, that is, adjectives that are the marked member of a pair of polar antonyms. Thus, (94) has the same two readings as the less than-comparative The helicopter was flying less high than a plane can fly:

(94) The helicopter was flying lower than a plane can fly.

This sentence can either mean that the helicopter was flying at an altitude below the minimal altitude at which a plane can fly, or that the helicopter was flying at an altitude below the maximal altitude at which a plane can fly. That this second reading is also available is quite unexpected. The first (less-than-minimal) reading can simply be represented as in (95a), where the adjective in the comparative clause is low. To represent the second (less-than-maximal) reading we have to replace low by its antonym high, as in (b):
(95)  

   a. The helicopter was flying lower than $\max(\lambda d[a\ plane\ can\ fly\ d\text{-low}])$.

   b. The helicopter was flying lower than $\max(\lambda d[a\ plane\ can\ fly\ d\text{-high}])$.

Apparently, the element that gets copied into the comparative clause can not only be the head of the comparative $low$, but also its unmarked antonym $high$. This situation is problematic for a strict syntactic copying account. We might save the copying approach if we allow ourselves to lexically decompose $lower$ as $less\ high$ and ultimately -$er\ little high$. I do not find this kind of lexical decomposition analysis very attractive, however, because it seems to lack any independent motivation.

To sum up, we can account for the ambiguity of $less\ than$-comparatives by making appeal to decomposition and copying. This account is at least explanatory to the extent that it allows us to give an explanation of why $less\ than$-comparatives are ambiguous, whereas $more\ than$-comparatives aren’t. However, the account makes use of syntactic processes which are of a somewhat dubious nature, and moreover the ambiguity seems to be more general in that it not only occurs with $less\ high$ but also with $lower$. Since at this point the copying account is the best I can offer, I will leave this issue for future research.  

\[\text{\textsuperscript{18}}\]

\[\text{\textsuperscript{18}}\ For\ a\ proposal\ that\ gives\ a\ semantic\ analysis\ of\ the\ ambiguity\ of\ less\ than-comparatives\ in\ terms\ of\ the\ reversal\ of\ scales,\ see\ Rullmann\ (to\ appear).\]
2.6 Equatives and Differential Comparatives

In the preceding sections I have shown how maximality plays a crucial role in the semantic analysis of two types of comparatives, *more than*-comparatives and *less than*-comparatives. In this section I will discuss two other comparative constructions, namely equatives and differential comparatives. As before, monotonicity and the distribution of negative polarity items will be important diagnostics. It will turn out that both equatives and differential comparatives exhibit a particular ambiguity, which is relevant for the question whether they are downward entailing or not.

2.6.1 Equatives

Equatives are comparative constructions of the form *as...as...*. That equatives involve maximality can be seen in examples such as (96a). This sentence means that the speed at which Marcus ran was equal to the maximal speed at which he could run, as indicated in (96b):

(96) a. Marcus ran as fast as he could (run).

b. Marcus ran as fast as \( max(\lambda d[\text{Marcus could run } d\text{-fast}]) \).
Because of maximality, we expect to find negative island effects in equatives, an expectation which is confirmed by the following data:

(97)  a. * Marcus ran as fast as he couldn’t.
     b. * Marcus ran as fast as he never did.
        (cf. Marcus ran as fast as he always did.)

Just like more than-comparatives and less than-comparatives, equatives can license negative polarity items. (98a) and (b) are two examples taken from McCawley (1981) (his 7.3.16 a and b), who credits J. R. Ross for the observation; to these I add (96c) and (d):

(98)  a. Two glasses was as much as I cared to drink.
     b. That was as much as he was willing to lift a finger to do.
     c. Marcus ran as fast as he had ever done before.
     d. Marcus ran as fast as anybody else did.

These examples lead us to conclude that equatives must be downward entailing. In order to see whether the analysis in terms of maximality can account for this, consider the set (99), where $Adj$ is a degree adjective and $D$ is a set of degrees:

(99) $\{x | x \text{ is as tall as } \max(D)\}$
If D is expanded, then $\text{max}(D)$ stays the same or goes up. But note that what happens to the set \{\text{x is as tall as } \text{max}(D)\} depends on exactly how the equative is understood. If ‘is as tall as’ is interpreted as ‘is at least as tall as’, then \{\text{x is as tall as } \text{max}(D)\} will shrink or stay the same. However, if we take ‘is as tall as’ to mean ‘is exactly as tall as’, then it is not possible to say whether \{\text{x is as tall as } \text{max}(D)\} will expand, shrink, or stay the same. We should therefore distinguish the following two interpretations of the equative construction:

\begin{align*}
(100) \quad \text{John is as tall as } \text{max}(D). \\
\text{a. } \text{‘at least as tall as’}: & \exists d [\text{tall}(j,d) \land d \geq \text{max}(D)] \\
\text{b. } \text{‘exactly as tall as’}: & \exists d [\text{tall}(j,d) \land d = \text{max}(D)]
\end{align*}

On the interpretation in (a) the equative is downward entailing, but on the interpretation in (b) it is non-monotone, i.e. neither upward not downward entailing.

\begin{align*}
(101) \quad \text{If } D \subseteq D', \text{ then} \\
\{x \mid \exists d [\text{tall}(x,d) \land d \geq \text{max}(D')]\} & \subseteq \{x \mid \exists d [\text{tall}(x,d) \land d \geq \text{max}(D)]\}. \\
(102) \quad \text{If } D \subseteq D', \text{ then we can neither conclude that} \\
\{x \mid \exists d [\text{tall}(x,d) \land d = \text{max}(D)]\} & \subseteq \{x \mid \exists d [\text{tall}(x,d) \land d = \text{max}(D')]\} \text{ nor that} \\
\{x \mid \exists d [\text{tall}(x,d) \land d = \text{max}(D')]\} & \subseteq \{x \mid \exists d [\text{tall}(x,d) \land d = \text{max}(D)]\}.
\end{align*}
We predict that negative polarity items can only occur in equatives if these are understood on the ‘at least’ interpretation. This is confirmed by the important observation that equatives which have an explicit occurrence of \textit{exactly} do not license negative polarity items, as was pointed out by Horn (1972) and by Seuren (1984). (103) and (104) are from Horn (1972), examples (1.76c) and (d); (105) is (30) from Seuren (1984):

(103) a. John is at least as tall as \textit{any} of his friends.
    b. * John is exactly as tall as \textit{any} of his friends.

(104) a. John is at least as tall as he \textit{ever} was.
    b. * John is exactly as tall as he \textit{ever} was.

(105) a. Jim is at least as competent as \textit{anybody} here could possibly be.
    b. * Jim is exactly as competent as \textit{anybody} here could possibly be.

I will not go into the question what the exact relationship is between the ‘at least’ interpretation and the ‘exactly’ interpretation of the equative. Horn (1972) argues that this is not a real truth-conditional ambiguity, but that ‘exactly’ interpretation is derived from the ‘at least’ interpretation as a result of a Gricean scalar implicature. The evidence for this is that this implicature can be canceled, as in the following examples (adapted from Horn’s (1.77)):
(106) John is as tall as Bill, if not taller.
(107) Not only is John as tall as Bill, he’s even taller.

The downward entailing nature of equatives on the ‘at least’ interpretation, and
the non-downward entailing nature of equatives on the ‘exactly’ interpretation, is
confirmed by the fact that the inferences in (108) is valid, whereas the one in (109) is
not:

(108) John makes at least as much money as a linguist can make ⇒
John makes at least as much money as a syntactician can make.

(109) John makes exactly as much money as a linguist can make ⇔
John makes exactly as much money as a syntactician can make.

On the ‘at least’ interpretation, equatives are also anti-additive, as shown in (110):

(110) \{x | \exists d [\text{tall}(x,d) \land d \geq \text{max}(D \cup D')]\} =
\{x | \exists d [\text{tall}(x,d)] \land \{x | \exists d [\text{tall}(x,d)] \land d \geq \text{max}(D')\}\}

Since equatives on the ‘exactly’ interpretation are not downward entailing, they cannot
be anti-additive either. The (in)validity the inferences in (111) and (112) confirms this
difference:
(111) John is at least as tall as Bill or Sam is ⇔
John is at least as tall as Bill is and John is at least as tall as Sam is.

(112) a. John is exactly as tall as Bill or Sam is ≠
John is exactly as tall as Bill is and John is exactly as tall as Sam is.
b. John is exactly as tall as Bill is and John is exactly as tall as Sam is ⇒
John is exactly as tall as Bill or Sam is.

Note that the inference in (112b) is valid because the antecedent can only be true if Bill and Sam are the same height, and in that case the consequent will also be true. Because of the invalidity of (112a), however, the ‘exactly’ reading is not anti-additive.

Summarizing, we have seen that the semantic properties of equatives depend on the way they are interpreted. On the ‘at least’ interpretation they are downward entailing and anti-additive and they license negative polarity items, whereas on the ‘exactly’ interpretation they are non-monotone and do not license negative polarity items. In the next section we will see that differential comparatives behave in exactly the same way.
2.6.2 Differential Comparatives

Differential comparatives (the term is from von Stechow (1984)) are *more than* or *less than*-comparatives which include an explicit measure phrase indicating the difference between the two degrees that are being compared. Some examples are given in (113):

(113)  

a. John is two inches taller than Bill is.  
b. John is two inches less tall than Bill is.  
c. John is two inches taller than a fighter pilot can be.

I will concentrate here on differential *more than*-comparatives; differential *less than*-comparatives don’t seem to introduce any new complications in addition to the ones already discussed in section 2.5.

Differential comparatives are ambiguous between an ‘at least’ interpretation and an ‘exactly’ interpretation in the same way that equatives are. (Again, I am non-committal as to whether this is a truth-conditional ambiguity or one that is due to a scalar implicature.) The two interpretations of (113a) are given in (114):

(114)  

a. $\exists d[\text{tall}(j,d) \land d \geq \max(\lambda d'[\text{tall}(b,d')]) + 2\text{inch}]$  
b. $\exists d[\text{tall}(j,d) \land d = \max(\lambda d'[\text{tall}(b,d')]) + 2\text{inch}]$
Here ‘+’ stands for the operation of addition defined on degrees. Thus, the expression ‘\( \text{max}(\lambda d'[\text{tall}(b,d')])+2\text{inch} \)’ denotes the degree we get by adding two inches to Bill’s (maximal) degree of tallness. The reader is referred to section 2.10 for some discussion of how addition on degrees can defined in a theory in which degrees are construed as equivalence classes of individuals. Note that differential comparatives are only possible with degree predicates for which a system of measurement exists, such as \textit{tall} or \textit{heavy}, but nor for other predicates like \textit{beautiful} or \textit{impressive}. We need the maximality operator in differential comparatives to account for examples like (113c), which means that John is two inches taller than the maximal height that a fighter pilot can be.

As with equatives, the logical properties of the differential comparatives depend on whether they get the ‘at least’ interpretation or the ‘exactly’ interpretation. On the ‘at least’ interpretation, differential comparatives are downward entailing and anti-additive; on the ‘exactly’ interpretation they are non-monotone. The downward entailing nature of the ‘at least’ interpretation is demonstrated in (115) and its anti-additivity in (116):

(115) If \( D\equiv D' \), then

\[
\{x \mid \exists d[\text{tall}(x,d) \land d\geq \text{max}(D')+2\text{inch}]\} \subseteq \{x \mid \exists d[\text{tall}(x,d) \land d\geq \text{max}(D)+2\text{inch}]\}
\]
(116) \( \{ x \mid \exists d [ \text{tall}(x,d) \land d \geq \max(D \cup D') + 2 \text{ inch}] \} = \) 
\( \{ x \mid \exists d [ \text{tall}(x,d) \land d \geq \max(D) + 2 \text{ inch}] \} \) \( \land \) \( \{ x \mid \exists d [ \text{tall}(x,d) \land d \geq \max(D' ) + 2 \text{ inch}] \} \)

The non-monotonicity of the ‘exactly’ interpretation is shown in (117):

(117) If \( D \subseteq D' \), then we can neither conclude that
\( \{ x \mid \exists d [ \text{tall}(x,d) \land d = \max(D) + 2 \text{ inch}] \} \) \( \subseteq \) \( \{ x \mid \exists d [ \text{tall}(x,d) \land d = \max(D') + 2 \text{ inch}] \} \)

nor that
\( \{ x \mid \exists d [ \text{tall}(x,d) \land d = \max(D') + 2 \text{ inch}] \} \) \( \subseteq \) \( \{ x \mid \exists d [ \text{tall}(x,d) \land d = \max(D) + 2 \text{ inch}] \} \).

These theoretical considerations are confirmed by the (in)validity of the corresponding natural language inferences:

(118) John makes at least $1000 more than a linguist can make \Rightarrow \) 
John makes at least $1000 more than a syntactician can make.

(119) John makes at least $1000 more than Bill or Sam can make \Leftrightarrow \) 
John makes at least $1000 more than Bill can make and John makes at least $1000 more than Sam can make.
(120) a. John makes exactly $1000 more than a linguist can make  
John makes exactly $1000 more than a syntactician can make.

b. John makes exactly $1000 more than a syntactician can make  
John makes exactly $1000 more than a linguist can make.

(121) a. John makes exactly $1000 more than Bill or Sam can make  
John makes exactly $1000 more than Bill can make and John makes exactly $1000 more than Sam can make.

b. John makes exactly $1000 more than Bill can make and John makes exactly $1000 more than Sam can make  
John makes exactly $1000 more than Bill or Sam can make.

Note that (121b) is valid because the antecedent entails that Bill and Sam make the same amount of money.

As with equatives, we predict that only the ‘at least’ reading can license negative polarity items but that the ‘exactly’ interpretation can’t. I believe the following examples do indeed show the expected contrast:

(122) a. John is at least two inches taller than any of his friends are.

b. John makes at least $1000 more than I ever did.
2.7 Universal Quantification vs. Maximality

In the preceding sections I have shown how a semantics of comparatives that makes use of the maximality operator \( \text{max} \) can account for the properties of various types of comparative constructions, in particular monotonicity patterns and the distribution of negative polarity items. Now I will briefly discuss an alternative approach to the semantics of comparatives according to which comparatives do not involve reference to a maximal degree, but universal quantification over degrees. This approach has been adopted in different ways by a number of authors including Cresswell (1976), Hoeksema (1983), and Pinkal (1989). The analyses of these authors differ in many ways from each other, and therefore I will keep the discussion at a very general level, ignoring many aspects of their analyses that are irrelevant for my present purposes. I believe their analyses all share the same basic problems which I will discuss below.

The basic idea of the universal quantification analysis is that a sentence like (124a) is analyzed along the lines of (124b):

\[(123) \quad \begin{align*}
    a. & \quad * \text{ John is exactly two inches taller than any of his friends are. } \\
    b. & \quad * \text{ John makes exactly $1000 more than I ever did. }
\end{align*}\]
(124)  a. John makes more money than a linguist can make.
    b. $\exists d [\text{John makes } d\text{-much } \land \forall d' [\text{A linguist can make } d\text{-much } \Rightarrow d > d']]$

According to this analysis the comparative contains a universal quantifier over degrees. The comparative clause forms the restrictor over this universal quantifier. Note that the universal quantification analysis and the maximality analysis are actually not that different from each other, because the definition of the maximality operator itself involves universal quantification. There are certain empirical differences between the two analyses, though, and as I will show these turn out in favor of the maximality analysis.

As long as we restrict ourselves to more than-comparatives, the universal quantification analysis seems to do just as well as the maximality analysis. The fact that the comparative clause is the restrictor of a universal quantifier, explains why it is downward entailing and anti-additive, and hence also why negative polarity items can appear there. As is well known, the restrictor of an explicit universal quantifier has precisely those properties as well:
(125) a. Everyone who owns a pet is happy ⇒
   Everyone who owns a dog is happy.

   b. Everyone who owns a dog or a cat is happy ⇒
   Everyone who owns a dog is happy and everyone who owns a cat is
   happy.

   c. Everyone who owns any kind of pet is happy.

Problems arise when we turn to less than-comparatives, however. As discussed
extensively in section 2.5, less than-comparatives are ambiguous between a less-than-
minimum and a less-than-maximum reading. The universal quantification analysis only
captures the less-than-minimum reading. (126a) for instance will be analyzed as (126b):

(126) a. The helicopter was flying less high than a plane can fly.

   b. ∃d[the helicopter was flying d-high ∧ ∃d'[A plane can fly d’-high →
       d<d’]]

On this analysis the sentence is true iff the helicopter was flying at an altitude below the
minimal altitude at which a plane can fly. Because the comparative clause is again the
restrictor of the universal quantifier, it is correctly predicted that on this reading less
than-comparatives are downward entailing and anti-additive and can license negative
polarity items. The problem for the universal quantification analysis is, though, that it
has no way of accounting for the less-than-maximum reading, which is upward entailing
and does not license negative polarity items. Of course, we can come up with a logical form that represents the less-than-maximum reading in a way that shows a superficial resemblance to (124b) and (126b). One possible way of doing this is (127):

\[(127) \quad \exists d[\text{the helicopter was flying } d\text{-high } \land \neg \forall d'[\text{A plane can fly } d'\text{-high } \rightarrow \neg d \geq d']]\]

But this would be completely ad hoc move, leaving us without a real explanation for why less than-comparatives are ambiguous, whereas more than-comparatives aren’t. The universal quantification account does not have a way of picking out the maximal degree from a set of degrees and therefore it cannot express the proposition that a certain degree is below this maximal degree, unlike the maximality account.

The universal quantification approach also has a problem with equatives and differential comparatives. It can adequately account for the ‘at least’ readings of these constructions, but not for the ‘exactly’ readings. First consider the ‘at least’ readings. The equative in (128a) and the differential comparative in (129a) are interpreted as in (128b) and (129b), respectively:

\[(128) \quad \begin{align*}
a. & \quad \text{John makes at least as much as a linguist can make.} \\
b. & \quad \exists d[\text{John makes } d\text{-much } \land \forall d'[\text{A linguist can make } d'\text{-much } \rightarrow \neg d \geq d']] \\
\end{align*}\]
(129)  a. John makes at least $1000 more than a linguist can make.
    b. $\exists d [\text{John makes } d\text{-much } \land \forall d'[\text{A linguist can make } d'\text{-much } \to \\
    d \geq d' + $1000]]$

To represent the ‘exactly’ interpretation we would have to replace the \( \geq \)-sign in these formulas by a \( = \)-sign:

(130)  a. John makes at least as much as a linguist can make.
    b. $\exists d [\text{John makes } d\text{-much } \land \forall d'[\text{A linguist can make } d'\text{-much } \to \\
    d = d' + $1000]]$

(131)  a. John makes at least $1000 more than a linguist can make.
    b. $\exists d [\text{John makes } d\text{-much } \land \forall d'[\text{A linguist can make } d'\text{-much } \to \\
    d = d' + $1000]]$

Now the problem is that these formulas will be self-contradictory in any situation in which all linguists do not all have the same salary. Suppose that a linguist can make anything between $1000 and $10,000 a month. Then according to (130b) John should make exactly $1000 and exactly $10,000 (and everything in between), which is a contradiction. The same problem arises for (131b).

To sum up, the universal quantification approach cannot adequately account for the less-than-maximum reading of less than-comparatives and for the ‘exactly’ reading
of equatives and differential comparatives. Note that these are precisely the cases in which comparatives are not downward entailing and do not license negative polarity items. The approach in terms of maximality is able to account for these cases.

2.8 Further Issues in the Semantics of Comparatives

2.8.1 Unexpected Wide Scope Readings

Every account of the semantic of clausal comparatives that I am aware of faces a problem involving the scope of certain quantifiers, such as everybody and most people, and conjoined NP's. Von Stechow’s account in terms of maximality, which I have defended in this chapter, is no exception to this. Take the following sentence:

(132) John weighs more than everybody else weighs.

As pointed out by Heim (1985), the maximality account predicts that (132) presupposes that everybody other than John has the same weight, as can be seen when we consider the interpretation of this sentence as given in (133):

(133) John weighs more than max(λd[everybody else weighs d-much]).
However, intuitively (132) is perfectly fine in a situation in which people have all kinds of different weights. It says that for every person \( x \) (other than John) John weighs more than \( x \).

The intended interpretation can be obtained if we allow the NP (everybody else) to be quantifier raised (or quantified in) so that it gets wide scope over the comparative. It is quite surprising that NP’s can take wide scope this way, since the scope of universal quantifiers is normally restricted to the minimal clause in which they appear.\(^{19}\) Wide scope readings can also be obtained for other upward entailing quantifiers, as shown in (134):

\[
\begin{align*}
(134) & \quad \begin{align*}
& a. \quad \text{John weighs more than most people weigh.} \\
& b. \quad \text{John weighs more than many people weigh.} \\
& c. \quad \text{John weighs more than some people weigh.}
\end{align*}
\end{align*}
\]

Essentially the same problem arises with conjoined NP’s, as in (135):

\[
\begin{align*}
(135) & \quad \begin{align*}
& a. \quad \text{John weighs more than Bill and Sam weigh.} \\
& b. \quad \text{John weighs more than } \max(\lambda d[\text{Bill and Sam weigh } d\text{-much}])
\end{align*}
\end{align*}
\]

\(^{19}\) Lerner (1992; 1993) develops a proposal in which the quantifier can be given wide scope by adjoining it to the IP of the comparative clause, analogous to May’s (1985) treatment of inverse linking and adopting the latter’s Scope Principle.
Again, the maximality account predicts that (135) presupposes that Bill and Sam have the same weight. The intended interpretation, however, is that John weighs more than Bill and John weighs more than Sam.\footnote{(135) also has another reading, which is irrelevant here, namely that John weighs more than Bill and Sam weigh together.}

Note, however, that downward entailing quantifiers are prevented from getting wide scope in this way. (136a), for instance, does not have an interpretation on which it means that for no person $x$ it is the case that John weighs more than $x$:

\begin{align*}
\text{(136) a.} & \quad \ast \text{ John weighs more than nobody else weighs.} \\
\text{b.} & \quad \ast \text{ John weighs more than few people weigh.}
\end{align*}

We not only have to explain why upward entailing quantifiers and conjoined NP’s inside comparative clauses can get wide scope, but also why the narrow scope reading does not seem to be available for these sentences. I would like to argue that this absence of the wide scope reading is only apparent. A sentence like (132) has two readings, but the narrow scope reading differs only from the wide scope reading in that it has a presupposition that the wide scope reading does not have, and for that reason the narrow scope reading is hard to detect intuitively. To show this, we need to look at cases in which the wide scope reading is unavailable for some reason. Interestingly, such cases can be found. The wide scope reading seems to disappear if the NP is more
deeply embedded inside the comparative clause:

\[(137) \quad \text{John weighs more than it seems that everybody else weighs.}\]

Unlike (132), (137) does presuppose that everyone other than John has (at least roughly) the same weight. This makes it plausible that a sentence like (132a) is actually ambiguous between the wide scope reading and the narrow scope reading. Because the latter differs from the former only in that it has a presupposition which the former lacks (namely that everybody other than John has the same weight), it is not surprising that (132a) may not feel like an ambiguous sentence intuitively.

In conclusion, we can say that for some ill-understood reason, certain quantifiers in certain positions can take scope outside the comparative clause. This is a problem for any theory of the comparative, including the one advocated here. The narrow scope reading is sometimes hard to detect for pragmatic reasons, but it does have the meaning that it is predicted to have on the maximality account of the semantics of the comparative.
2.8.2 The Russell Ambiguity

In addition to the argument based on maximality, von Stechow gives another argument in support of the idea that the comparative clause is a degree-denoting expression. Although the argument is attractive at first sight, Hoeksema (1984) and Heim (1985) have shown that it is seriously flawed, and that an alternative analysis of the relevant data is possible, which does not depend on the idea that the comparative clause is a denoting expression. In this section, I will review von Stechow’s argument and Hoeksema’s and Heim’s counter arguments. The upshot of the discussion will be that the data discussed in this section do not offer any additional support for von Stechow’s claim (and mine) that comparative clauses denote degrees, neither do they undermine it.

Von Stechow’s argument is based on the ambiguity of sentences like (138), which was originally pointed out by Bertrand Russell (1905), and rediscovered by Postal (1974):

(138) I thought your yacht was larger than it is.

The easiest way to make clear that this sentence is indeed ambiguous is by citing Russell’s original anecdote:
I have heard of a touchy owner of a yacht to whom a guest, on first seeing it, remarked, “I thought your yacht was larger than it is”; and the owner replied, “No, my yacht is not larger than it is.” What the guest meant was, “The size that I thought your yacht was is greater than the size your yacht is”; the meaning attributed to him is, “I thought the size of your yacht was greater than the size of your yacht.” (Russell (1905), p. 489)

Following Russell, von Stechow analyzes the ambiguity of (138) as a scope-ambiguity. Since in his analysis the comparative clause is a denoting expression analogous to NP’s, it can undergo Quantifier Raising (QR), taking the comparative clause outside the scope of the propositional attitude verb thought. Thus, the intended (non-contradictory) reading of (138) has a Logical Form along the lines of (139):

(139) \[ \text{CP} \text{Op} \text{IP} [\text{your yacht is } d\text{-large}] \text{I thought your yacht was larger than } d] \]

The interpretation of (139) is obtained by means of standard techniques. (140a) gives the translation for the raised comparative clause, and (140b) that for the main clause. Note that the variable \(d\) (corresponding to the trace of the raised comparative clause) is bound by a \(\lambda\)-operator. Note that here we are switching to an intensional language

21 Giving the comparative clause wide scope in this manner has also been proposed by Pinkal (1989), and, in a quite different framework, Larson (1988).

22 Von Stechow’s actual proposal is slightly different in that he assumes that than is pied-pied when the complement undergoes QR, but that difference is inconsequential for our purposes.
The denotation of the sentence as a whole is then obtained by applying the function denoted by (140b) to the degree denoted by (140a). It is crucial that (140a) is outside the scope of the verb thought. It therefore denotes the size of the yacht in the actual world. As a result the sentence expresses the proposition that I believed de re about the actual size of the yacht that the yacht is larger than that.

As von Stechow himself admits, there is an alternative solution to the Russell-problem which does not rely on scope but on the way in which the comparative clause is interpreted. According to this solution, the verb is inside the comparative clause is evaluated not with respect to the speaker’s belief-worlds but with respect to the actual world in which the sentence is uttered. That is, (138) is interpreted as if it were (141):

\[
\text{(141) I thought your yacht was larger than it actually is.}
\]

The adverb actually can be analyzed as an indexical modal operator that takes the sentence back to the actual world, analogous to the way in which now takes the sentence back to the present time (Kamp (1971)). We may assume that on its non-
contradictory reading (138) has an implicit modal operator analogous to *actually*.

The technique that Kamp uses in his analysis of *now* is known as ‘double indexing’. Von Stechow argues that double indexing is not sufficient to account for all cases of the Russell-ambiguity. The argument is based on the fact that the expressive power of the double indexing technique is limited. Double indexing always results in a reading that corresponds to giving the comparative widest scope. Von Stechow gives an example intended to show that sometimes the comparative clause has intermediate scope:

(142) I thought Plato could have been more boring.

This sentence has a reading that von Stechow paraphrases as follows:

(143) For each of my belief worlds $w$, there is a world $w'$ accessible from $w$, such that the degree of Plato’s boringness in $w'$ exceeded the degree of Plato’s boringness in $w$.

This is the kind of ‘intermediate scope’ reading that can not be accounted for with the double indexing technique.
Although von Stechow’s argument against double indexing is valid, it does not lend support to his conclusion that we should therefore adopt his QR analysis. Recently, Cresswell (1990) has shown that double indexing is not powerful enough to account for indexical operators like *now* and *actually* and that we need full quantification over possible worlds. (Earlier, van Benthem (1983) had given parallel arguments for the domain of tense logic.) The kind of logical language assumed by von Stechow (Montague’s IL), does not allow that, because it does not have explicit reference to possible worlds. In this dissertation I am assuming a logical language that does have explicit reference to and quantification over possible worlds. In such a language, the intended reading of (138) can be represented as in (144):

\[
\text{thought}(w)(1,\lambda w' [\exists d [\text{large}(w')(\text{yacht},d) \land \\
\quad d > \max (\lambda d [\text{large}(w)(\text{yacht},d)])]]) \]
\]

Here \(w\) and \(w'\) are variables over possible worlds. The free variable \(w\) is interpreted as referring to the actual world. Because the variable \(w\) in the subformula \(\text{large}(w)(\text{yacht},d)\) does not get bound by the lambda-operator, (144) expresses the proposition that in all world that are compatible with what I thought, the size of the yacht is greater than its size in the actual world.

Hoeksema (1984) and Heim (1985) have shown that von Stechow’s QR analysis cannot be correct, because it is possible to construct sentences parallel to (138).
in which the comparative clause demonstrably must have narrow scope with respect to
the matrix verb, but which nevertheless retain the non-contradictory reading. Hoeksema
gives the following example:

(145) We expect that every schoolboy, thinks he, is brighter than he, is.

This sentence has a reading in which every schoolboy is in the scope of expect (i.e.
there are no expectations about particular schoolboys) while at the same time the
comparative clause than he is is interpreted with respect to the actual world. On this
reading, the sentence is true if in all worlds that are compatible with our expectations it
is the case that every schoolboy ascribes to himself a degree of brightness that is greater
than his degree of brightness in the actual world. Since the comparative clause contains
a pronoun that is bound by the quantifier every schoolboy, it is impossible to account
for this reading by giving the comparative clause wide scope, as in (146), because that
would take the pronoun out of the scope of the quantifier that is supposed to bind it:

(146) [Op, [hej is d,-bright]] [the teacher expect that every schoolboyj thinks
hej is brighter than dj]

Heim (1985) draws the same conclusion as Hoeksema on the basis of a similar example
involving a counterfactual:
If at least one member of the team were faster than he is, we could win the game.

This sentence has a reading on which the indefinite NP *one member of the team* is inside the scope of the counterfactual (that is, the NP is non-specific) whereas the *than*-clause is interpreted with respect to the actual world.

Von Stechow used the QR analysis of the Russell ambiguity as an argument supporting the treatment of the comparative clause as a degree denoting expression. Since there are good arguments against the QR analysis and in favor of allowing full quantification over possible worlds, this particular argument does not go through. However, the discussion in this section does not provide an argument against treating comparative clauses as degree denoting expressions. The evidence is simply neutral in this respect.

2.9 Conclusion

In this chapter, I have shown how von Stechow’s theory that comparative clauses denote maximal degrees can explain the negative island effect in comparatives. The maximality account explains various semantic properties of not only *more than*-comparatives, but also of *less than*-comparatives, equatives, and differential
comparatives, constructions which have not been widely discussed in the literature. These constructions exhibit ambiguities which affect their monotonicity behavior and their ability to license negative polarity items. In the next chapter, I will show that maximality not only plays a role in comparatives, but also in free relatives and questions, and that the maximality account of negative island effects carries over to these constructions as well. The final section of this chapter, section 2.10, is an appendix which discusses the question how degrees can be analyzed as equivalence classes.
2.10 Degrees as Equivalence Classes (Appendix)

2.10.1 Introduction

In the literature various proposals can be found to define degrees in terms of other, less abstract, sorts of things. The approach which has gained the most widespread acceptance in the literature is to construe degrees as equivalence classes of ordinary objects (Cresswell (1973), Klein (1980; 1991), and Hoeksema (1983)). So, for instance, the degree of Jim’s tallness can be defined as the set of all people who are the same height as Jim, and the degree of beauty of the Golden Gate Bridge as the set consisting of all objects which are equally beautiful as the Golden Gate Bridge. In this section, I will first spell out this idea more formally. I will then sketch how a semantics for measure phrases (like two feet) can be given in a theory that defines degrees as equivalence classes (based on Klein (1980;1991)), and finally I will briefly discuss, but not resolve, some complications that arise.
2.10.2 Constructing Degrees

Suppose that associated with any gradable adjective A (like tall or beautiful) there is a two-place relation \( \preceq_A \) and a set \( D_A \). \( D_A \) is a subset of the universe of discourse; intuitively, it is the set containing all and only those objects of which the adjective can be sensibly predicated. \( D_{\text{tall}} \), for instance, will contain people and buildings, among other things, but not ideas or colors, because ideas or colors cannot sensibly be said to be tall or not tall. The relation \( \preceq_A \) (intuitively, at least as tall as or at least as beautiful as) will be assumed to be reflexive, transitive, and connected, on \( D_A \):

- **Reflexivity:** \( \forall x \in D_A [x \preceq_A x] \)
- **Transitivity:** \( \forall x, y, z \in D_A [x \preceq_A y \land y \preceq_A z \rightarrow x \preceq_A z] \)
- **Connectedness:** \( \forall x, y \in D_A [x \preceq_A y \lor y \preceq_A x] \)

Given \( \preceq_A \), we can define an equivalence relation \( \simeq_A \) (as tall as or as beautiful as) as follows:

\[
x \simeq_A y \iff x \preceq_A y \land y \preceq_A x
\]

The degree of ‘A-ness’ (tallness, beauty) of an object \( x \) (notated as \( \deg_A(x) \)) will now be defined as the set of all objects \( y \in D_A \) that stand in the relation \( \simeq_A \) to \( x \), that is:

\[
\deg_A(x) = \{y \in D_A | x \simeq_A y\}.
\]

Thus, \( \deg_{\text{tall}}(\text{john}) \) will be the set of people who are just as tall as John. Let \( \text{DEG}_A \) be the set of all degrees obtained in this way, that is, the set of equivalence classes into

---

\(^{23}\) For the sake of simplicity I will restrict myself to adjectives here, but in principle the account can be extended to other degree predicates, like verbs (e.g. weigh, cost) or adverbs (e.g. often, quickly).
which $D_A$ is partitioned by $\sim_A$. $\succeq_A$ induces a relation $\succeq_A$ on the members of $\text{DEG}_A$, which is defined as follows:\footnote{In the main text I throughout omit the subscript referring to the degree predicate, writing $\succeq$ rather than $\succeq_A$ and $\text{DEG}$ rather than $\text{DEG}_A$.}

\[
\text{deg}_A(x) \succeq_A \text{deg}_A(y) \iff x \succeq_A y
\]

It can be shown that $\succeq_A$ is a linear order, that is, a relation which is reflexive, transitive, connected, and anti-symmetric:

\begin{align*}
\text{Reflexivity:} & \quad \forall d \in \text{DEG}_A[d \succeq_A d] \\
\text{Transitivity:} & \quad \forall d,d',d'' \in \text{DEG}_A[(d \succeq_A d' \land d' \succeq_A d'') \to d \succeq_A d''] \\
\text{Connectedness:} & \quad \forall d,d' \in \text{DEG}_A[d \succeq_A d' \lor d \succeq_A d'] \\
\text{Anti-Symmetry:} & \quad \forall d,d' \in \text{DEG}_A[(d \succeq_A d' \land d' \succeq_A d) \to d = d']
\end{align*}

It is easy to see why $\succeq_A$ is anti-symmetric, whereas $\succeq_A$ isn’t. If Mary and Sue are equally tall, then this doesn’t imply that Mary and Sue are the same person, but it does imply that Mary’s degree of tallness is the same as Sue’s degree of tallness.

Given $\succeq_A$ we can define other relations on degrees in the usual manner:

\[
\begin{align*}
d &\sim_A d' \iff d' \succeq_A d \\
d &\succeq_A d' \iff d' \succeq_A d \land d \neq d' \\
d &\preceq_A d' \iff d' \succeq_A d
\end{align*}
\]
A gradable adjective like tall or beautiful can now be analyzed as a two-place relation between entities and degrees. The proposition that Mary is tall to degree \( d \) can then be expressed by the atomic formula given in (148a); its truth conditions are given in (148b):

\[(148)\]
\[
\begin{align*}
\text{a. } & \text{tall(mary,} d\text{)} \\
\text{b. } & [\text{tall(mary,} d\text{)}] = 1 \iff [d] \in \text{DEG}_{\text{tall}} \text{ and } [\text{mary}] \in [d].
\end{align*}
\]

### 2.10.3 Measuring Degrees

As indicated in the main text, measure phrases can be interpreted as expressions which denote degrees. Measure phrases can be used with the positive form of the adjective (as in (149a)), but also with the comparative (as in (149b-d)):

\[(149)\]
\[
\begin{align*}
\text{a. } & \text{Mary is six feet tall.} \\
\text{b. } & \text{Mary is taller than six feet.} \\
\text{c. } & \text{Mary is two inches taller than Sue is.} \\
\text{d. } & \text{Mary is twice as tall as Sue.}
\end{align*}
\]

Just saying that a measure phrase like six feet denotes a certain degree of tallness is not sufficient. We need some account of what makes it possible to add one degree to
another, as in (149c), and of how degrees can be multiplied, as in (149d). Moreover, rather than leaving the complex measure *six feet* semantically unanalyzed, we would like to derive its meaning compositionally from the meaning of its parts *six* and *feet*.

Klein (1980; 1991) gives an outline of how a rudimentary system of measurement for length can be constructed. Suppose we have some standard object, such as a bar, that is exactly one foot long. (Think of the metal bar conserved in Paris that used to be the standard meter.) Let’s call this standard object ‘foot’. Now let ft be the equivalence class of all objects that are the same length as foot, that is ft = \text{deg}_{\text{long}}(\text{foot}). Suppose furthermore that we have an operation of concatenation, denoted by the symbol ∪, defined on physical objects like rods and bars. Two objects \(a\) and \(b\) are concatenated by putting them end to end in a straight line; the result is another object \(a \cup b\). Given this operation of physical concatenation, we can define addition on degrees of length as follows:

Let \(d\) and \(d'\) ∈ \text{DEG}_{\text{long}}. Then:

\[
d + d' = \{d'' : d'' \in \text{DEG}_{\text{long}} | \exists x \in d \exists y \in d' [x \cup y \in d'']\}.
\]

In other words, the sum of two degrees of length \(d\) and \(d'\) is the set of all objects that are exactly as long as the object we get by concatenating one element of \(d\) with one element of \(d'\).

Once we can add degrees, we can also multiply them. For instance, for any degree \(d \in \text{DEG}_{\text{long}}\), \(2d = (d + d)\). More generally, we can give the following recursive
definition of multiplication of degrees:

Let \( d \in \text{DEG}_{\text{long}} \). Then:

(a) \( 1d = \text{def} \ d \)

(b) \( nd = \text{def} (n-1)d + d \) (for any \( n > 1 \)).

So 6ft is the length of the object we get by concatenating six objects that are each one
foot in length.

The system of measurement constructed this way is rather simplistic, of course,
but it shows how in principle a semantics for measure phrases can be given in a theory
in which measure phrases denote degrees and degrees are equivalence classes of
objects.

The semantics for measure phrases just outlined is specific for the dimension of
length, but can be generalized to other dimensions, such as height and width, if instead
of concatenation (defined above as putting two objects end to end in a straight line) we
use other operations, like putting two objects on top of each other, or putting two
objects side by side. Constructing a system of measurement for other properties like
speed, temperature or price will be more complicated.\(^{25}\) I won’t speculate on how one
could go about doing this, but it is important to note that it will involve the application

\(^{25}\) Such systems of measurement would be involved in the semantics of sentences
like these:

(i) I was driving fifty miles an hour.
(ii) Today it is five degrees colder than it was yesterday.
(iii) A cheese burger costs twice as much as a hot dog.
of basic concepts from physics and economics, respectively. Systems of measurement are always based on (perhaps rudimentary) scientific theories. Before there were thermometers, people could compare temperatures and for instance say that things like ‘Today it is colder than it was yesterday’, but they didn’t have a way of measuring temperatures yet, nor the natural language expressions (‘Today it is five degrees colder than it was yesterday’) that go with it. Even measuring lengths involves some basic knowledge of physics, like the fact that objects don’t get shorter if you concatenate them.

Of course, for many gradable properties, like beauty, there are no exact measure phrases, simply because we don’t have any reliable way of measuring them. An important advantage of construing degrees as equivalence sets is that it does not make the notion of degree itself dependent on ways of measuring them. This is crucial because in natural language comparatives can be used even if there is no way of measuring the property involved, for instance in a sentence like ‘The Golden Gate Bridge is more beautiful than the World Trade Center’. In the theory outlined above, the only formal difference between a predicate that can be measured (tall) and one that cannot (like beautiful), is that in the former case some additional mathematical structure is imposed on the set of degrees. Addition and multiplication are defined for degrees of tallness, but not for degrees of beauty.
2.10.4 Some Complications

The first complication for the semantics of degrees I want to discuss is that of comparison across different dimensions. Consider the following example:

(150) The table is longer than it is wide.

In this sentence the length of the table is compared to its width. The problem is that, the way they were defined in above, degrees are adjective-specific. The sets $\text{DEG}_{\text{long}}$ and $\text{DEG}_{\text{wide}}$ are disjoint, as are the relations $\geq_{\text{long}}$ and $\geq_{\text{wide}}$. We have no way of comparing a degree of tallness to a degree of width, yet that is what is going on in (150). To deal with such examples we must have a way of comparing degrees across dimensions. Formally, we need an isomorphism between the two sets of degrees $\text{DEG}_{\text{long}}$ and $\text{DEG}_{\text{wide}}$, that is, a function $i$ from $\text{DEG}_{\text{long}}$ to $\text{DEG}_{\text{wide}}$ which is one-to-one and for which the following holds:

$$\forall d, d' \in \text{DEG}_{\text{long}} [d \geq_{\text{long}} d' \iff i(d) \geq_{\text{wide}} i(d')]$$

Given this isomorphism $i$, we can define a relation $\geq$ between degrees of length and degrees of width as follows:

For any $d \in \text{DEG}_{\text{long}}$ and $d' \in \text{DEG}_{\text{wide}}$:

$$d \geq d' \iff_{\text{def}} i(d) \geq_{\text{wide}} d'.$$

Our example (150) will be true iff the degree to which the table is long stands in the relation $\geq$ to the degree to which the table is wide.
The existence of an isomorphism is not a sufficient condition for comparing degrees belonging to different dimensions, however. The length of the table can be compared to its width only because length and width are measured ‘in the same way’. Length and width have a common system of measurement; they are both expressed in terms of feet and inches. It is impossible to compare the length of the table with its weight, simply because we measure length in feet, but weight in pounds, and we can’t compare feet to pounds. This incommensurability of length and weight explains the oddness of the following example, which can only be understood in some metaphorical sense:

(151) The table is longer than it is heavy.

I will not try to spell out what it means for two dimensions to be measured ‘in the same way’.

The second problem I want to mention is that of antonyms. Adjectives typically come in pairs like tall and short, heavy and light, big and small, warm and cold. If we assume that \( x \geq_{\text{tall}} y \) iff \( x \leq_{\text{short}} y \), then tall and short will determine the same equivalence classes, that is, \( \text{DEG}_{\text{tall}} = \text{DEG}_{\text{short}} \). So degrees of tallness will also be degrees of shortness. This is problematic because of the well known fact that of a pair of antonyms one is marked (the ‘negative’ antonym), while the other is unmarked (the ‘positive’ antonym). This is reflected in the following data:
The problem of cross world comparison of degrees has been emphasized by Cresswell (1976).

(152) a. Mary is five feet tall.
   b. ? Mary is five feet short.

(153) a. How tall is Mary?
   b. ? How short is Mary?

The (b) sentences, with the negative antonym, are marked in the sense that they presuppose that Mary is short, whereas the (a) sentences are neutral in that respect. The semantics of degrees outlined above gives us no way of accounting for this difference.

The final problem I would like to discuss is due to the fact that it is possible to compare degrees across possible worlds, as shown by the following example:26

(154) a. If Mary had eaten more as a child, she would have been taller than she is.
   b. Mary was taller than anyone had thought possible.

Each of these examples involves comparison of degrees across different possible worlds. In (154a) Mary’s degree of tallness in the actual world is compared to her degree of tallness in worlds where as a child she ate more than she did in the actual world, while in (b) Mary’s degree of tallness in the actual world is compared to the

26 The problem of cross world comparison of degrees has been emphasized by Cresswell (1976).
degree of tallness she has in other people’s belief worlds. Identifying degrees simply with equivalence classes of individuals will not work for such examples, because Mary does not have the same degree of tallness in each world. This problem can be solved by analyzing degrees not as equivalence classes of individuals, but as equivalence classes of ordered pairs consisting of an individual and a possible world, as proposed by Cresswell (1976). So if Mary is taller in world \( w \) than she is in world \( w' \), there will be two distinct degrees \( d \) and \( d' \) such that \( <\text{mary}, w> \in d \), \( <\text{mary}, w'> \in d' \), and \( d > \text{tall} d' \). Because the issue is largely irrelevant for the issues that are the focus of this dissertation, I will not attempt to work this suggestion out in greater detail.
3.1 Introduction

In the previous chapter I have shown that the semantics of comparatives crucially involves reference to maximal degrees. Using maximality we can explain a number of semantic properties of the comparative, such as their ability to license negative polarity items. Maximality can also account for the negative island effect in comparatives. The purpose of this chapter is to demonstrate that maximality also plays an important role in two *wh*-constructions, namely questions and free relatives. I will begin by examining the negative island effect.
3.1 Negative Islands and Maximality in Questions

Let’s try to see whether von Stechow’s explanation of negative island effects in comparatives can be carried over to questions. Suppose that a question like (1a) actually means something like (1b):

(1)  
a. How tall is John?
b. What is the maximal degree $d$ such that John is $d$-tall?

I will use the following quasi-logical abbreviation for (1b):

(2) $\,?d[d = \max(\lambda d'[\text{John is }d'-\text{tall}])]$

For the moment I will refrain from giving an explicit semantics for the $?-operator. In section 3.4 I will discuss the intended semantics for the $?-operator in the two most influential theories for the semantics of questions, namely those of Karttunen (1977) and Groenendijk and Stokhof (1982; 1984).

Notice that if we assume the interpretation given in (2) then the negative island effect is indeed explained in exactly the same way that it was in the case of the comparative. Consider for instance (3):
(3)  
a.  * How tall (do you think) John isn’t?

b.  $\exists d [d = \max(\lambda d'[\text{John isn’t } d’-\text{tall}])]$

Because there is no maximal degree $d’$ such that John is not $d’$-tall, the denotation of the question is undefined. The same explanation will hold for other downward entailing elements, such as nobody and few people, for exactly the same reason that it did in the case of comparatives. In downward entailing contexts scales are reversed and as a result the maximum point on the scale is not defined.

Further evidence for the hypothesis that questions involve maximality comes from (4), which is parallel to the kind of examples von Stechow uses as evidence for the presence of the maximality operator in comparatives:

(4)  How fast can John run?

(5)  
a.  $\exists d [d = \max(\lambda d'[\text{John can run } d’-\text{fast}])]$

b.  $\exists d [d = \max(\lambda d' [\exists w [\text{John runs } d’-\text{fast in } w]])]]$

Suppose that the maximal speed at which John can run is 7 miles per hour. Then the only truthful answer to (4) could be 7 miles per hour, even though it is of course true that John is also capable of running 6 or 5 miles per hour. This is captured in (5a) by means of the max-operator. (5b) spells out more precisely what that means, on the (simplifying) assumption that the modal can corresponds to existential quantification
over possible worlds. If John can run 7 miles per hour, then there will be at least one possible world at which he runs 7 miles per hour, but there will also be worlds at which he runs 6 or 5 miles per hour, or at which he doesn’t run at all. Thus $\lambda d'[\exists w[\text{John runs } d'-\text{fast in } w]]$ will denote the set of all degrees $d'$ such that there is a world at which John runs $d'$-fast. In the given situation, this will be the set of all degrees of speed up to and including 7 miles per hour. The $max$-operator selects the maximal member from this set, that is, 7 miles per hour.

The account of negative islands in questions sketched above raises some important problems and questions that have to be dealt with. One issue that is raised by the account sketched above is that the maximality requirement for questions appears to be rather ad hoc. Is maximality something that only shows up in degree questions such as (1a), or do we also find it in non-degree questions like Which people did John invite? In this chapter, I will argue that maximality is indeed a much more general phenomenon. To set the stage, in section 3.3 I will discuss the role played by maximality in definites (following Link’s (1983) analysis) and in free relatives (Jacobson (1990)). In section 3.4 I will then return to questions. It will be argued that maximality plays a role in all wh-questions. In questions involving individuals rather than degrees maximality takes the form of exhaustiveness, as has been suggested by Jacobson (1990). The idea that questions are exhaustive (in a specific, well-defined sense) has been argued for by Groenendijk and Stokhof (1982; 1984).
This brings us to a second issue raised by my analysis, namely what kind of more general theory for the semantics of questions the account should be implemented in. In my discussion of questions in section 3.4, I will start out by adopting the semantic theory of questions proposed by Karttunen (1977). Then I will show that by adding exhaustiveness (maximality) to this theory, we end up with an analysis that is very close to that of Groenendijk and Stokhof.

3.3 Maximality in Definite NP’s and Free Relatives

3.3.1 Link’s Semantics for the Definite Determiner

The definite determiner the can be used in (at least) two ways which at first sight appear to be unrelated. When used with a singular count noun, the expresses the uniqueness of the referent of the NP. The sentence The boy is sick, for instance, can only be used felicitously in a situation in which there is only one (salient) boy; the sentence asserts that this unique boy has the property of being sick. When used with a plural count noun, on the other hand, the appears to have the same function as a universal quantifier. The sentence The boys are sick is true iff every boy in the domain of discourse is sick.
Link (1983) has shown that it is possible to give a unified semantics for the definite determiner which accounts for both aspects (uniqueness and universality) of the meaning of the determiner. Link proposes a semantics for plurals which is based on an ontologically rich model structures for natural language interpretation.\footnote{Link’s semantics also applies to the domain of mass nouns where the definite determiner has the same maximizing effect as with count nouns. I will not discuss mass nouns, however.} These model structures include not only run-of-the-mill \textit{atomic} objects, but also the \textit{plural (non-atomic) objects} which are the \textit{sums} of atomic objects. For instance, if John and Bill are two atomic objects in the domain of discourse D, say \textit{j} and \textit{b}, then D also includes their sum \textit{j+b}. The set of atomic objects, which I will call \textit{AT}, is a proper subset of D. Atomic and non-atomic objects are ordered by the \textit{part-of relation} \(\leq\), which is reflexive, transitive and anti-symmetric (\textit{i.e.} \(\leq\) is a (weak) partial order). To give an example, suppose that the domain of discourse D contains exactly three atomic objects, say John, Bill, and Mary. Then D is the partially ordered set depicted in the diagram in (6):

\begin{equation}
\begin{array}{c}
\begin{array}{c}
\bullet j+b+m \\
\quad \bullet j+b \\
\quad \bullet j+m \\
\quad \bullet b+m \\
\quad \bullet j \\
\quad \bullet b \\
\quad \bullet m
\end{array}
\end{array}
\end{equation}

\textbf{(6)} \quad \text{sums}

\textbf{(6)} \quad \text{atoms}

More technically, the structure \(<D,+, \leq, \text{AT}>\) is a \textit{complete atomic join semilattice},
which means that:

1. D is *partially ordered* by $\preceq$ ($\preceq$ is reflexive, transitive, and antisymmetric).

2. $+$ (the *join operation*) maps any non-empty subset A of D onto an element of D. $+A$ is the *least upper bound* of A with respect to $\preceq$; that is, $+A$ is the smallest element of D such that for all $a \in A$, $a \preceq +A$. If A is a set with two elements $\{a, b\}$, we write $a+b$ rather than $+\{a, b\}$.

3. AT is the set of all *atoms* in D, that is, the set of all those elements $a$ of D such that for all $b \in D$, if $b \preceq a$, then $b = a$.

4. D is *atomic* in the sense that for all $a \in A$, there is a $b \in AT$ such that $b \preceq a$.

Given this kind of model structure, we can give a semantics for the plural marking on nouns. A count noun like *boy* will denote a set of atoms, that is, $[\text{boy}] \subseteq AT$. The plural *boys* denotes the set of all non-atomic sums of boys. This is formalized in (7):

\[
\text{(7)} \quad [\text{boys}] = \{x \in (D-AT) | \exists X \subseteq [\text{boy}] \text{ such that } x = +X\}
\]

(7) says that the denotation of *boys* is the set of non-atomic objects $x$ such that $x$ is the result of taking the sum of all the elements in some subset of [boy]. This gives us the right truth-conditions for sentences in which *boys* is used as a predicate, such as (8a), as

---

$^2$ Here I am following Landman’s (1989) adaptation of Link’s analysis, which differs from Link’s own formalization in that it does not include a zero-element.
is shown in (8b):

(8)  a. John and Bill are boys.
    b. \( j+b \in [\text{boys}] \) iff \\
        \( j+b \in \{x \in (\text{D-AT}) \mid \exists X \subseteq [\text{boy}] \text{ such that } x = +X \} \) iff \\
        \( j+b \notin \text{AT} \) and \( \exists X \subseteq [\text{boy}] \text{ such that } j+b = +X \) iff \\
        \( j \neq b \) and \( \{j,b\} \subseteq [\text{boy}] \)

To take a concrete example, suppose that there are three boys in the domain of discourse, John, Bill, and Sam. The denotation of the singular noun \( \text{boy} \) will then be the set \( \{j, b, s\} \), and \( \text{boys} \) will denote the set \( \{j+b, j+s, b+s, j+b+s\} \). In the form of a diagram:

(9)

On the basis of such denotations for singular and plural count nouns, Link shows that it is possible to give an elegant and unified semantics for the definite determiner \( \text{the} \) which encompasses both the uniqueness requirement of singular definite NP’s and the universal quantificational force of plural definites. Let’s first consider the meaning of \( \text{the} \) in plural definites, such as \( \text{the boys} \). Intuitively, the function of the
definite plural article is to collect all the individual boys into the plural individual that is their sum. In the example given above, \textit{the boys} denotes the sum of all boys, \(j+b+s\). This sum-formation can be achieved by defining the denotation of \textit{the boys} as the \textit{maximal} element of the set [boys], that is, the unique element of the set of which all other elements of the set are parts. Thus, the denotation of an NP of the form \textit{the N} can be defined as follows:\footnote{Link’s actual semantics for \textit{the} is somewhat different in that he makes all NP’s denote generalized quantifiers. Following Partee (1987), I assume that referential NP’s have type \(e\) as their basic type, but can be shifted to higher types such as \(<<e,t>,t>\) if necessary.}

\begin{equation}
(10) \begin{align*}
\text{a.} & \quad [\text{the } N] = \max([N]) \\
\text{b.} & \quad \max(A) = \max[x \in A \land \forall x' \in A[x' \leq x]]
\end{align*}
\end{equation}

What about singular definite NP’s? The beauty of Link’s semantics for the definite determiner is that it accounts for those as well. Consider the singular NP \textit{the boy}. The noun \textit{boy} denotes the set of atoms which are boys. Since all the elements of this set are atomic, none of them stands in the \(<\text{-relation to any of the others. Therefore, \(\max([-\text{boy}])\) will only be defined if \([-\text{boy}]\) contains exactly one member, or in other words, if there is exactly one boy in the domain of discourse. If there are no boys or if there is more than one boy, \(\max([-\text{boy}])\) is not defined. Thus, by translating the definite determiner as the \textit{max}-operator, we account for the fact that the definite determiner corresponds to sum-formation in case the noun is plural, but enforces}
uniqueness if the noun is singular.

The definition of the maximality operator in (10b) is formally identical to the definition I gave in chapter 2. The only difference is that in that definition \( \text{max} \) was intended to apply to sets of degrees, whereas here it applies to sets of (atomic and/or non-atomic) individuals. Because the previous chapter dealt exclusively with comparatives, whereas in this section we have only been concerned with the semantics of the definite determiner, it might perhaps appear that we are dealing with two separate notions of maximality, one applying to degrees and the other to individuals. However, in the rest of this chapter I will discuss two \( wh \)-constructions, namely free relative clauses and \( wh \)-questions, which exhibit maximality effects with respect to both degrees and individuals. From this, I conclude that it would not be useful to try to keep the two notions of maximality separate. There is one maximality operator which applies both to sets of degrees and to sets of individuals. When applied to a set of degrees, this operator picks out the highest degree in this set, if there is one. When applied to a set of individuals, the same operator picks out the unique individual from the set that is the sum of all the elements in the set, if there is one. The fact that superficially these appear to be quite different operations is a reflection of the fact that degrees and individuals are ordered in different ways. Individuals are ordered by a (merely) partial ordering (the part-of relation), whereas degrees are ordered by a linear relation (the smaller-than relation).
3.3.2 Maximality in Free Relatives

The apparent double function of the definite determiner (uniqueness and universality) can also be found in free relatives. This has been observed by Jacobson (1990), who illustrates this with the following examples (Jacobson’s (7) and (8)):

(11)  a. I ordered what he ordered for desert.
      b. I ordered the thing he ordered for desert.

(12)  a. John will read whatever Bill assigns.
      b. John will read everything/anything Bill assigns.

In (11a) the free relative clause *what he ordered for desert* appears to have the uniqueness typical for a singular indefinite, as indicated by the paraphrase in (11b). The free relative *whatever Bill assigns* (12a) on the other hand is most appropriately paraphrased with a universal quantifier or free choice *any*. It might be thought that the difference between (11a) and (12a) is due to some lexical difference between the wh-words *what* and *whatever*, and that free relatives headed by *what* require a unique referent and that those headed by *whatever* have the quantificational force of a universal. Jacobson shows that this is not the case, however. For one thing, as she points out, there are free relatives with *what* that appear to have universal force (Jacobson’s (14) and (15)):
(13)  
   a. I read what was on the reading list.  
   b. I read everything that was on the reading list.  

(14)  
   a. Do what the babysitter tells you to do.  
   b. Do everything the babysitter tells you to do.  

Conversely, there are free relatives with whatever that clearly do not have universal quantificational force, such as the following examples in which the only function that -ever appears to have is to indicate ignorance on the part of the speaker (Jacobson’s (9)-(10)):

(15)  
   John read whatever Bill assigned - although I don’t remember what it was, but I do know that it was long and boring.  

(16)  
   Everyone who went to whatever movie the Avon is now showing said it was boring.  

Jacobson argues that the ambiguity between a definite reading and a universal reading of free relatives is only apparent, and that in fact free relatives have a uniform semantics that is very similar to that of definite NP’s. She adduces some interesting facts showing that free relatives are not real universal quantifiers. One argument for this position is the possibility of anaphoric reference by means of the pronoun it
demonstrated in (13) and (14), something which is impossible for real universals (Jacobson’s (11) and (12)):

(17) * John read everything/anything that Bill assigned, although I don’t remember what it was, but I do know that it was long and boring.

(18) * Everyone who went to every/any movie the Avon is now showing said it was boring.

Another argument is that, unlike real universals and free choice any, free relatives cannot be modified by almost (Jacobson’s (78) and (79)):

(19) a. * For years, I did almost whatever you told me to do.
    b. For years I did almost everything/anything you told me to do.

Furthermore, free relatives do not license negative polarity items, but real universals do (Jacobson’s (80) and (81)):

(20) a. * I can read whatever Bill ever read.
    b. I can read everything/anything that Bill ever read.

If free relatives are not universal quantifiers, what are they? According to Jacobson’s analysis, free relatives denote maximal individuals in the same way that
definites do. Their apparent universal force derives from sum-formation and not from universal quantification. Apart from the negative evidence cited above showing that free relatives are not universal quantifiers, Jacobson supplies positive evidence supporting the idea that they denote maximal individuals. This evidence comes from the interpretation of specificational pseudo-clefts, in the sense of Higgins (1973). Consider the following examples (Jacobson’s (53) and (54)):

(21)  a. What is on my plate are beans, rice, and tacos.
    b. Beans, rice, and tacos are on my plate.

(22)  a. What books I read last year are *Moby Dick*, *The Brothers Karamazov*, and *Ted and Sally*.
    b. I read *Moby Dick*, *The Brothers Karamazov*, and *Ted and Sally* last year.

Jacobson points out that there is a difference in interpretation between the (a) and the (b) sentences. (21a) implies that beans, rice, and tacos are the only things that are on my plate, but there is no such implication in (21b). Similarly, (22a), but not (22b), is an exhaustive listing of the books I read last year. The exhaustiveness found in the (a) sentences must be contributed by the semantics of the free relative. (See Higgins (1973) for arguments that the precopular constituent in a pseudo-cleft sentence is indeed simply a free relative.)
The exhaustiveness of examples like (21) and (22) is explained by Jacobson’s proposal that free relatives denote maximal individuals. This idea can be implemented easily with the help of the maximality operator $\text{max}$ introduced above. The free relative in (23a), for instance, will get the denotation given in (23b), which is equivalent to (23c):

(23)

a. what John ordered  
b. $\text{max}(\lambda x[\text{ordered}(j,x)])$  
c. $\exists x[\text{ordered}(j,x) \land \forall x'[\text{ordered}(j,x') \rightarrow x' \preceq x]]$

Thus, the free relative *what John ordered* denotes the maximal individual of which it is true that it is a thing and that John ordered it, or in other words, the sum of all the things that John ordered.

It’s not hard to see how this kind of denotation can be derived from the syntactic structure of free relatives. Essentially following Groos and van Riemsdijk (1979) (and contra Bresnan and Grimshaw (1978)) I will assume that the *wh*-phrase of a free relative clause occupies a complementizer position, more specifically, the Spec-of-CP. Thus (23a) has the syntactic structure given in (24a). (24b)-(e) give the steps in the derivation of the logical translation of this free relative.
Some clarifying comments may be in order. I am assuming that the trace \( t_i \) in the object position of the verb ordered translates as a variable. The complementizer head C is filled with an empty element \( e \), which is coindexed with the phrase in Spec-of-CP and which semantically acts like a \( \lambda \)-abstractor binding the variable corresponding to the trace. The wh-word what acts as a function that takes a set of entities and maps it onto the maximal element of that set.

Jacobson points out that although definite NP’s and free relatives both denote maximal individuals, there is also an important difference. As we saw in section 3.3.1, singular definites denote atomic individuals and plural definites denote non-atomic individuals (sums). For free relatives like what(ever) John ordered this distinction is irrelevant, because what is not overtly marked as either singular or plural. The denotation of the free relative can be either atomic or non-atomic, depending on the situation. If John ordered exactly one thing, then what(ever) John ordered will denote the atomic individual that is the unique thing that John ordered. If John ordered more than one thing, however, it will denote the individual that is the sum of everything he
ordered. Hence the apparent ambiguity of free relatives between unique and universal readings. In being free to denote either atomic or non-atomic maximal individuals, free relatives have something in common with definite NP’s in languages that do not mark the distinction between singulars and plurals. In such a language, a noun denotes a set that includes both atomic individuals and their sums, and a definite NP will therefore denote the maximal individual in the set denoted by the noun. If this set happens to be a singleton, then the NP denotes the (unique) member of the set, but if the set has more than one element, then the NP denotes the sum of all the elements in the set.

Not all free relatives are unmarked for the singular/plural distinction, however. The free relative in (25) is marked singular and the one in (26) is marked plural:

(25) I’ll read whatever book John read.
(26) I’ll read what(ever) books John read.

The denotation of (25) is like that of a singular definite: if John read exactly one book, it denotes that book; otherwise the denotation is undefined. (26) is like a plural definite in that it denotes the sum of all the books that John read, provided he read at least two. Given the translations in (27) and (28), this difference between (25) and (26) will follow on the assumption that the predicate book only applies to atomic individuals, whereas books applies to non-atomic ones:
Jacobson’s own analysis is slightly different. She assumes that free relatives have $\langle e, t \rangle$ as their basic type, and are shifted down to type $e$ when they are used referentially. 

To conclude this review of Jacobson’s proposal, let’s return to the specificational pseudo-clefts in (21) and (22) (repeated below) that were cited as a source of evidence for her analysis, to see in detail how the exhaustive listing effect is accounted for.

(29) What is on my plate are beans, rice, and tacos.
(30) What books I read last year are *Moby Dick*, *The Brothers Karamazov*, and *Ted and Sally*.

Williams (1983) and Partee (1986) have argued for an analysis of specificational pseudo-clefts according to which the free relative functions as a predicate that is applied to the postcopular NP. I have treated free relatives here as referential expressions, of type $e$, but I will assume that in specificational pseudo-clefts they are shifted to type $\langle e, t \rangle$. This kind of type shifting is a natural process, as argued by

---

4 Jacobson’s own analysis is slightly different. She assumes that free relatives have $\langle e, t \rangle$ as their basic type, and are shifted down to type $e$ when they are used referentially.
Partee (1987). The specific type shift we need is the one she calls 'ident', defined as in (31), which maps an entity onto the singleton set containing it:

\[(31) \quad \text{ident}(x) =_{\text{def}} \lambda y[y = x]\]

Applying this type shift to the precopular free relative of (30) yields the predicate denoted by the following expression:

\[(32) \quad \lambda y[y = \max(\lambda x[\text{books}(x) \land \text{read}(I,x)])]\]

This predicate in turn is applied to the postcopular NP, which gives us the proposition in (33) (where m, b, and t are constants denoting *Moby Dick*, *The Brothers Karamazov*, and *Ted and Sally*, respectively):

\[(33) \quad m + b + t = \max(\lambda x[\text{books}(x) \land \text{read}(I,x)])]\]

This proposition is true if the sum of *Moby Dick*, *The Brothers Karamazov*, and *Ted and Sally* is identical to the sum of all the books that I read. Since this implies that those three books are the only books I read, the exhaustive listing effect associated with these specificational pseudo-clefts is accounted for.
3.4 Maximality and Exhaustiveness in Questions

After this excursus into the semantics of definite NP’s and free relatives, I will now return to \textit{wh}-questions. In section 1 of this chapter we have seen how we can account for negative island effects in degree questions if we assume that they involve maximality. In this section, I have two objectives. The first is to show that maximality in questions is not restricted to degree questions only, and that \textit{wh}-questions in general involve maximality. In fact, Jacobson (1990) already suggests that her account of maximality in free relatives should be extended to questions. My second objective is to relate my analysis to existing theories of the semantics of questions that can be found in the literature, in particular those of Karttunen (1977) and Groenendijk and Stokhof (1982; 1984). I will take Karttunen’s theory as my starting point, and will then show that once we introduce maximality into Karttunen’s account we end up with a theory that is very close to that of Groenendijk and Stokhof.

3.4.1 Karttunen’s Theory of Questions

Karttunen’s (1977) semantics for questions is based on the idea that a question denotes the set of propositions that are its true answers. Take a simple \textit{wh}-question like (34):
(34) What did John read?

Suppose that we are in a situation in which John read *Moby Dick* and *The Brothers Karamazov* and nothing else. Then, according to Karttunen, (34) denotes the set that has as its members the two propositions 'John read *Moby Dick*’ and 'John read *The Brothers Karamazov*’. In general, the denotation of (34) will be the set of all true propositions of the form 'John read x’ for some x. This denotation is expressed by the formula in (35):\(^5\)

\[
\lambda p \exists x[p(w) \land p = \lambda w[read(w)(j,x)]]
\]

Here p is a variable ranging over propositions of type \(\langle s,t \rangle\) and w is a variable of type s ranging over possible worlds (or indices). When free, w is intended to refer to the actual world. Hence, the subformula p(w) in (35) means that p is true in the actual world. \(\lambda\)-abstracting over w amounts to taking the intension of an expression. The expression \(\lambda w[read(w)(j,x)]\) denotes the set of possible worlds in which John read x, or in other words, the proposition that John read x. So (35) denotes the set of propositions which are of the form 'John read x’ (for some x) and which are true in w, the actual world.

\(^5\) The question *Which book did John read?* will be translated as \(\lambda p \exists x[p(w) \land book(w)(x) \land p = \lambda w[read(w)(j,x)]]\). Note that the translation of the noun *book*, that is, the subformula book(w)(x), is outside the scope of the abstractor \(\lambda w\). As a result the noun is interpreted in a *de re* manner (cf. Groenendijk and Stokhof (1982; 1984)).
3.4.2 Adding Maximality to Karttunen’s Theory

Jacobson (1990) suggests that her account of the semantics of free relatives should be extended to questions. Adding maximality to Karttunen’s theory is rather straightforward, but it has some interesting and far-reaching consequences. We can start by replacing (35) as the translation of the question *What did John read?* by (36):

\[
\lambda p \exists x [p(w) \land p = \lambda w [x = \text{max}(\lambda y [\text{read}(w)(j,y)])]]
\]

This expression denotes the set of all true propositions of the form ‘\(x\) is the maximal individual that John read’. To see what this means, consider again a concrete example. Suppose that in the actual world, John read the following three books and nothing else: *Anna Karenina, Buddenbrooks, and Crime and Punishment*. Then the proposition denoted by the \(\lambda\)-expression in (37a) is true in the actual world (where \(a\), \(b\) and \(c\) are constants denoting the three books just mentioned). An informal paraphrase of this proposition is given in (37b):

\[
\begin{align*}
\text{(37a)} & \quad \lambda w [a+b+c = \text{max}(\lambda y [\text{read}(w)(j,y)])] \\
\text{(37b)} & \quad \text{John read } \textit{Anna Karenina, Buddenbrooks, and Crime and Punishment, and nothing else.}
\end{align*}
\]
Moreover, (37a) is the only proposition of the form $\lambda w[x = \text{max}(\lambda y[\text{read}(w)(j,y)])]$ ('John read x and nothing else') that is true in the actual world. In other words, in the given situation, (36) will denote the singleton set that has (37) as its only member.

In general, by adding maximality to Karttunen’s theory of questions, we end up with a theory in which all questions denote singleton sets which contain exactly one proposition. Now, as Jacobson (1990) points out, because there is a one-to-one correspondence between propositions and singleton sets of propositions, the resulting theory is essentially equivalent to one in which questions denote propositions. Thus, instead of (36) we may as well write (38) as the translation of *What did John read*:

$$
\text{(38)} \quad \text{up} \exists x[p(w) \land p = \lambda w[x = \text{max}(\lambda y[\text{read}(w)(j,y)])]]
$$

A theory in which questions denote propositions has in fact been proposed by Groenendijk and Stokhof (1982; 1984). In the next subsection, I will briefly outline their theory and compare it to the theory we just arrived at by adding maximality to Karttunen’s account.
3.4.3 Groenendijk and Stokhof’s Theory

In Groenendijk and Stokhof’s theory, the question *What did John read?* gets the following translation:

\[
\lambda w'[\lambda x[\text{read}(w)(j,x)] = \lambda x[\text{read}(w')(j,x)]]
\]

\(\lambda x[\text{read}(w)(j,x)]\) denotes the set of all things that John read in \(w\), the actual world, and \(\lambda x[\text{read}(w')(j,x)]\) denotes the set of all things that John read in \(w'\). Therefore, (39) denotes the set of worlds \(w'\) such that the set of things that John read in \(w'\) is equal to the set of things that he read in \(w\), the actual world. In other words, it denotes the set of worlds in which John read the same things that he read in the actual world. Suppose again that in the actual world \(w\) John read *Anna Karenina*, *Buddenbrooks*, and *Crime and Punishment*, and nothing else. Then (39) will denote the set of worlds in which John read *Anna Karenina*, *Buddenbrooks*, and *Crime and Punishment*, and nothing else, or in other words, the proposition that John read those three books, and nothing else. In effect, then, (39) denotes the same proposition as (38). This means that adding maximality to Karttunen’s semantics for questions results in a theory which, for all intents and purposes, is equivalent to Groenendijk and Stokhof’s theory.\(^6\)

---

\(^6\) An important difference between Groenendijk and Stokhof’s theory and the analysis presented here is that in my analysis in a question like *Which book did John read?* the common noun *book* will be interpreted *de re*, whereas in Groenendijk and Stokhof’s analysis it is interpreted *de dicto* (cf. Groenendijk and Stokhof (1982; 1984)). Groenendijk and Stokhof can only derive the *de re* reading by the questionable
3.4.4 Exhaustiveness

One of the main arguments that Groenendijk and Stokhof give for their theory is based on what they call the exhaustiveness of questions. They distinguish two kinds of exhaustiveness, weak exhaustiveness and strong exhaustiveness. Karttunen’s (1977) theory incorporates weak exhaustiveness, but not strong exhaustiveness. Groenendijk and Stokhof’s (1982) theory incorporates both weak and strong exhaustiveness.

Weak exhaustiveness means the following. If Mary knows what John read, then for every \( x \) such that John read \( x \), Mary knows that John read \( x \). Weak exhaustiveness guarantees that inferences like the following obtain:

\[(40)\quad \text{Mary knows what John read.}
\]

\[\text{John read } War and Peace.\]

Therefore: Mary knows that John read \( War and Piece \).

In Karttunen’s theory, weak exhaustiveness is guaranteed by a meaning postulate that says in effect that knowing a question (=a set of propositions) implies knowing every proposition that is an element of it. Recall that in Karttunen’s theory \textit{what John read} denotes the set of true propositions of the form ‘John read \( x \)’. So knowing what John
read implies knowing all true propositions of the form 'John read x'.

In Groenendijk and Stokhof’s theory no such meaning postulate is needed. In a world where John read \( a \), \( b \), and \( c \), and nothing else, the question what John read denotes the proposition that John read \( a \), \( b \), and \( c \), and nothing else. Knowing this proposition implies knowing the proposition that John read \( a \), and knowing the proposition that John read \( b \), and knowing the proposition that John read \( c \) (on the assumption that knowing a proposition implies knowing its entailments).

Now let’s turn to strong exhaustiveness. Strong exhaustiveness means that in addition to weak exhaustiveness the following condition holds: If Mary knows what John read, then, for every \( x \) such that John did not read \( x \), Mary knows that John did not read \( x \). Strong exhaustiveness guarantees that the following inference is valid:

\[
\begin{align*}
&\text{(41) Mary knows what John read.} \\
&\text{John did not read Dead Souls.} \\
&\text{Therefore: Mary knows that John did not read Dead Souls.}
\end{align*}
\]

Strong exhaustiveness is not guaranteed in Karttunen’s theory. If Mary knows what John read, she knows all propositions of the form 'John read \( x \)’, but this does not imply that she knows any propositions of the form 'John did not read \( x \)’. Groenendijk and Stokhof’s theory on the other hand does guarantee strong exhaustiveness. In their
theory, if John read \(a, b,\) and \(c,\) and nothing else, then knowing what John read means knowing the proposition that John read \(a, b,\) and \(c,\) and nothing else. It follows that if John did not read \(d,\) and Mary knows what John read, then she knows that John did not read \(d,\)

I have shown that by adding maximality to Karttunen’s theory, we get a theory that is very close to Groenendijk and Stokhof’s theory. Such a theory has the desirable property of guaranteeing (strong) exhaustiveness.\(^7\) So, in a sense, maximality and exhaustiveness are two sides of the same coin. In the beginning of this chapter, I introduced maximality in degree questions for the sole purpose of explaining the negative island effect. Now we see that postulating maximality is not an \textit{ad hoc} move. Maximality in degree questions is part of a larger pattern of exhaustiveness found in all \textit{wh}-questions.

Finally, let me spell out the interpretation of the degree question \textit{How tall is John?} which in the beginning of this chapter had been given as (42a), using the informal question operator ‘?’ (42a) can now be spelled out as (42b):}

\(^7\) Groenendijk and Stokhof argue that strong exhaustiveness holds for verbs like \textit{know} and \textit{find out}, but not for verbs like \textit{wonder}. The first class of verbs they call extensional, and the second intensional. Intensional verbs take the intension of a question as their argument, that is, a relation between worlds, rather than a proposition (a set of worlds). Berman (1990) has argued against strong exhaustiveness on the basis of what he calls the quantificational variability effect in questions. For alternative analyses of this phenomenon, see Lahiri (1991) and Groenendijk and Stokhof (1992).
(42)  a.  \(?d[d = \text{max}(\lambda d'[\text{tall}(w)(j,d')])])\]

b.  \(\text{up} \exists d[p(w) \land p = \lambda w[d = \text{max}(\lambda d'[\text{tall}(w)(j,d')])])]\]

In the next chapter I will continue using ‘?’ but merely for the purposes of abbreviation:

\(?\nu \Phi\) is short for \(\text{up} \exists \nu[p(w) \land p = \lambda w \Phi]\).
CHAPTER 4
SCOPE AMBIGUITIES IN *HOW MANY*-QUESTIONS

4.1 Introduction

Cinque (1990) and Kroch (1989) have pointed out that questions involving *how many* are potentially ambiguous in scope.¹ For instance, (1) has two readings depending on whether the *wh*-phrase *how many books* has wide or narrow scope with respect to the verb *want*.

(1) How many books does Chris want to buy?
   a. What is the number $n$ such that there are $n$ books that Chris wants to buy?
   b. What is the number $n$ such that Chris wants it to be the case that there are $n$ books that he buys?

¹ An earlier version of the analysis presented in this chapter was presented at the 1993 annual meeting of the LSA in Los Angeles. Completely independently, Cresti (1993) has worked out an analysis that is strikingly similar to my own. We both owe important insights to Frampton (1990; 1991).
The wide scope, or *de re*, reading of the sentence is paraphrased in (1a). Under this reading, it is assumed that there are certain books which Chris wants to buy and the speaker asks how many such books there are. (1b) paraphrases the narrow scope, or *de dicto*, reading of the sentence. On this reading it is not assumed that there are any specific books that Chris wants to buy, but rather that he wants to buy a certain number of books, without having any particular books in mind.\(^2\)

The ambiguity in a question like (1) is reminiscent of the more familiar *de re/de dicto* ambiguity in declarative sentences which contain an indefinite NP, such as (2):

(2) Chris wants to buy five books.
   a. There are five books that Chris wants to buy.
   b. Chris wants it to be the case that there are five books that he buys.

(2a) paraphrases the *de re* or wide scope reading of the indefinite, and (2b) the *de dicto* or narrow scope reading.

\(^2\) Note that in the (b) paraphrase the number *n* is still understood *de re*. As Angelika Kratzer has pointed out to me, there is also reading of (1) in which the number itself is taken *de dicto*, as shown by the possibility of giving (i) as an answer:

(i) Chris wants to buy as many books as you do.
I will not give a complete analysis of this ‘superintensional’ reading in this dissertation, but I will offer a suggestion on how to deal with it in footnote 5.
The parallelism between (1) and (2) might suggest that there is no need for a special explanation of scope ambiguities in *how many*-questions. Doesn’t the mechanism that explains the ambiguity of (2) (quantifier raising, quantifying in, Cooper-storage, or whatever) also automatically account for the ambiguity of (1)? I believe things are not quite that simple, and that we do have to say something extra to account for the ambiguity of *how many* questions. However, I will postpone giving arguments for this position until after I have formulated my explanation for the ambiguity.

*How many*-questions bear an obvious resemblance to the degree question that were studied in chapter 3, and therefore we would expect *how many*-questions to be sensitive to selective islands as well. That this is indeed the case, becomes clear when we consider the behavior of *how many*-phrases that are extracted out of a negative island or a *wh*-island. Take (3) and (4), for instance:

(3) How many books did no student want to buy?
   a. What is the number \(n\) such that there are \(n\) books that no student wants to buy?
   b. * What is the number \(n\) such that no student wants it to be the case that there are \(n\) books that s/he buys?
(4) How many books do you wonder whether Chris wants to buy?
   a. What is the number \( n \) such that there are \( n \) books that you wonder
      whether Chris wants to buy?
   b. * What is the number \( n \) such that you wonder whether Chris wants to buy
      \( n \) books?

As originally noted by Kroch (1989), these sentences are not ambiguous. Unlike (1),
they only allow the wide scope reading, which is paraphrased in (3a) and (4a). (3) can
only be used to ask how many books there are of which it is true that no student wants
to read them. It cannot be interpreted as asking for the number \( n \) such that no student
wants to read \( n \) books (*de dicto*). Similarly for (4).³

The disappearance of the *de dicto* reading only occurs in questions and other
wh-movement constructions. Compare (3) and (4) to the corresponding declarative
with an indefinite NP:

(5) No student wants to buy five books.
(6) I wonder whether Chris wants to buy five books.

These sentences seem to be just as ambiguous as (2). (Perhaps they even show a bias in

³ The superintensional reading mentioned in footnote 2 also disappears. (3), for
instance, cannot be answered with (i):
   (i) No student wants to buy as many books as you do.
favor of the *de dicto* reading. If there is any loss of ambiguity in (5) and (6) as compared to (2), it is the *de re* reading that is weakened, not the *de dicto* reading.)

So what we see here is an interesting interaction between extraction out of a selective island and scope. If a *how many*-phrase is moved out of a selective island, it can only get wide scope. (See section 4.9 for a more precise formulation of this generalization.) Obviously, the disappearance of the narrow scope reading under extraction out of a selective island is related to the fact that selective islands block the extraction of degree expressions altogether. My primary goal in this chapter is to show how scope ambiguities in *how many*-questions like (1) can be accounted for in such a way that the non-ambiguity of (3) and (4) falls out of the analysis of selective island effects presented in this dissertation.

The chapter is organized as follows. 4.2 takes a closer look at the scope ambiguity of *how many*-questions and makes a first step towards formalization. In sections 4.3-4.7, I present an explicit model-theoretic semantics for *how many*-questions. I will demonstrate that the various readings of *how many*-questions can be derived without a syntactic operation of ‘reconstruction’ which lowers the *wh*-phrase (or part of it) back to the position of the trace. I formulate explicit rules which translate S-structure representations into expressions of intensional logic. This interpretation procedure is compositional and in principle the logical language is dispensable. To a certain extent the translation proceeds in a type-driven fashion.
In my account, the semantics of the traces left by *wh*-movement plays a central role. The ambiguity of *how many*-questions hinges on the distinction between two types of traces: those which represent individual variables of type $e$ (‘small traces’) and those which stand for variables of a higher type (‘big traces’). A noteworthy feature of the analysis is that more complicated cases in which the *how many*-phrase can take scope between two scope-bearing elements are derived by taking into account the semantic function of intermediate traces (see section 4.7). Intermediate traces are usually viewed as a purely syntactic device without any semantic import. If my analysis is correct, they play an important role for semantic interpretation as well. However, intermediate traces are interpreted in exactly the same way that traces in argument positions are, and no semantic rules are introduced that apply specifically to intermediate traces.

In section 4.8, I return to the question of selective islands. It will be shown that, if a *how many*-phrase is extracted out of a selective island, it will have wide scope over the island, but may still have narrow scope with respect to scope-bearing elements outside the island. The next question is how to account for this generalization. One approach is to allow the syntax to make reference to the distinction between small traces and big traces. Frampton (1990) and Cresti (1993) are two representatives of this line of research. Their analyses, which are geared toward *wh*-islands rather than negative islands, are discussed in section 4.9. In section 4.10, I will discuss an alternative analysis, one that uses the theory of maximality developed in this dissertation to account for the ‘wide scope only’ interpretation of *how many*-phrases which are
extracted out of negative islands. In this type of analysis there is no need for syntactic rules or principles that make reference to the semantic type of the trace. This approach represents a more restrictive view of the syntax-semantics interface. Whether it can be extended to cover other selective islands (such as \textit{wh}-islands) as well is a question I will postpone until chapter 5.

4.2 A Closer Look at the Scope Ambiguities

Consider again the two readings of sentence (1), repeated here for convenience:

(7) How many books does Chris want to buy?
   a. What is the number $n$ such that there are $n$ books that Chris wants to buy?
   b. What is the number $n$ such that Chris wants it to be the case that there are $n$ books that he buys?

(8) gives two slightly more formal paraphrases of the two readings of this sentence.
(8) a. What is the number \( n \) such that \( n \) is the cardinality of the set \( \{ x \mid x \text{ is a book and Chris wants to buy } x \} \)?

b. What is the number \( n \) such that Chris wants the cardinality of the set \( \{ x \mid x \text{ is a book and you buy } x \} \) to be equal to \( n \)?

The wide scope reading of the sentence asks what the cardinality is of the set of objects \( x \) such that \( x \) is a book and Chris wants to buy \( x \). This paraphrase is given in (8a). (8b) paraphrases the narrow scope reading.

Sentences that are more complex than (1) may have more than just two readings. Sentence (9), for instance, is three ways ambiguous. *How many books can* have wide scope over both the verb *think* and the verb *want* (9a), or it can have narrow scope with respect to *want*, but wide scope with respect to *think* (9b), or it can have narrow scope with respect to both verbs (9c):

(9) How many books do you think Mary wants to buy?

a. What is the number \( n \) such that there are \( n \) books which you think Mary wants to buy?

b. What is the number \( n \) such that you think there are \( n \) books which Mary wants to buy?

c. What is the number \( n \) such that you think that Mary wants it to be the case that there are \( n \) books that she buys?
(10) gives more formal paraphrases of these three readings:

(10)  

a. What is the number $n$ such that $n$ is the cardinality of the set $\{x \mid x$ is a book and you think Mary wants to buy $x\}$?

b. What is the number $n$ such that you think that $n$ is the cardinality of the set $\{x \mid x$ is a book and Mary wants to buy $x\}$?

c. What is the number $n$ such that you think that Mary wants the cardinality of the set $\{x \mid x$ is a book and Mary buys $x\}$ to be equal to $n$?

4.3 The Interpretation of How Many

Before formulating interpretation rules for how many-questions, we should consider the semantics of simple numeral constructions like five books. Krifka (1986; 1989) has proposed to analyze count nouns as relations between entities and numbers. The noun books denotes a relation which holds between an object $x$ and a number $n$ if $x$ consists of $n$ books. The NP five books is then translated as follows:

---

4 In fact, Krifka further decomposes $\text{book}(x,n)$ as $\text{BOOK}(x) \land \text{NU}($BOOK$')(x) = n$, where $\text{BOOK}$ is a function that is true of (atomic or non-atomic) objects consisting of books and $\text{NU}($BOOK$)$ is a function that maps an object onto the number of ‘natural book units’ it consists of (NU stands for ‘natural unit’). The reason behind this analysis is that this allows a unified account of simple numeral constructions like five books and measure constructions such as five pounds of coffee and twenty head of cattle (and also classifier constructions in languages like Chinese).
(11) \[\text{five books} \to \lambda P \exists x [\text{book}(w)(x,5) \land P(w)(x)]\]

Five books are on the table \(\to \exists x [\text{book}(w)(x,5) \land \text{on_the_table}(w)(x)]\)

This means that the sentence *Five books are on the table* is true if there is a plural object consisting of five books that is on the table. Note that the sentence will be true if there are in fact *more* than five books on the table.

*How many books* will be translated in essentially the same way as *five books*, except that the number 5 is replaced by a variable \(n\), ranging over numbers.

(12) \[\text{how many books} \to \lambda P \exists x [\text{book}(w)(x,n) \land P(w)(x)]\]

In a question, the variable \(n\) is bound by the question operator. A simple (non-ambiguous) question like *How many books did John read?* is going to be translated as in (13), completely analogous to the way degree questions were treated in chapter 3:

(13) \[\text{How many books did John read?} \to \]

\[? n[n = \text{max}(\lambda n' \exists x [\text{book}(w)(x,n') \land \text{read}(w)(j,x)])]\]

Note that the presence of the maximality operator is crucial here. If John read exactly five books, it is also true that he read four books (or three, or two, or one). Nevertheless, answering (13) with ‘four’ would be false in that situation. The
maximality operator makes sure that the maximal number of books that John read is selected.

4.4 Deriving the Narrow Scope Reading: Semantic Reconstruction

The wide and narrow scope readings of the ambiguous question (14) are going to be represented as in (a) and (b), respectively:

(14) How many books did Mary believe John read?

a. wide: \( \exists n [n = \max (\lambda n' \exists x [\text{book}(w)(x,n') \land \text{believe}(w)(m,\lambda w'[\text{read}(w')(j,x)])])] \)

b. narrow: \( \exists n [n = \max (\lambda n' [\text{believe}(w)(m,\lambda w'[\exists x [\text{book}(w')(x,n') \land \text{read}(w')(j,x)])])] \)

These representations make clear that the two readings differ in the scope of the existential quantifier that is part of the translation of the how many-phrase. Deriving the wide scope reading does not pose any special problems, because in that reading the

\[ \]

\[ ^{5} \text{I believe the superintensional reading mentioned in footnote 2 can be dealt with by abstracting over a variable of type } <s,n> \text{ (a function from possible worlds to numbers), rather than a variable of type } n \text{ (numbers). Let } N \text{ be a variable of type } <s,n>. \text{ The superintensional reading can then be represented as follows (ignoring maximality):} \]

\[ (i) \quad \exists N [\text{believe}(w)(m,\lambda w'[\exists x [\text{book}(w')(x,N(w')) \land \text{read}(w')(j,x)])]] \]

Under this reading, the question asks for a function from numbers to worlds \( N \) (we might call it a 'numerical concept', in analogy to Montague’s individual concept) such that, in each of Mary’s belief-worlds \( w' \), John read \( N(w') \) books. If the answer is Mary believes that John read as many books as Jim did, then \( N \) would be the function that maps every world \( w \) onto the number of books that Jim read in \( w \).
existential quantifier takes its scope in the surface position of the *how many*-phrase. To account for the narrow scope reading, however, we need a way to give the existential quantifier narrow scope with respect to the verb *read*. One possible way of doing this would be by means of a syntactic operation of reconstruction which moves the *wh*-phrase back to the position of its trace at the level that is relevant for semantic interpretation (say, LF). Although I have no knock-down argument against this kind of analysis, I think it is a valuable methodological principle to let semantic interpretation take place on syntactic structures that are as close to surface structure as possible.

There is a semantic alternative to syntactic reconstruction, namely λ-conversion. All we need to do is translate the *wh*-trace as a variable of the appropriate type, bind it with a λ-operator and then apply the resulting function to the denotation of the *wh*-phrase.

Following Cresti (1993), I will call this 'semantic reconstruction’ in order to emphasize both the similarity and the difference with syntactic reconstruction.

---

6 Syntactic reconstruction as a mechanism to deal with scope ambiguities in *how many*-questions has in fact been proposed by Dobrovie-Sorin (1992) and Heycock (1992).

7 A problem for any theory that assumes syntactic reconstruction is that scope reconstruction does not always pattern together with the reconstruction of binding relations. As Cinque (1991) and Kroch (1989) note, scope reconstruction of a *how many*-phrase is blocked by a selective island, but reconstruction of binding relations is not:

(i) How many books did the editor wonder whether to publish next year *(only wide scope)*

(ii) It’s of herself that I don’t wonder whether she thinks

But more recently Heycock (1992) has drawn attention to an extremely interesting set of cases in which reconstruction for the purposes of binding is affected in a subtle but systematic way by weak islands.
In my analysis a central role is played by the way in which \textsl{wh}-traces are translated as variables. Traces will sometimes be translated as variables of type \textit{e} which range over individuals in the domain of discourse, and sometimes as variables of higher types, in particular the type of NP-intensions, \textit{<s,<<s,\textit{e},t>>,t>>}. To distinguish these two kinds of traces typographically, I will indicate a trace of type \textit{e} by a lower case \textit{t} (referred to as ‘small trace’), but a trace of a type higher than \textit{e} by an upper case \textit{T} (‘big trace’).\footnote{In the examples discussed in this paper the only ‘high’ type we are dealing with is \textit{<s,<<s,\textit{e},t>>,t>>}, the type of generalized quantifier intensions. However, in a more inclusive fragment big traces could also stand for variables of other high types such as \textit{<s,\textit{e},t>>}, the type of predicative NP’s (Partee (1987)) and possibly intensional objects (Zimmermann (1992)).} I will assume that the type of the trace does not have to be the same as that of its antecedent. So a \textsl{wh}-phrase which itself denotes a generalized quantifier, is free to leave either a big trace or a small trace. This ability to leave either type of trace is what brings about the scope ambiguity. For instance, the two readings of the embedded question \textit{how many books John needs} will be represented as in (15) (\textit{e} is the phonetically empty head C which is coindexed with its Specifier \textit{how many books}, under Spec-Head agreement):

\begin{itemize}
\item a. \textit{[CP [NP how many books], [C [e] [IP John needs \textit{t}]]]} (\textit{wide scope})
\item b. \textit{[CP [NP how many books], [C [e] [IP John needs \textit{T}]]]} (\textit{narrow scope})
\end{itemize}

(15) \ldots how many books John needs

\begin{itemize}
\item a. \textit{[CP [NP how many books], [C [e] [IP John needs \textit{t}]]]} (\textit{wide scope})
\item b. \textit{[CP [NP how many books], [C [e] [IP John needs \textit{T}]]]} (\textit{narrow scope})
\end{itemize}

On the wide scope reading, represented in (15a) the object of \textit{need} (\textit{i.e.} the trace \textit{t}) is
an individual variable, and the trace is therefore a small one. On the narrow scope reading given in (15b), John bears the need-relation to some higher type intensional object, and hence we have a big trace $T$. The presence of this big trace, which is going to be translated as a variable of a higher type, makes it possible for us to give the wh-phrase narrow scope without making appeal to a syntactic reconstruction rule.

An important question is what the possible types of a trace are and what determines the type of a given trace. Without suggesting that I have anything like a worked-out theory in this area, I would like to put forward the hypothesis that in principle the type of the trace is completely free, but that there are two semantic factors that restrain it. First of all, the trace must be able to combine semantically with its syntactic sister(s) by the normal semantic combination rules, which include functional application, type shifting, function composition, etc. Second, the trace must end up bound in the final, top-level representation and the result must be semantically well-formed (that is, make sense). I suspect that this will be sufficient to rule out any unwanted type assignments.
4.5 Spelling Out the Analysis

I will now state the semantic rules that are central to my analysis. First the translation rule for traces:

(16) Trace Translation Rule
a. $[\text{NP } t_i] \rightarrow x_i$ (where $x_i$ is of type $e$)
b. $[\text{NP } T_i] \rightarrow X_i(w)$ (where $X_i$ is of type $<s,<<s,<e,t>,t>,t>$)

Small traces are translated as variables of type $e$ and big traces as variables of the type of generalized quantifier intensions. In the logical translation language I will also use the convention that lower case variables are of type $e$ and upper case variables of type $<s,<<s,<e,t>,t>,t>$. It is important to realize that, since I am assuming that $wh$-phrases are free to leave either a big trace or a small trace, I might as well have used just one kind of trace in the syntax which could then be translated as either a variable of type $e$ or as a variable of type $<s,<<s,<e,t>,t>,t>$. I will keep distinguishing between small traces and big traces in the syntax, however, for two reasons. The first reason is convenience. Keeping big and small traces typographically distinct allows me to make the semantic function of traces of different types perspicuous in the syntactic representations. The second reason is a more principled one. There might be syntactic rules or principles that
are sensitive to the distinction between the two kinds of traces. In section 4.9, I will discuss two candidates for syntactic rules of this sort which have been proposed by Frampton (1990) and Cresti (1993) to account for the fact that a *how many*-phrase that is extracted out of a selective island necessarily has wide scope over the island. Whether we actually need syntactic rules of this kind is a question which will be discussed in section 4.10.

The next rule we need is one that translates $\overline{C}$. It embodies the idea that the semantic function of the (usually empty) head $C$ is to act as a $\lambda$-operator, exploiting the fact that $C$ is coindexed with the XP in the Spec of CP. This $\lambda$-operator binds the variable corresponding to the trace left by this XP. If this trace is a small one, then the $\lambda$-operator abstracts over a variable of type $e$; if the trace is big, then the abstraction is over a variable of type $<s,<<s,<e,t>>,t>$. IP' is the translation of IP:

\[ \text{IP'} \]

---

9 It might be objected that the $\overline{C}$ translation rule as stated here is not strictly compositional, because it has to look inside the syntactic representation of the IP to find the type of the trace. There are two possible solutions to this problem. One possibility is to introduce a special mechanism in the syntax which indexes each moved phrase with the type of the trace it left behind when it was moved. Another, perhaps more attractive option is to leave the type of the variable bound by the $\lambda$-abstractor unspecified. Any type mismatches between the lambda operator and the variable it is supposed to bind must then be filtered out as semantically or pragmatically deviant because of the presence of ‘left-over’ variables which did not get bound in the course of the derivation. This latter option would require us to make sure that in cases of type mismatch between the variable and its binder, the variable cannot somehow get bound by accident later in the derivation, producing a pragmatically acceptable (but wrong) interpretation of the sentence. I leave this issue for further research.
(17) \( \overline{\text{C Translation Rule}} \)

a. \([e \text{ C} \text{ IP}] \rightarrow \lambda x \text{IP}'\) if \(\text{C}_i\) locally binds a small trace \(t_i\) in IP.

b. \([e \text{ C} \text{ IP}] \rightarrow \lambda X \text{IP}'\) if \(\text{C}_i\) locally binds a big trace \(T_i\) in IP.

We next move up to the CP level. The phrase in the Spec of CP will either act as the functor that takes the \(\overline{\text{C}}\) as its argument or vice versa, depending on the types of the two expressions. In the formulation of the rule we can simply allow for both options, making the translation procedure type-driven in the sense of Klein and Sag (1985).

(18) \( \text{CP Translation Rule} \)

\([_{\text{CP}} \text{XP} \overline{\text{C}}] \rightarrow \text{XP}'(\lambda w \overline{\text{C}}')\) or \(\overline{\text{C}}'(\lambda w \text{XP}')\)

Finally, we need a rule that binds the number-variable in the translation of *how many*. This rule applies on the CP level, after the CP translation rule has applied. To make sure that this rule binds the right variable, we have to take some care in the way the indexing of the variable is set up. We have to distinguish the index on the determiner *how many* in (12) from the index on the wh-phrase *how many books* as a whole. This is shown explicitly in the tree representation in (19).
We would probably need the same distinction between wh-index and NP-index to deal with possessive wh-phrases like whose book.

To distinguish the two indices, I will refer to the index on the NP as a whole as the ‘NP-index’ and to the index on the determiner how many as the ‘wh-index’. The motivation for this is that the wh-index can be thought of as corresponding to the wh-feature on the determiner how many. In (19a) the wh-index is 1 and the NP-index is 2. I will adopt the convention of writing the wh-index as a subscript on how many and the NP-index as a subscript on the noun, as indicated in (19b).  

```
(19)  a.             NP₂      b. how many₁ books₂
                      Det₁        N₂
                       how many₁ books₂
```

The question-rule that binds the number-variable can now be formulated as follows:

(20) **Question Rule**

\[
\text{CP} \rightarrow \text{?n}[n = \text{max}(\lambda n, \text{CP}')]]
\]

where \(i\) is the wh-index of the constituent in the Spec of CP.

Note that the Question rule needs to know what the wh-index of the constituent in the Spec of CP is in order to bind the right variable. This is where the correspondence between the wh-index and the wh-feature that I mentioned above becomes important. I

---

10 We would probably need the same distinction between wh-index and NP-index to deal with possessive wh-phrases like whose book.
will assume that the wh-feature of the determiner, being a foot feature in the sense of Gazdar et al. (1985), must percolate up to the NP level, characterizing the whole NP as a wh-phrase. By saying that the wh-index corresponds to the wh-feature I mean that the wh-index percolates up together with the wh-feature. We can think of the wh-index as part of the wh-feature. I will furthermore assume that the wh-feature, and hence also the wh-index, are ‘visible’ on the CP-level, the level at which the Question rule applies. Note by the way that identifying the wh-index with the wh-feature gives the right result for wh-phrases like how many people’s parents. This whole NP is a wh-phrase because of the wh-feature which percolates up from how many. As a result, the free variable that gets bound by the question rule will be the number variable in the translation of how many.
4.6 Deriving the Wide and Narrow Scope Readings

The derivation of the wide scope reading of (15) is given in (21):

(21) ... how many, books, John needs t,? \textit{(wide scope)}

Some comments are in order here. On the wide scope (or \textit{de re}) reading of the sentence the \textit{wh}-trace is ‘small’ and therefore translated as a variable of type \(e\). Since the verb \textit{needs} requires a NP of type \(<<s,<e,t>>,t>\) as its argument, the interpretation of the trace has to be type-raised. I will assume the following rule that raises NP’s from type \(e\) to type \(<<s,<e,t>>,t>\). Presumably, this rule follows from more general type raising principles (Partee and Rooth (1982), Partee 1987).
Under Spec-Head agreement, the empty head C is coindexed with the \textit{wh}-phrase in the Spec of CP. Because $t_2$, the trace locally bound by C, is a small trace, clause (a) of the $\overline{C}$ translation rule (17) applies. The translation of the CP is obtained in two steps. First, the CP translation rule (18) is applied. Because of type driven translation, the NP \textit{how many books} which sits in the Spec of CP, takes the $\overline{C}$ as its argument. The resulting translation of the CP has a free variable $n$, which stands for the number of books. In the second step, this free variable gets bound by the question rule given in (20) (this rule can be regarded as a special kind of type raising).

The narrow scope reading of (15) is derived as follows:
(23) ... how many, books₂ John needs T₂?

\[
\text{CP: } \lambda n \, [n = \text{max}(\lambda n_{1}[\text{need}(w)(j, \lambda w \lambda P \exists y[\text{book}(w)(y, n_{1}) \land P(w)(y)])])]
\]

\[
\text{CP: } \lambda X_{2}[\text{need}(w)(j, X_{2})](\lambda w \lambda P \exists y[\text{book}(w)(y, n_{1}) \land P(w)(y)])
\]

\[
= \text{need}(w)(j, \lambda w \lambda P \exists y[\text{book}(w)(y, n_{1}) \land P(w)(y)])
\]

\[
\text{NP: } \lambda P \exists y[\text{book}(w)(y) \land P(w)(y)]
\]

On the narrow scope reading of the sentence, \textit{how many books} leaves a big trace. In accordance with the trace interpretation rule (16), this big trace is translated as a variable ranging over generalized quantifier intensions. Since the object of \textit{need} has the right type now, type-raising is unnecessary. This time, the empty head C locally binds the big trace T₂, and hence clause (b) of the C’ translation rule applies. The interpretation of the CP is again derived in two steps. This time, the type-driven translation procedure makes the NP \textit{how many books} the argument of the C’. By \(\lambda\)-conversion the intension of \textit{how many books} ends up as the argument of the predicate \textit{need}. Subsequently, the question rule (20) applies and the free variable \(n_{1}\) gets bound by the ?-operator.
4.7 The Role of Intermediate Traces

So far, we have considered the two readings of a simple sentence involving only one intensional verb. More interesting cases arise if we make the sentence more complex. Consider for instance (24) which has three readings, a wide scope reading (a), an intermediate scope reading (b), and a narrow scope reading (c):

(24) How many books did Mary say John needs?
   a. What is the number $n$ such that there $n$ books which Mary says John needs?
   b. What is the number $n$ such that Mary says there are $n$ books which John needs?
   c. What is the number $n$ such that Mary says John needs $n$ books?

We can account for these three readings if we take seriously the role played by the intermediate trace left in the Spec of CP of the embedded clause. Just like traces in argument positions, intermediate traces can be either small or big. This gives us the following four logically possible syntactic representation of the sentence:
(25)  

a. How many books did Mary say \textit{t} John needs \textit{t}?

b. How many books did Mary say \textit{t} John needs \textbf{T}?

c. How many books did Mary say \textbf{T} John needs \textit{t}?

d. How many books did Mary say \textbf{T} John needs \textbf{T}?

With these four syntactic representations, the translation and type raising rules we already have are sufficient to derive the three readings of the sentence that are distinguished in (24). (25a) and (b) both result in the wide scope reading, (c) represents the intermediate scope reading, and (d) gives the narrow scope reading. The four derivations are given in (26) below (the question is given in its embedded form in the derivations). It should be noted that in the two derivations of the wide scope reading in (26a) and (b) the intermediate trace in the Spec of CP undergoes type raising, which is forced by the type driven translation procedure of the CP translation rule.
(26)

a. ... how many books Mary says John needs (wide scope)

b. ... how many books Mary says John needs (wide scope)
c. ... how many books Mary says John needs (intermediate scope)

\[
\begin{align*}
\text{CP: } & n = \max(\lambda n_1, [\text{say}(w)(m, \lambda w \exists y [\text{book}(w)(y, n_1) \land \text{need}_s(w)(j, y)])]) \\
\text{CP: } & \lambda X_2[\text{say}(w)(m, \lambda w [X_2(w)(\lambda w \lambda x_1[\text{need}_s(w)(j, x_1)])])(\lambda w \lambda P \exists y [\text{book}(w)(y, n_1) \land P(w)(y)])] \\
& = \text{say}(w)(m, \lambda w \exists y [\text{book}(w)(y, n_1) \land \text{need}_s(w)(j, y)]) \\
\text{NP}_2 & \quad \overline{\lambda} X_2[\text{say}(w)(m, \lambda w [X_2(w)(\lambda w \lambda x_1[\text{need}_s(w)(j, x_1)])])] \\
\text{Det} & \quad N_2 \quad C \quad \text{IP: say}(w)(m, \lambda w [X_2(w)(\lambda w \lambda x_1[\text{need}_s(w)(j, x_1)])]) \\
\text{how many books} & \quad e_2 \quad \text{NP: m} \quad \text{VP: say}(w)(\lambda w [X_2(w)(\lambda w \lambda x_1[\text{need}_s(w)(j, x_1)])]) \\
\text{Mary} & \quad \text{V: say} \quad \text{CP: } X_2(w)(\lambda w \lambda x_1[\text{need}_s(w)(j, x_1)]) \\
& \quad \text{says} \quad \text{NP: } X_2(w) \quad \overline{\lambda} X_2[\text{need}_s(w)(j, x_1)] \\
T_2 & \quad \text{John needs } t_2
\end{align*}
\]

188
It is worthwhile to stress once more that no other syntactic or semantic devices are necessary to obtain the three readings of this sentence than those that were introduced earlier in order to account for the two readings of (15). Intermediate traces are interpreted in exactly the same way that traces in argument positions are. It is obvious, moreover, that the analysis will work equally well in more complex sentences that have more than three potential sites where the how many-phrase can take scope. In any given sentence the scope of the how many-phrase is determined by the distribution of big and small traces in the chain it heads. In general, given a chain (how many \(N, \tau_1, \ldots, \tau_n\)), where for each \(i \leq i \leq n\) \(\tau_i\) is either a small trace \(t\) or a big trace \(T\) coindexed with how many \(N\), the wh-phrase will take scope in the position of the big trace \(\tau_m\) such that: (i) for all \(i \leq i \leq m\) \(\tau_i\) is a big trace; (ii) \(\tau_{m+1}\) is a small trace. Thus in a chain like (27), how many \(N\) has scope in the position of the trace indicated by the star:

(27) \(\text{how many } N \ldots T \ldots T \ldots T^{*} \ldots t \ldots T \ldots t \ldots t \ldots\)

Note that all traces c-commanded by the leftmost small trace in the chain are irrelevant for the determination of scope, in the sense that it doesn’t make a difference for the truth conditions of the sentence whether these traces are big or small; cf. the derivations of the equivalent (25a) and (b).
Because of the importance of intermediate traces in determining the scope of wh-phrases, the analysis proposed here has some important consequences for their theoretical status. In Government and Binding Theory, the existence of intermediate traces has always been motivated on purely syntactic grounds, such as locality constraints on movement. It has even been assumed that intermediate traces can be deleted at the level of Logical Form because they supposedly lack ‘semantic content’ (Lasnik and Saito (1992)). If my analysis is correct, however, intermediate traces play a crucial role in semantic interpretation. For one thing, this means that intermediate traces cannot arbitrarily be deleted without affecting the range of possible interpretations of the sentence. A further implication, and a potentially more interesting one, is that, on the basis of the present analysis, we can start looking for empirical evidence for the presence or absence of intermediate traces in certain constructions. In the remainder of this chapter, I will turn to one potential source of such evidence: the scope of how many-phrases under extraction out of selective islands.

To conclude this section, let me point out that the analysis doesn’t use any special techniques to derive the different readings of how many-questions. All the semantic techniques used in this chapter are perfectly standard (the translation of traces as variables, λ-abstraction and conversion, type raising, type driven translation, etc). The only minor innovation may be the notion that traces can be of more than one possible type, but even that is not a very revolutionary idea in a general framework that allows a flexible relationship between syntactic categories and semantic types, in the
spirit of Rooth and Partee (1982) and Partee (1987). Therefore, the theoretical ‘cost’ of my analysis is very low.

### 4.8 Extraction Out of Selective Islands

Kroch (1989) has shown that, when extracted out of a selective island such as a *wh*-island or a negative island, *how many*-phrases lose their narrow scope reading (Cinque (1990)):

(28) a. How many books did the editor decide to publish next year?  
*(ambiguous)*

b. How many books did the editor wonder whether to publish next year?  
*only wide scope*

c. How many books did no editor want to publish next year?  
*only wide scope*

It is important to observe that a *how many*-phrase that is extracted out of a selective island does not necessarily get maximal scope. It can get wide scope with respect to the selective island, but narrow scope with respect to another element which itself has scope over the island. An example of this is (29):
(29) How many books did you tell him that the editor wonders whether to publish next year?

In contrast to (28b) and (c), (29) is still ambiguous. *How many books* can have either wide or narrow scope with respect to the verb *tell*, although it can only have wide scope with respect to *wonder*.

Based on this observation, we can formulate the following empirical generalization:

(30) A *how many*-phrase extracted out of a selective island has wide scope over the island.

The question is how to account for this generalization. In the next section I will explore approaches suggested by Frampton (1990; 1991) and Cresti (1993) which are based on the idea that there are certain syntactic rules or principles which require traces in certain positions to be of type *e.* In section 4.9, I will discuss an alternative approach based on the maximality account of negative islands developed in chapters 2 and 3.
4.9 Syntactic Approaches

In response to Rizzi’s and Cinque’s work, a very interesting suggestion has been made by Frampton (1991). Frampton argues that what is crucial in the examples cited above is the semantic type of the wh-trace inside the selective island. In (28b) and (c) the trace inside the selective island behaves as an individual variable. In terms of the analysis proposed in this chapter, this means that extraction out of a selective island is possible only if the trace inside the island is a small one. Frampton’s suggestion, then, is captured by the following filter:\(^{11}\)

\[(31) \text{ The following structure is ruled out:} \]
\[ ... \alpha_i \ldots [\ YP \ldots \ T_i \ldots \ ] \ldots \]
\[\text{where YP is a selective island and } \alpha_i \text{ locally binds } T_i.\]

According to this filter, the options for the distribution of traces in (28a) and (b) are as follows:

\(^{11}\) Kiss (to appear) proposes another filter, which bears an obvious formal similarity to (31):

(i) Specificity Filter:
If \(\text{Op}_i\) is an operator which has scope over \(\text{Op}_j\) and binds a variable in the scope of \(\text{OP}_j\), then \(\text{OP}_j\) must be specific.

The notion of specificity argued for by Kiss is an extension of that of Enç (1991). The main difference between Kiss’s Specificity Filter and (31) is that the Specificity Filter is a condition on the \text{operator} (the extracted \text{wh-phrase}) whereas (31) is a condition on the \text{trace}. 193
(32) a. How many books, did the editor decide $t/T_1$ to publish $t/T_1$ next year.
    b. How many books, did the editor wonder whether to publish $t/(*T_1$, next year.

It should be clear that, in conjunction with the interpretation procedure for traces proposed in this chapter, (31) makes the correct predictions concerning the scope possibilities of the *how many*-phrase.

Frampton provides a very suggestive piece of evidence for his proposal. Heim (1987) has argued that the subject position in an existential *there*-sentence cannot be filled by an individual variable. This can be shown most easily with definite pronouns:

(33) a. Every suspect denied he was at the scene of the crime.
    b. * Every suspect denied there was he/him at the scene of the crime.

Since small traces represent individual variables, this means that small traces cannot fill the subject position of a *there*-sentence. The prediction, then, is that in cases of extraction out of this position only the narrow scope reading is available. This is borne out by the contrast between (34a) and (b):

194
Note by the way that it would not be sufficient to state the prohibition against individual variables in the subject position of there-sentences as a syntactic restriction against small traces. This is because an intermediate trace that is small would still result in the wide scope reading, even if the trace in the subject position is big (cf. the derivation in (22b)):

(34)  

<table>
<thead>
<tr>
<th></th>
<th>How many police officers did they claim [IP T/t were at the scene of the crime]? (ambiguous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>How many police officers did they claim [IP there were T/*t at the scene of the crime]? (only narrow scope)</td>
</tr>
</tbody>
</table>

The presence of there in the embedded clause in (34) forces the narrow scope reading.\(^{12}\)

Another prediction is that extraction out of a weak island from the subject position of a there-sentence should be bad. This is also confirmed by the data:\(^{13}\)

(35)  

* How many police officers did they wonder whether [IP there were T/t at the scene of the crime]?

\(^{12}\) Note by the way that it would not be sufficient to state the prohibition against individual variables in the subject position of there-sentences as a syntactic restriction against small traces. This is because an intermediate trace that is small would still result in the wide scope reading, even if the trace in the subject position is big (cf. the derivation in (22b)):

(i)  

How many police officers did they claim [CP \( t \) that [IP there were T at the scene of the crime]]?

There are independent arguments to assume that the prohibition against individual variables in the subject position of there-sentences must be semantic rather than syntactic, since it is just a special case of the general prohibition against strong quantifiers in that position (Barwise and Cooper (1981)), and the latter constraint is clearly of a semantic nature.

\(^{13}\) The same observation is made by Kiss (to appear).
Because of the presence of *there*, the trace in (35) can’t be big; because of the weak island it can’t be small. As a result, the sentence is ungrammatical.

Of course the filter (31) is still not more than a descriptive generalization. There are various ways in which (31) could be incorporated in recent theories of locality within the Government and Binding framework. Frampton (1990) has suggested one approach, based on Rizzi (1990) (see also chapter 1). Frampton proposes the following reformulation of the Empty Category Principle:

(36)  **ECP** (Frampton (1990)):

A trace must either be identified as an individual variable, or be chain identified.

By an ‘individual’ variable Frampton means a variable of type $e$. A trace is chain identified if it is joined to the head of its chain by a sequence of links which each satisfy antecedent government. So what (36) comes down to is that a big trace (a trace that is not of type $e$) must bear a more local relation, namely antecedent government, to its antecedent than a small trace. The antecedent government requirement on big traces makes it impossible to skip a *wh*-island, because of relativized minimality (cf. Rizzi (1990) and chapter 1).¹⁴

¹⁴ Note that Frampton’s notion of chain identification is stronger than necessary in that it requires each link of the chain between a big trace and the head of the chain to satisfy antecedent government. As I have shown above, any occurrence of an intermediate small trace has the automatic effect of turning all subsequent traces into small ones. We therefore only need to stipulate that a big trace must be antecedent...
A different syntactic proposal comes from Cresti (1993). Cresti argues that a
*wh*-phrase that escapes from a *wh*-island must pass through a special position, referred
to as δ. This position is adjoined to CP, and it has the special semantic property that it
can only be filled by expressions of type e. This constraint is expressed by the filter in
(37):

\[(37) \quad \ast \left[ \left[ \delta \, X \right] \left[ CP \ldots \right] \right] \quad \text{where } X \text{ is not of type } e.\]

Cresti gives evidence for the existence of such a position from Greek, based on work by

Of course various other candidates for the syntactic principle behind (31) could
be considered, depending on one’s preferred theory of locality. All syntactic
approaches, including Frampton’s and Cresti’s, have one fundamental property in
common, namely that they presuppose that the distinction between small and big traces
is not only a semantic distinction, but also a syntactic one. Since there is no other
reason why big and small traces should be distinguished syntactically (apart from
typographical convenience), adopting a principle like (36) or (37) amounts to opting for
a more powerful theory, namely one that allows syntactic rules and principles to refer to
the distinction between small and big traces. This leads to the question whether a
syntactic principle like (36) or (37) cannot be dispensed with altogether in favor of a

governed by its immediate predecessor in a chain.
semantic theory of selective islands. This issue will be addressed in the next section and in chapter 5.

4.10 A Semantic Approach Based on Maximalitivity

One problem for syntactic approaches like those of Frampton (1990) and Cresti (1993) is that they are geared specifically towards wh-islands and have trouble dealing with negative islands. Yet, it is clear that negative islands force the how many-phrase to have wide scope just like wh-islands do. Consider the following examples:

(38)  
\begin{align*}  
\text{a.} & \quad \text{How many books was John able to read?} \\
\text{b.} & \quad \text{How many books was John not able to read?} 
\end{align*}

The (a) sentence is ambiguous between a wide scope reading and a narrow scope reading, but the presence of the negation in the (b) sentence blocks the narrow scope reading. The negative island effect is not restricted to negation; it can be observed with other downward entailing elements as well:

(39)  
\begin{align*}  
\text{a.} & \quad \text{How many books did nobody read?} \\
\text{b.} & \quad \text{How many books did few people read?} 
\end{align*}
These sentences are both unambiguous: they only allow a reading in which *how many books* has wide scope. Contrast this with the following ambiguous examples:\textsuperscript{15}

(40) a. How many books did everybody read?
    b. How many books did many people read?

Frampton (1990) and Cresti (1993) do not extend their analysis to the negative island cases. Rizzi (1990) faces essentially the same problem in his theory. Recall that Rizzi tries to account for these contrasts by stipulating that *not* is an A’-specifier and therefore prevents the *wh*-phrase from antecedent-governing its trace, under his principle of Relativized Minimality. He furthermore assumes that monotone decreasing quantifiers such as *nobody* and *few people* move to an A’-specifier position (the specifier of CP) at LF, and thereby block antecedent-government of the trace at that level. In chapter 1, I have given strong arguments against such a proposal.

These considerations are enough of an incentive to consider semantic alternatives to Frampton and Cresti’s accounts. I will argue in this section that to

\textsuperscript{15} In fact, (44) seems to be three ways ambiguous. Besides the wide scope and the narrow scope readings, it has a pair-list reading. The three readings can be paraphrased as follows:

(i) \textit{wide scope}: what is the number $n$ such that there are $n$ books $x$ such that for every person $y$, $y$ read $x$?
(ii) \textit{narrow scope}: what is the number $n$ such that for every person $y$ there are $n$ books $x$ such that $y$ read $x$?
(iii) \textit{pair list}: for every person $y$, what is the number $n$ such that there are $n$ books $x$ such that $y$ read $x$?

See Rullmann and de Swart (1992) for some discussions of these three readings.
account for the obligatory wide scope readings for how many-phrases which are
extracted out of selective islands we don’t need any syntactic principles or rules that
refer to the semantic type of the trace, but that this phenomenon can be accounted for
by the theory of selective island effects I developed in chapters 2 and 3. Consider again
the contrast in (38), repeated here as (41):

(41)  a. How many books was John able to read?
    b. How many books was John not able to read?

(42) and (43) give simplified representations of the wide and narrow scope
interpretations of both sentences:

(42)  Wide and narrow scope readings of (41a):
   a. \(?n[n = max(\lambda n' \exists x[book(w)(x,n') \land able_to_read.(w)(j,x)])]
   b. \(?n[n = max(\lambda n'[able_to_read(w)(j,\lambda w \exists x[book(w)(x,n') \land P(w)(x)])]

(43)  Wide and narrow scope readings of (41b):
   a. \(?n[n = max(\lambda n' \exists x[book(w)(x,n') \land \neg able_to_read.(w)(j,x)])]
   b. \(?n[n = max(\lambda n' [\neg able_to_read(w)(j,\lambda w \exists x[book(w)(x,n') \land P(w)(x)])]

The problematic (absent) reading is the narrow scope reading of (41) represented in
(43b). Why is this reading excluded? I want to argue that this reading is excluded for
the same reason that extraction of a degree phrase out of a negative island is blocked: namely, because of the maximality requirement associated with questions. If John is not able to read \( n \) books (for any number \( n \)) then he won’t be able to read more than \( n \) books, either. Thus, there is no maximal number \( n \) such that John is not able to read \( n \) books. Hence, (43b) is not interpretable.

What about the wide scope reading, which is represented in (43a)? Here we need the maximal number \( n \) such that there are \( n \) books which John is not able to read. Since in a given situation the number of books is finite, there is guaranteed to be a maximal \( n \). It should be noted though, that (43b) does require a situation in which there is a salient set of books. *How many books* must be D-linked, in the sense of Pesetsky (1987). The relevance of this notion will be further discussed in chapter 5.

The question is whether the maximality account can be extended to the other selective island cases which are the main focus of the syntactic theories. If a maximality theory of selective islands in general turns out to be feasible, this will have an important consequence for the theoretical status of the distinction between big and small traces. Unlike a syntactic approach, a purely semantic/pragmatic account doesn’t need to assume that the big/small distinction is in any way reflected in the syntax and thus has no need for a syntactic rule or principle that is sensitive to the distinction between big and small traces. Thus, in the syntax we could have just one kind of trace, which would translate as a variable of either type \( e \) or \( \langle s, \langle s, \langle e, t, t \rangle, t, t, t \rangle \rangle \). If there is only one kind
of trace in the syntax, there could be no syntactic rule or principle that is sensitive to the
semantic type of the trace. This hypothesis leads to a more restrictive and therefore less
powerful theory of the syntax/semantics interface. This kind of theory is preferable on a
priori grounds. Whether it is indeed possible is one of the topics of the next chapter.
5.1 Introduction

In the three preceding chapters I have shown that the notion of maximality is central to the semantics of \textit{wh}-constructions. I have also argued that maximality makes it possible to account for the negative island effect in a straightforward way, on the assumption that in downward entailing contexts maximal degrees are not defined. The empirical coverage of this account, however, has been limited in at least three respects:

- The analysis has focused on \textit{wh}-expressions that denote degrees or numbers, but adjuncts such as \textit{why} and \textit{how}, which are also blocked by negative islands, have not been discussed;

- It has been shown that negative islands can be explained in terms of maximality, but the question as to whether this account can be extended to other types
of selective islands, such as *wh*-islands, has been left open;

- The discussion has been limited to *wh*-constructions, and the possibility that other constructions might also show maximality effects has not been addressed.

The goal of this chapter is to expand the coverage of the account in all three areas. This chapter is more speculative than the preceding three chapters, and the conclusions that will be reached have a rather tentative character.

In section 5.2, I will discuss how the maximality account of negative island effects can be extended to cover the behavior of adjuncts such as *how* and *why*. The explanation will make essential use of the notion of Discourse-linking, proposed by Pesetsky (1987). The same notion can also explain certain prima facie counterexamples against negative islands.

Section 5.3 is concerned with the question whether the maximality account can be extended to other types of selective islands (such as *wh*-islands) and whether a unified account can be given of all selective islands, including negative islands. I will argue that this may not be the case, and that the other selective islands may have to receive a syntactic explanation, along the lines proposed by Frampton (1990) and Cresti (1993).
In section 5.4, finally, I show that maximality effects do not only show up in *wh*-movement constructions, but can also be induced by focus. This is of theoretical importance since it shows that negative island effects really aren’t island effects at all. It is possible to get the same effect in a construction without *wh*-movement whose semantics involves maximality.

5.2 Adjuncts

5.2.1 *How and Why Versus When and Where*

In the preceding chapters I have discussed the sensitivity to negative islands of one type of expressions: *wh*-phrases denoting degrees (like *how tall*) or numbers (like *how many books*). It is well-known from the literature, however, that the class of expressions that are sensitive to negative islands is wider. It also includes certain kinds of adjuncts, such as *how* and *why*, as is demonstrated by the following examples:

(1) a. I wonder how Jill/everybody/most people fixed the car.
   b. * I wonder how Jill didn’t fix the car.
   c. * I wonder how nobody/few people fixed the car.
(2)  a. I wonder why Jill/everybody/most people came to the party.

   b. # I wonder why Jill didn’t come to the party.

   c. # I wonder why nobody/few people came to the party.

(2b) and (c) are grammatical, but not under the intended reading. The reading under which they are grammatical is one in which *why* has scope over the negation or downward entailing quantifier. This reading can be paraphrased as ‘I wonder what the reason is that Jill did not come to the party,’ and ‘I wonder what the reason is that nobody/few people came to the party.’ In the intended, but ungrammatical, reading of (2b) and (c) *why* has narrow scope. This reading can be given the somewhat cumbersome paraphrase ‘I wonder what reason *x* is such that Jill did not come to the party for *x*,’ and ‘I wonder what reason *x* is such that Jill did not come to the party for *x*.’ The fact that *why* cannot be extracted across negation or any other downward entailing quantifier can also be demonstrated with examples such as these:

(3)  a. I wonder why, Jill/everybody/most people think Anna was arrested *t*.  

   b. * I wonder why, Jill didn’t think Anna was arrested *t*.

   c. * I wonder why, nobody/few people think Anna was arrested *t*.

In these sentences, *why* has been extracted out of the embedded clause as indicated by the position of the trace. (3b) and (c) are ungrammatical, because negation or a downward entailing quantifiers intervenes between *why* and its trace.
The ungrammaticality of such examples cannot be ascribed to the status of \textit{why} and \textit{how} as adjuncts. For one thing, as was pointed out by Rizzi (1990), extraction of \textit{how} across a negative island is ungrammatical even if it is not a adjunct but an argument, as with the verb \textit{behave}:

\begin{enumerate}
\item I wonder how Jill/everybody/most people behaved.
\item * I wonder how Jill didn’t behave.
\item * I wonder how nobody/few people behaved.
\end{enumerate}

Moreover, it appears that not all adjuncts are alike. \textit{Where} and \textit{when}, for instance, are much less sensitive to negative islands than \textit{how} and \textit{why}:

\begin{enumerate}
\item I wonder how Judy played with her dog.
\item * I wonder how Judy didn’t play with her dog.
\item I wonder why Judy played with her dog.
\item # I wonder why Judy didn’t play with her dog.
\item I wonder where Judy played with her dog.
\item I wonder where Judy didn’t play with her dog.
\end{enumerate}
(8)   a.  I wonder when Judy played with her dog.
     b.  I wonder when Judy didn’t play with her dog.

(5b) and (6b) are ungrammatical (under the intended reading), but (7b) and (8b) are much better. Why this difference?

5.2.2 D-Linking in Negative Questions

It is important to note that although (7b) and (8b) are grammatical, they do require a special discourse context, namely one which provides a salient set of places or times at which Judy could have played with her dog. That is, these sentences are only acceptable if the wh-phrase is ‘Discourse-linked’ (or ‘D-linked’ for short) in the sense of Pesetsky (1987). A wh-phrase is D-linked if it quantifies over a set that is preestablished in the discourse context. Wh-phrases differ in the extent to which they require or allow for D-linking. Which-phrases are always D-linked, whereas who and what can be D-linked, but don’t have to be. I will argue below that among adjuncts where and when can be D-linked quite easily, but why and how resist D-linking. Comorovski (1989a; 1989b) and Cinque (1991) have pointed out the relevance of this notion for the possibility of extraction of wh-phrases out of selective (‘weak’) islands, in particular wh-islands. Here I will concentrate on negative islands (other selective islands will be discussed in the next section).
D-linking is not just relevant for the adjuncts, but also for arguments. Consider the following pair of questions:

(9)  
\begin{align*}
  \text{a.} & \quad \text{Who was at the party yesterday?} \\
  \text{b.} & \quad \text{Who was not at the party yesterday?}
\end{align*}

There is a clear difference in the pragmatics of these two sentences. (9a) can be answered by simply listing all the people who were at the party, but in the case of (9b) the set of all the people in the universe who were not at the party is much too large to list. Moreover, it wouldn’t be very informative to mention, say, Boris Yeltsin as one of the people who were not at the party, unless there was a reason to think that he might have been there. So to answer (9b) one has to know who was expected to be at the party in the first place. The same point can be made in another way. (9a) asks for the specification of the set of people who were at the party, (9b) asks for a specification of the complement of that set. But the complement of a set is only defined with respect to a given domain, and therefore, to be able to answer (9b), one has to know what this domain is. In other words, who in (9b) must be D-linked.

The necessity for D-linking in negative questions like (9b) can be understood in terms of the maximality (or exhaustiveness) of questions argued for in chapter 3. (9a) asks for the maximal plural individual that attended the party, that is, the sum of all the people who attended the party. (9b) asks for the sum of all people who were not at the
party. The latter sum will contain many ‘irrelevant’ individuals (like Boris Yeltsin or Aristotle) if the set of people under consideration is not contextually restricted. Although the sum of all people in the universe who did not attend the party is an object that is formally defined, it does not make sense pragmatically because it is much too ‘big’.

To explain the necessity for D-linking in (9b) we don’t need to stipulate any special principles concerning the pragmatics of negative questions. It is simply a consequence of the fact that the set of people who were at the party is much smaller than its complement with respect to the total universe of all people. D-linking is also possible in the case of (9a), especially when it is about a party that was attended by many people. In that case, the question could for instance be interpreted as pertaining to only a subset of all the people at the party, such as the friends or acquaintances of the speaker.

5.2.3 D-linking and Adjuncts

Now let’s return to the adjunct data. The fact that how and why resist extraction out of negative islands can be understood as a consequence of the inability of these wh-phrases to be D-linked. Consider the contrast between (10a) and (b) (again under the
intended interpretation that gives *why* narrow scope with respect to the negation):

(10) a. Why (do you think) Carol came to the party?
    b. # Why (do you think) Carol did not come to the party?

(11) gives quasi-formal translations of these sentences:

(11) a. ?x[x = max(λx[Carol came to the party because of x])]
    b. ?x[x = max(λx¬[Carol came to the party because of x])]

(11a) asks for the sum of all the reasons *x* such that Carol came to the party because of *x*. The maximality (or exhaustiveness) requirement built into the meaning of the question is crucial here. Suppose that Carol had two reasons for coming to the party: she was bored and she wanted to meet her friend Max. Then someone who knows why Carol came to the party must know both reasons, and not just one of them. Knowing why Carol came to the party is equivalent (in the given situation) to knowing that Carol came to the party because she was bored and because she wanted to meet Max, and for no other reason.

Now consider (11b). Here we are asked to construct the sum of all the reasons *x* such that Carol did not come to the party because of *x*. This sum will include reasons such as the following: because the earth is not flat, because the capital of China is
Beijing, because there is no grass in my back yard, because January 31st is a Monday this year, and lots of other completely irrelevant reasons that have nothing whatsoever to do with Carol or the party. Now again, this sum of reasons is formally well defined, but pragmatically it is an unwieldy monster. The problem would not arise if it were possible to restrict the set of relevant reasons contextually so that why would be D-linked. However, as pointed out in the literature, it appears that why cannot be D-linked, at least not easily, which explains the unacceptability of (11b). How is similar to why in that it resists D-linking, but when and where can be D-linked.

The question then arises why some wh-expressions can be D-linked, but others can’t. I think this has to do with the ontological status of their referents and their role in discourse. Things like people, places, and times (the referents of who, when, and where) are the basic entities that discourse is about. (See Szabolcsi and Zwarts (1993) for some remarks in this vein.) They make up the domain of discourse, and the discourse referents of Discourse Representation Theory refer to entities from these domains. As a routine matter, language users can limit the domain of discourse to certain relevant subsets of people, places, and times. Reasons and manners, on the other hand, are much more in the background. Normally, discourse is not centered around them in the way it can be centered around people, places and times. It is much harder therefore to restrict a discourse to a small set of reasons or manners that is made salient by the context than to restrict it to a salient set of people, places, or times. It is important to note, though, that it is not completely impossible to contextually restrict reasons and manners, and to
the extent that it is possible to do so, extraction out of negative islands becomes possible. Two examples are given in (12) and (13):

(12) There are several potential reasons why Carol might have come to the party. Perhaps she came because she was bored, or because she wanted to meet her friend Max, or perhaps she wanted to show off her new dress. Now, for which of these reasons/why do you think Carol did NOT come to the party?

(13) There are several ways you can travel to the West Coast. The easiest way is to take a plane, but that’s also expensive. You can also drive across the continent, but that is usually very exhausting. Another alternative is to take a train. That way you can relax and also enjoy the magnificent scenery. Now, in which of these ways/how do you NOT want to travel to the West Coast?

D-linking is much easier with expressions like for which of these reasons and in which of these ways than it is for why and how.
5.2.4 D-linking and Degree Expressions

To conclude this section, I would like to discuss cases of the kind pointed out by Kroch (1989), in which a degree expression can escape from a selective island because of D-linking. Kroch gives the example of a game show in which the rules specify that for certain infractions certain numbers of points must be deducted. Suppose for instance that a participant can be penalized by deducting either 20, or 50, or 100 points from his total. In a situation like that, an example like the following appears to be acceptable:

(14) How many points did the judges decide NOT to deduct?

A similar example can be constructed involving the conventional distances in sports like running or swimming:

(15) Rob is a super athlete who has run international races of almost any distance. How far has Rob never run?

What is characteristic of examples like these is that the degrees (numbers of points, distances) are not viewed as being ordered on a linear scale, but rather as if they were ordinary entities that are partially ordered by a Linkian part-of relation (see chapter 3). This means that the notion of maximality will be interpreted differently than in other
cases. It will not refer to the highest degree on the scale, but to the Linkian sum of degrees. Since the context provides a salient set of degrees in these examples, this sum will always be defined. Thus D-linking can save not only adjunct questions, but also degree questions from ungrammaticality due to a negative island.

5.3 Other Types of Selective Islands

5.3.1 Introduction

In chapter 1, I mentioned that negative islands are part of a larger class of selective (‘weak’) islands. In this section I will address the question whether the maximality account of negative islands that was developed in the preceding chapters can and should be extended to other types of selective islands, in particular *wh*-islands. Recall that selective islands are defined as islands which block degree expression like *how tall* and adjuncts like *how* and *why*, but not other *wh*-phrases such as *which man*. Besides negative islands, Cinque (1990) mentions the following weak types of selective islands:
(16) \textit{Wh}-islands:
  a. ? Which man did you wonder whether Bill invited?
  b. * How did you wonder whether Bill behaved?
  c. * How tall did you wonder whether Bill had become?

(17) Factive islands:
  a. ? Which man did you regret that Bill invited?
  b. * How do you regret that Bill behaved?
  c. * How tall did you regret that Bill had become?

(18) Extraposition islands:
  a. ? Which man is it time to invite?
  b. * How is it time to behave?
  c. * How high is it time to jump?

Selective islands contrast with absolute (‘strong’) islands which block the extraction of any kind of \textit{wh}-phrase. Some examples of absolute islands are given in (19) and (20):

(19) Complex NP islands:
  a. * Which man did Bill meet someone who invited \( t \)?
  b. * How did Bill meet someone who behaved \( t \)?
  c. * How tall did Bill meet someone who had become \( t \)?
Subject islands:

a. * Which man did [inviting \(t\)] seem inappropriate?
b. * How did [behaving \(t\)] seem inappropriate?
c. * How high did [jumping \(t\)] seem inappropriate?

5.3.2 Can the Maximality Account be Extended to Other Selective Islands?

The fact that negative islands and other selective islands, such as \(wh\)-islands and extraposition islands, block the same class of \(wh\)-expressions suggests strongly that a unified account should be given. Can the maximality account of negative islands be extended to the other selective islands? Let’s consider \(wh\)-islands. Can the unacceptability of (21a), for instance, be explained in terms of maximality?

(21) a. * How tall do you wonder who is?
b. \(?d[d = \text{max}(\lambda d[\text{you wonder who is } d\text{-tall}])]\)

According to the semi-formal translation given in (21b), this question asks for the maximal degree \(d\) such that you wonder who is \(d\)-tall. The ungrammaticality of this sentence would be explained if we could argue that such a maximum is not defined. Unfortunately, I don’t think that such an assumption is very plausible. Imagine a situation in which I wonder who is seven feet tall, for instance because I was told there...
is one player on the basketball team who is seven feet tall, without being told who he is. This does not imply that I should also wonder who is seven feet and one inch tall, since the identity of the seven footer may be the only question I am interested in. Consequently, in the given situation, seven feet will be the maximum degree \(d\) such that I wonder who is \(d\)-tall. We are forced to conclude that the maximality account does not obtain for \(wh\)-islands.

Similar considerations apply to the factive and extraposition islands. Consider for instance (17c). If Bill is in fact (exactly) 6 feet tall and you regret the fact that he is six feet tall, then the maximal degree \(d\) such that you regret that Bill is \(d\)-tall is six feet. In that situation, it is not true, for instance, that you regret that he is 7 feet tall. Therefore, there clearly is a maximum, and maximality can’t explain the ungrammaticality of (17c). The same kind of reasoning will show that maximality can’t be responsible for the extraposition island in (18c).

5.3.3 Is a Unified Account of Selective Islands Possible?

It appears, then, that a unified account of all selective islands in terms of maximality is not feasible. Perhaps that should not be too big a surprise, after all. There are some reasons to suppose that the selective islands in (16)-(18) are of a different nature than negative islands. For one thing, extraction of a \(which\)-phrase out of these
islands leads to at least some decrease in acceptability (see the (a) examples), but this is not the case for negative islands. (22) is perfectly fine, for instance:

(22) Which man do you think Bill didn’t invite?

Moreover, *wh*-islands appear to be more amenable to a syntactic analysis than negative islands. It is no coincidence that syntactic accounts of selective islands (like those of Rizzi (1990) and Cinque (1990)) focus primarily on *wh*-islands and then try to extend their analysis to negative islands, whereas semantic accounts (such as Szabolcsi and Zwarts (1991) and Szabolcsi and Zwarts (1993)) take negative islands as their point of departure (cf. chapter 1).

It is also relevant to note that there is a lot of cross-linguistic variation with respect to syntactic islands. In Dutch, the equivalents of (16a) and (18a) are ungrammatical for instance. The negative island effect, on the other hand, appears to be quite robust cross-linguistically, as we would expect from a semantic phenomenon.

I conclude that a uniform analysis of selective islands may be less attractive than it might appear at first sight. This leads us to consider the alternative possibility, namely that a hybrid account of selective islands is the correct one. The question we are dealing with is one of the proper division of labor between semantics and syntax. On the one hand we have seen that negative islands cannot easily be explained in syntactic terms,
but can be given a semantic account in terms of maximality. *Wh*-islands on the other hand cannot be explained in terms of maximality, but do seem to be amenable to a syntactic account.

In chapter 4 I discussed proposals by Frampton (1990) and Cresti (1993) to account for the obligatory wide scope reading of *how many*-phrases when extracted out of a *wh*-island by stipulating that a trace which is inside a *wh*-island, but locally bound from outside the island, can only be a small trace, that is, a trace which is translated as a variable of type $e$. This idea is embodied in the descriptive generalization (31) in section 4.9. Frampton and Cresti each offer different proposals for how to account for this generalization, but the exact nature of the syntactic principle responsible for (31) need not concern us here.\(^1\) Now notice that a principle of this kind will automatically rule out extraction of adjuncts and degree phrases out of *wh*-islands on the assumption that adjuncts and degree phrases are of a higher type than $e$ and therefore under movement leave a big trace. The *wh*-island facts can therefore be explained in a rather elegant way, provided we are willing to adopt a syntactic principle that refers directly to the semantic type of a *wh*-movement trace. In chapter 4 I said that such an account should be dispreferred on *a priori* grounds, because it requires a less restrictive view of the syntax/semantics interface. It is not impossible however that the less restrictive theory

\(^{1}\) Cinque (1990) and Postal (1992) have argued that the trace left under extraction out of an island cannot be a normal trace, but only an empty resumptive pronoun. Assuming that pronouns are always of type $e$, this would be yet another way of accounting for the generalization that traces inside islands can only be of type $e$. 220
turns out to be the correct one. At this point I would not want to draw any firm conclusions from this discussion, but it seems to me that a hybrid account may very well be the most feasible.

In the next section I return once more to the issue of the proper division of labor between syntax and semantics, that time in light of the fact that negative ‘island’ effects can also be induced by focus.

5.4 Maximality Effects Induced by Focus

In this section, I show that negative island effects are not limited to instances of movement. Essentially the same effects can be observed in constructions involving focus. The observations that lead to this conclusion go back to Jackendoff (1972), but they seem to have been overlooked in recent discussions of negative islands.

5.4.1 The Structured Meaning Approach to Focus

Before I can demonstrate how maximality effects can be induced by focus, I need to lay out some assumptions about focus. I will adopt the structured meaning approach to focus (as developed by Jacobs (1983) and Krifka (1991)) which can be
regarded as an implementation of Jackendoff’s ideas. According to this approach, focusing a constituent in a sentence induces a partition of the meaning of the sentence into two parts, background and focus. (The background corresponds to what Jackendoff calls the ‘presuppositional set’ of the sentence.) This is captured formally by interpreting the sentence as a ‘structured meaning’, i.e. an ordered pair $<B,F>$, where $B$ is the background and $F$ the focus. The standard interpretation of the sentence is obtained by applying $B$ to $F$. The sentence (23) for instance, will be associated with the structured meaning (b), while its standard interpretation is (c):

(23) a. Sue took JOHN to the movies.
   b. $<\lambda x[\text{took_to_the_movies}(s,x)],j>$
   c. $\lambda x[\text{took_to_the_movies}(s,x)](j) =$
      $\text{took_to_the_movies}(s,j)$

Not only sentences, but also subsentential constituents can be interpreted as structured meanings. Krifka (1991) gives compositional rules for the assignment of structured meanings to a fragment of English. In what follows I will largely ignore what happens at the subsentential level, though.

It has often been observed that interpretation of focused sentences is often stronger than is indicated in (23c). (23a) can be interpreted as implying that John is the only one that Sue took to the movies. I would like to suggest that what is involved in
this strengthening is again maximality. On the stronger interpretation, the sentence says that John is the maximal individual that Mary took to the movies. With the help of the $max$-operator this reading can be represented as in (24):

\begin{equation}
    j = max(\lambda x [\text{took\_to\_the\_movies}(s,x)])
\end{equation}

5.4.2 Negation and Focus

Jackendoff (1972) points out that sentences containing a negation are often ambiguous, depending on the way the sentence is divided into background and focus. Take for instance the following example:

\begin{equation}
    \text{Sue didn’t invite BILL.}
\end{equation}

a. Bill wasn’t one of the people Sue invited.

b. Bill was one of the people Sue didn’t invite.

(25) has two readings which are paraphrased in (a) and (b). The reading paraphrased in (a) is felicitous in a context such as the following:
(26) Sue was organizing a party for a very select group of people. Unfortunately, none of us knew who she had invited. When I saw Bill in the movie theater on the night of her party, I knew at least that Sue didn’t invite BILL.

A context that brings out the (b) reading is given in (27):

(27) Sue was organizing a party for a very select group of people. One hour before the party started she called me up in panic. She told me that there was someone very important that she had forgotten to invite, but she was so ashamed that she didn’t want to tell me who it was. When I saw her good friend Bill in the movie theater that night, I figured that Sue didn’t invite BILL.

The difference between the two readings is that on the (a) reading the property under discussion is that of being invited by Sue, while on the (b) reading it is the property of NOT being invited by Sue that is under discussion. The two readings are spelled out in (28a) and (b), respectively:
(28)  a.  NOT(<λx[invited(s,x)],b>)
    Truth conditions: [b] ∈ [λx[invited(s,x)]]

b.  <λx[¬invited(s,x)],b>
    Truth conditions: [b] ∈ [λx[¬invited(s,x)]]

Here NOT is an operator (introduced here merely for expository purposes) that indicates that the background does not apply to the focus. Note that the truth conditions of the two sentences end up being the same. They only differ in whether the negation is part of the background or not.

The two readings of (25) may be distinguishable by means of intonation. If the sentence is pronounced with a falling pitch accent at the end, it tends to get the (a) interpretation. With this intonation, the sentence has the flavor of being ‘incomplete’. A possible completion is given in (29):

(29)  Sue didn’t invite BILL, she invited SAM.

If the sentence is read with a falling pitch accent at the end, however, the (b) reading is preferred. With this intonation the sentence has a ‘complete’ flavor.
5.4.3 Negation and Degree Expressions

With this much as background, we can now turn to examples that contain a negation and a focused degree expression. It turns out that in such cases the ambiguity disappears. An example is (30) which contains the focused measure phrase 300 pounds:

(30) Bill doesn’t weigh [f 300 POUNDS].

a. 300 pounds isn’t what Bill weighs.

b. * 300 pounds is what Bill doesn’t weigh.

(a) is a possible paraphrase of (30), but (b) isn’t. It might be objected that the (b) paraphrase is ungrammatical because a non-referential wh-phrase is extracted from a negative island. That is true, but it doesn’t affect my point, which is that (30) does not have the interpretation that the (b) paraphrase would express if it were grammatical. (30) can be used to deny that Bill’s weight is 300 pounds, but it cannot be used to assert that 300 pounds is what Bill does not weigh.

2 Jackendoff (1972) makes the same observation with an example involving an instrumental:

(i) Maxwell didn’t kill the judge [f with a silver HAMMER].

a. It wasn’t with a silver hammer that Maxwell killed the judge.

b. * It was with a silver hammer that Maxwell didn’t kill the judge.

3 Another objection that might be raised is that the focus cases could be analyzed as involving islands after all if we assume that focused constituents are obligatorily moved
If we pay attention to the intonation of the sentence, we observe again that the (a) reading is associated with a rising pitch accent at the end of the sentence (as in \textit{Bill doesn’t weigh 300 pounds; he weighs 250 pounds}), whereas the ungrammatical (b) reading is characterized by a falling pitch accent at the end.

The two readings are represented by the background-focus structures in (31):

\begin{align}
\text{(31)} & \\
& \text{a. NOT}(\lambda d[\text{weigh}(b,d)],300\text{lb}) \\
& \text{b. } \lambda d[\neg\text{weigh}(b,d)],300\text{lb}
\end{align}

We can understand the ungrammaticality of the (b) reading if we assume the stronger interpretation of focus, involving maximality. Under that interpretation, the truth conditions of (31b) are as follows:

\begin{align}
\text{(32)} & \\
& 300\text{lb} = \max(\lambda d[\neg\text{weigh}(b,d)])
\end{align}

There is no maximal degree \(d\) such that Bill doesn’t weigh \(d\)-much, which explains why the (b) reading does not obtain. Note that the ungrammaticality of the (b) reading is explained in exactly the same way that the negative island data were accounted for in the preceding three chapters. Because of the downward entailing context, there is no

at LF. For convincing arguments against this view, see Rooth (1985).
maximal degree, and therefore the sentence cannot be interpreted. The fact that
maximality effects can be induced by focus, and not just by \textit{wh}-movement, shows that
the ‘negative islands effect’ is not strictly speaking an island effect at all. Exactly the
same effect shows up in a construction where there is no movement but whose
semantics involves maximality. In terms of the issue of the proper division of labor
between syntax and semantics as discussed in section 5.3, the data discussed in this
section most clearly show that the ‘negative islands effect’ can only be accounted for in
semantic terms.


Joly, A (1967). *Negation and the Comparative Particle in English*. Presses Université Laval, Quebec.


McCawley, J.D. (1981). *Everything that Linguists Have Always Wanted to Know about Logic* (*but were ashamed to ask).* Basil Blackwell, Oxford.


