

What Moves Where When in Which Language?

by

Norvin W. Richards III

B.A. Linguistics
Cornell University, 1993

Submitted to the Department of Linguistics and Philosophy
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Linguistics
at the
Massachusetts Institute of Technology

September 1997

©1997 Norvin W. Richards III. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document in whole or in part.

Signature of Author: Norvin W. Richards III
Department of Linguistics and Philosophy
September 3, 1997

Certified by: David Pesetsky
David Pesetsky
Professor of Linguistics
Thesis Supervisor

Accepted by: Robert Stalnaker
Robert Stalnaker
Professor of Philosophy
Chairman, Department of Linguistics and Philosophy

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

SEP 30 1997

LIBRARIES

ARCHIVES

What Moves Where When in Which Language?

by

Norvin W. Richards III

Submitted to the Department of Linguistics and Philosophy
on September 3, 1997 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Linguistics

ABSTRACT

Much work in syntax has used the properties of *wh*-movement as a probe into the nature of the derivation. One perennial issue is the nature of *wh*-in-situ. Is *wh*-in-situ related to its scopal position by an operation like movement or by an entirely different process? If *wh*-in-situ does undergo invisible movement, why is this movement invisible? If we assume a derivational model, what is the relation between overt and covert movement in the derivation?

In this thesis I will investigate the properties of multiple-*wh* questions in a number of languages (particularly Bulgarian, Serbo-Croatian, Chinese, and Japanese), in an attempt to find evidence for a particular answer to these questions. I will argue that the classic model assumed by the Extended Standard Theory is essentially correct; there is covert movement, and all covert movement follows all overt movement in the derivation (and is therefore invisible because it takes place after the point in the derivation at which the representation is interpreted by the phonological component).

One crucial aspect of the argument will involve investigation of the nature of additional-*wh* effects. I will claim that additional-*wh* effects only appear when certain structural and derivational conditions on the relation between the *wh*-movements involved are met, and additional-*wh* effects can therefore be used to determine which *wh*-movement operations precede which others.

Chapter 1 is an overview of some competing claims about the architecture of the grammar, and a discussion of the nature of evidence that might help us to choose among these claims. In Chapter 2 I discuss the distribution of *wh*-island effects in a number of languages, arguing that the overt/covert distinction is in fact irrelevant to the distribution of *wh*-islands. Chapter 3 is a discussion of the nature of Superiority effects in several languages. In Chapter 4 I investigate the nature of feature strength and develop a version of Procrastinate which is empirically distinct in several desirable ways from that developed by Chomsky (1993). Finally, Chapter 5 discusses additional-*wh* effects in some detail.

Thesis Supervisor: David Pesetsky
Title: Professor of Linguistics

Contents

Chapter 1: Introduction	12
1. Proposals	13
1.1 “T Model”	13
1.2 No Covert Movement	15
1.3 Single Output Syntax	15
1.4 Invisible Overt Movement	16
1.5 Theoretical summary	17
2. Assumptions	18
Chapter 2: Subjacency Forever	21
1. Problems: Levels and Taxes	21
2. CP-Absorption and IP-Absorption	25
3. Diagnostics for CP- and IP-Absorption	27
3.1 Wh-islands	27
3.2 Scrambling	27
3.3 Superiority	29
3.4 Weak Crossover	32
3.5 Wh-movement and QR	34
3.6 Interacting wh-dependencies	37
4. “Heterogeneous Movement” languages: German and English	44
5. Interlude: Hungarian	47
6. Conclusion	50
Chapter 3: Featural Cyclicity and the Ordering of Multiple Specifiers	53
1. Multiple wh-movement	60
2. Multiple A-specifiers	76
2.1 Idiom chunks	77
2.2 Quantifier scope	79
2.3 Scrambling generally	81
3. Object shift	88
4. Spec NegP	97

5. Cliticization	100
6. Freedom of ordering	101
6.1 Morphological effects on syntax: clitic ordering	104
6.2 Syntactic effects on syntax: Bulgarian <i>wh</i> -words	106
7. Shortest Move	109
8. Equidistance	113
9. Conclusion	117
 Chapter 4: In Full Pursuit of the Unspeakable	 118
1. Movement through a “weak” position	124
1.1 French participles	124
1.2 Object Shift in Mainland Scandinavian	126
1.3 V-to-I in Mainland Scandinavian	128
1.4 Japanese <i>wh</i> -movement	129
1.5 Tagalog vs. Chamorro	131
1.6 Agreement in Mohawk, Chicheŵa, and Irish	135
1.6.1 Mohawk	135
1.6.2 Chicheŵa	139
1.6.3 Irish	142
1.7 Derived Object Condition	145
1.8 Bulgarian vs. Serbo-Croatian	149
1.9 Conclusions	151
2. Destroying the evidence: ellipsis	159
2.1 (Pseudo)gapping	159
2.2 Multiple Sluicing	162
2.3 Conclusions	165
3. “Bottom-heavy” chains	166
3.1 Japanese	167
3.2 Malay	170
4. An embarrassment of riches: movement to two strong features	173
4.1 “Weakening” the subject position	174
4.1.1 Anti-anti-agreement	181
4.1.1.1 Morphological anti-anti-agreement	181
4.1.1.2 Anti-anti-agreement and anaphora	184
4.1.1.3 Person-animacy hierarchies and anti-agreement	186
4.1.2 Anti-agreement and ergativity	189
4.1.3 Interlude: Chamorro <i>-um-</i>	195
4.1.4 Anti-object-agreement	199
4.2 Subject-complementizer interactions	199
4.3 Improper movement, and proper improper movement	208
4.4 Japanese subject scrambling	215
4.6 Conclusion and expansions	219
5. Overall conclusions	222
 Chapter 5: The Principle of Minimal Compliance	 225
1. Introduction	225
2. Applications	228
2.1 Reflexivity	229
2.2 Weak Crossover	231
2.3 VP-ellipsis	232
2.4 <i>that</i> -trace effects	233
2.5 Subjacency, CED	235
2.5.1 Dependencies	237

2.5.2 Links; Connectedness	240
2.5.3 Interlude: Subjacency/CED vs. Connectedness	245
2.6 Attractors and Islands	250
2.6.1 Islands	251
2.6.1.1 Additional-wh effects in Bulgarian and Japanese	251
2.6.1.2 Inside jobs; additional additional-wh effects	262
2.6.2 Attractors	270
2.6.2.1 Superiority	272
2.6.2.1.1 Bulgarian	276
2.6.2.1.2 Japanese	277
2.6.2.2 Path Containment Condition	292
2.6.2.2.1 The PCC in other languages	294
2.6.2.2.2 Volvo-sentences	318
3. Some possible further expansions	325
3.1 Stylistic Fronting	325
3.2 Holmberg's Generalization	328
3.3 Lower-wh effects	329
4. Is the PMC recursive?	333
4.1 Path Containment Condition	333
4.2 <i>that</i> -trace effects	335
4.3 Parasitic gaps	337
5. Conclusion	337
Chapter 6: Conclusion	340
Bibliography	348

*Ang di lumilingon
sa pinanggalingan
ay di makararating
sa paroroonan.*

--Tagalog proverb

Who does not look back
at where he came from
will not reach
where he is going.

Acknowledgments

Before I wrote this thesis I used to believe that papers could in principle be finished. I guess I still believe that, in principle, but it certainly hasn't happened here; there's lots of work still to do, including a lot of internal contradictions still to be resolved, and at least one section which I am now pretty sure is false. I think I prefer that, in a way, to the alternative; this thesis is (hopefully) an outline for a research program I'll be able to follow for a while, rather than something I'll now abandon for greener pastures. Anyway, time's up. Time to thank all the people who made this possible, starting with my committee.

Working with David Pesetsky has been a real pleasure. David has a rare gift for seeing a problem from many angles, from which I've benefited on more than one occasion; in earlier versions of this thesis I attempted to footnote all the ideas I've gotten from him, but eventually gave up in embarrassment. Meetings with him have typically involved brainstorming past various problems with my analyses, creating numerous possible solutions where I could only see one (or none) before (or, just as helpfully, eliminating superfluous "solutions" that I could see no good way of choosing between). At its best, conversation with him has the quality not so much of wisdom received from on high but of collaborative effort, giving me the feeling that I could do this syntax stuff myself, given time. He has also frequently saved me from my own theoretical cowardice, encouraging me to seek evidence for seemingly crazy results of my theories, when my first impulse would have been to wallpaper over the problems.

I'm very fortunate to have had the chance to work with Noam Chomsky, and the influence of his ideas on the approach developed in this thesis will be obvious. Besides being a fairly intelligent guy, Noam is free of the rhetorical bluster which so often seems to characterize people in prestigious positions in their field. My appointments with him often radically changed the way I was looking at my subject, always (I think) for the better.

I'd also like to thank Shigeru Miyagawa, my fellow Tuscaloosan, for his help and encouragement in the writing of this thesis. The Japanese portions of this thesis, in particular, are much richer for Shigeru's vast knowledge of the Japanese syntactic

literature, as well as for his willingness to consider and render judgments on any example sentence, no matter how convoluted. His friendly enthusiasm has also been an important driving force behind my work here. I remember running into Shigeru once as he was jogging along Memorial Drive, shortly after I had arrived in Boston. He stopped to exuberantly welcome me to the department and tell me that he looked forward to working with me, thereby making the department seem considerably less scary and remote than it had a few seconds before.

Scheduling problems prevented Ken Hale from being a member of my committee, in the end, so I join the large number of people who thank Ken in their acknowledgments as a “member in spirit” of the committee. Ken’s spiritual contributions to this thesis have been considerable, particularly in the discussion of Mayan languages in chapter 4, which is largely based on a conversation I had with him. More generally, he is one of the foremost practitioners I know of the art of doing rigorously formal linguistics which pays crucial attention to data from as wide a selection of languages as possible, and this is an art I have tried to learn from him. I’m also grateful to Ken for giving me the chance to work with him in Australia in 1996 on a dictionary of Lardil, thus allowing me to learn fieldwork in the field from a master. Ken is also one of the best human beings I know; it’s good to know that someone as nice as he is can be as successful as he’s been.

On to my classmates. It’s been fun exchanging ideas, deep thoughts, declaimed beliefs, books, puzzled stares, rants, zwrites, and foodstuffs with Martha McGinnis, although she has an irrational hatred of relative pronouns and does in fact cut her bananas funny (cf. Harley 1995 for discussion). David Braun, Yiddish *maven* extraordinaire, was a steadying influence on all of us before he gallivanted off to Harvard to teach Yiddish. I’ve enjoyed both agreeing and disagreeing with Danny Fox; I’ve even enjoyed disagreeing with him at first and then agreeing with him later, which is a neat trick on his part (and every flag does too, as does its mother). Paul Hagstrom, besides being a great officemate and a responsible guinea-pig owner, has been fun and interesting to talk to about linguistics; I find that we often converge on the same answers to complicated questions, which suggests that we’re both deeply right, as far as I’m concerned. Ingvar Löfstedt introduced me to Tarkovsky and King Crimson, defined in precise phonological terms the difference between a duck and a ghost, checked out the Sigma, and composed a number of ditties about desks, hardware, the light of the moon, and Zero Syntax, all while maintaining a perfectly straight face. Rob Pensalfini, the only actor/singer/songwriter/linguist I’ve ever known, prepared the best pasta I’ve ever had over a campfire, as well as sharing his kangaroo with me (with generous side helpings of interesting conversation about fieldwork, morphology, endangered languages, Shakespeare, constellations, male and female emus (yeah, emus!), and life in general). Uli Sauerland, my other great officemate, when not busy biking to foreign lands, has cheerfully supplied me with German judgments, German chocolate, and access to his immense library. Finally, Judy Yoo-Kyung Baek was a mystic and seemingly infallible guide to things Korean, especially during the field methods course on Korean, where she represented her native tongue with uncanny calm and poise.

It was a lucky day for me (though maybe not so lucky for her) when Roumi Izvorski decided to visit here last year, thus giving me near-constant access to Bulgarian wh-movement intuitions. Roumi’s clear and consistent intuitions and her careful examination of them have added immeasurably to this thesis (several sections of which arose from her noticing facts about her own intuitions and asking if I had an account of them); it helps, of course, that she’s fun to be around. Heidi Harley was always able to cheer me up when things seemed dismal, and she has a name which is fun to say for metrical reasons, besides. I’m proud to strongly resemble her uncles, as I apparently do.

Andrew Carnie is partly responsible for introducing me to Civilization™, and thus for shortening this dissertation considerably (or else making it possible at all, I'm not sure which), and has also helped to maintain my hope that one can become rich and famous by working on the grammar of an obscure V-initial language. It's been a great pleasure talking linguistics with and occasionally being fed by Hooi Ling Soh, queen of Malay wh-movement, Chinese NP movement, and pickled vegetables. Colin Phillips' boundless energy, rigorously syntactic mind, and unfailing good humor were high on the list of the many pleasant things about living and working in this department; also, it was fun exchanging basic Italian greetings with him (*ciao, Colino!*).

This thesis was brought to you in large part by the efforts of a number of people who lent me their expertise in a variety of languages. For their help with the facts of Bulgarian, I'd like to thank Roumyana Izvorski, Ani Petkova, Virginia Savova, Roumyana Slabakova, and Kamen Stefanov. Sandy Chung helped me with the facts of Chamorro, for which I am grateful. My Chinese informants were Lisa Cheng, Hooi Ling Soh, and Wei-Tien Dylan Tsai; many thanks. Thanks, too, to Sten Vikner, for his help with Danish, and to Sjeff Barbiers and Iris Mulders, for their Dutch intuitions. I'm also very grateful to Marie Claude Boivin for her French judgments, and to Uli Sauerland and Susi Wurmbrand, who gave me a lot of help with understanding the facts of German. For his help with some Icelandic data, many thanks to Höskuldur Thráinsson. I'd also like to thank Takako Aikawa, Minoru Fukuda, Shigeru Miyagawa, Masao Ochi, Satoshi Oku, and Kazuko Yatsushiro for their help with the Japanese facts. Thanks (again) to Hooi Ling Soh for some Malay facts. I am grateful to Arild Hestvik and Øystein Vangsnes for their Norwegian judgments. For their assistance with the facts of Serbo-Croatian, many thanks to Željko Bošković, Wayles Browne, Damir Cavar, Steven Franks, Snjezana KordiĆ, and Milan Mihaljević. Finally, I would like to thank Lama Ngawang Jorden for his Tibetan judgments, and for teaching me some of his language.

There isn't much Tagalog in this thesis, but I'd like to thank Marlon Abayan, Alex Diaz, Nicole Lazo, and Luis Sarmento, my local Tagalog informants, for their hard work with me over the years, and all the other folks in MIT's Filipino Students Association, for their friendly and easy acceptance of me as the only blond Filipino in the group (and for not breaking my ankles with bamboo poles when they had the chance). *Maraming salamat sa inyong lahat*. Thanks, too, to all the folks who've made it fun to be an Austronesianist, including Jan Voskuil, Lisa Travis, Hooi Ling Soh, Andrea Rackowski, Viviane Phillips, Matt Pearson, Ileana Paul, Masanori Nakamura, Anne MacLachlan, Ed Keenan, Dan Finer, and Sandy Chung.

There is no Lardil at all in this thesis. But I'm very grateful to the Lardil people for their hospitality and support of me and the other linguists who worked on the Lardil Dictionary during July and August of 1996; thanks, especially, to my *thabu* and namesake Cyril Moon (Birdibir), my *mimay* Kenneth Jacobs (Kulthangarr), and my *wunga* Lindsay Roughsey (Burrurr), for all the time they spent making sure we got their language down right. *Waa, ngithun kubarda jika! Banda nyerrwi ngakulmu merri ngalumerrinjithu*. Thanks, too, to Anna Ash, David Nash, Jane Simpson, and especially Ken Hale, whose work now and in the past has been an inspiration to all of us.

I'd also like to thank a number of people for friendship, food, deep linguistic and non-linguistic discussions, and other essentials of life, especially Anna Ash, Artemis Alexiadou, Ben Bruening, Chris Collins, Brett Baker, Masanori Nakamura, Roumyana Slabakova, Jay Keyser, Mark Newson, Bridget Copley, Brigit Cosgrove, Orin Percus, David Embick, Máire Noonan, Meltem Kelepir, Masao Ochi, Jason Lenicheck, Calixto Agüero-Bautista, Jim Huang, Masatoshi Koizumi, Taylor Roberts, Vijayanthi Sarma,

Matt Pearson, Elena Benedicto, Peter Svenonius, Wei-tien Dylan Tsai, Susi Wurmbrand, Elena Anagnostopoulou, Iris Mulders, John Frampton, Jon Nissenbaum, Ngawang Jorden, Satoshi Oku, Ruben van de Vijver, Howard Lasnik, Martin Hackl, Father Angg, Morris Halle, Maya Honda, Gaurav Mathur, Seth Minkoff, Dianne Jonas, Monica Santa Maria Somohano, Jen Purdy, Gereon Müller, João Costa, Luciana Storto, Ileana Paul, Renate Musan, Jonathan Bobaljik, Linda Thomas, Cornelia Krause, Wayne O'Neil, Lisa Travis, David McKay, Michel DeGraff, Jane Simpson, Joachim Sabel, David Nash, Marie-Hélène Côte, Vivian Lin, Masha Babyonyshev, Marie Claude Boivin, Hubert Truckenbrodt, Hidekazu Tanaka, Sjef Barbiers, Alec Marantz, Sarah Gordon, Hiroyuki Ura, Kazuko Yatsushiro, Pilar Barbosa, Yoonjung Kang, Rajesh Bhatt, Fleur Veraart, Bev "Veb Lhots" Stohl, Sveva Besana, Sheila Anderson, Takako Aikawa, Ayumi Ueyama, Akira Watanabe, Željko Bošković, Andrea Zukowski, and all the people who were left out of this list for no good reason.

Thanks, too, to Building 20, for being a great place to think, sleep, play video games, and occasionally work. I miss it already.

My undergraduate life at Cornell was excellent preparation for my life as a grad student (besides being a lot of fun). Thanks for that are due to a number of people, especially Wayne Harbert, my first linguistics professor, who was always willing to carefully and completely answer an ill-informed question or discuss an outlandish theory, and whose example prompted me to become a syntactician. Thanks, too, to (in no particular order) Carol Rosen, Molly Diesing, Erika Mitchell, Jim Huang, Lynn Santelmann, Luis Lopez, Michael Bernstein, Zelmira Nuñez del Prado, Tomoyuki Yoshida, John Bowers, John Whitman, Jay Jasanoff, G.N. Clements, Beverley Goodman, Gennaro Chierchia, and all the folks at Watermargin.

Finally, I'd like to thank my family, especially my parents, Norvin and Janet Richards, and my sister, Genevieve Stevens, for their love, support, and English judgments. This thesis is dedicated to them.

Abbreviations

The following is a list of the abbreviations used in the glosses in this thesis.

A	actor (Austronesian; ≈logical subject)
ABS	absolutive
AC	agentive construction (Mayan; indicates extraction of subject)
ACC	accusative
AGR	agreement
AP	antipassive
APPL	applicative
ASP	aspect
AT	actor-topic
AUG	augment
AUX	auxiliary
CAUS	causative
CL	(noun) class
COMP	complementizer
CONDIT	conditional
COP	copula
DAT	dative
DEF	definite
DET	determiner
DU	dual
DUP	duplicative
EMPH	emphatic
ERG	ergative
FACT	factual
FEM	feminine
FOC	focus
FUT	future
G	goal (Austronesian; ≈logical object)

11		
GEN	genitive	
GT	goal-topic	
HAB	habitual	
HON	honorific	
IMP	imperfective	
INDIC	indicative	
INST	instrumental	
IRR	irrealis	
LOC	locative	
M	masculine	
NEG	negative, negation	
NM	nominalizer	
NOM	nominative	
NONFUT	non-future	
OBJ	object	
OBL	oblique	
OM	object marker	
OT	object topicalization (Karitiana; indicates extraction of object)	
PART	participle	
PERF	perfective	
PL	plural	
POSS	possessive	
PRES	present	
PREV	preverb	
PROG	progressive	
PUNC	punctual	
Q	interrogative	
R	realis	
REL	relative	
SG	singular	
SM	subject marker	
SRFL	semireflexive	
STAT	stative	
SUBJ	subject	
SUBJN	subjunctive	
T	topic (Austronesian)	
TNS	tense	
TOP	topic	
TRANS	transitive	
UNM	unmarked case	

In the discussion of Mohawk I have used Baker's (1996) abbreviations for the Mohawk agreement morphemes; the relevant abbreviations are:

1	first person	s	singular	S	subject
2	second person	p	plural	O	object
M	masculine			P	possessor
F	feminine				
N	neuter				

Chapter One: Introduction

We can distinguish three main ways of forming multiple-wh questions in the world's languages. In some languages, such as Bulgarian, Romanian, Serbo-Croatian, and Polish, all wh-words move to their scopal positions:

Bulgarian (Bošković 1995a, 13)

- (1) Kogo kakvo e pital Ivan?
whom what AUX asked Ivan
'Who did Ivan ask what?'

In another class of languages, which includes Japanese, Korean, Chinese, and Tibetan, wh-words typically remain in situ:

Japanese

- (2) Taroo-ga dare -ni nani -o ageta no?
Taroo NOM who DAT what ACC gave Q
'Who did Taroo give what?'

Finally, there is a class of languages, including English, which mixes these two strategies, moving one wh-word and leaving the rest in situ:

(3) Who did John give *t* what?

The right way of characterizing the differences between these languages has been a topic of much discussion in the literature, and will be the main topic of this thesis.

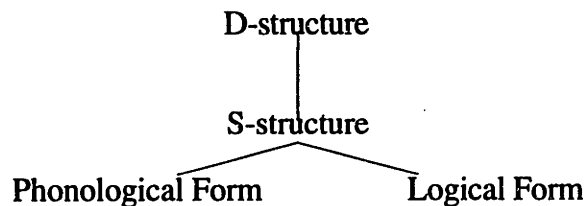
1 Proposals

In this section I will outline some of the proposed accounts of the relevant differences between the three classes of languages. I will try to show in an abstract way the different predictions these proposals might make.

1.1 “T Model”

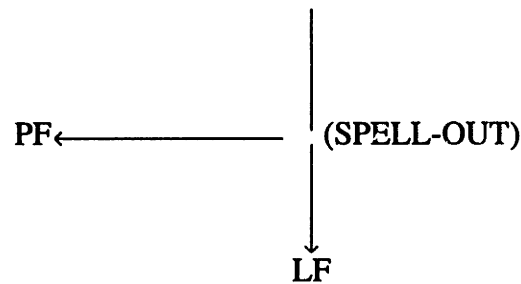
One approach to the facts in (1-3) is defended in Huang (1982) and Lasnik and Saito (1984), among many others. The approach is based crucially on certain aspects of the architecture of the grammar given in (4), one of the basic assumptions of the Extended Standard Theory:

(4)



In (4), the syntactic component consists of a mapping from D-structure, the level of lexical insertion, onto S-structure by means of transformations. The S-structure generated by this process is then sent to Phonological Form (henceforth PF), where it is given a representation usable by the articulatory-perceptual mechanisms responsible for speech, and to Logical Form (henceforth LF), where it receives a semantic representation. This approach has a more or less direct descendant in the architecture assumed by Chomsky (1995):

(5)



(5) differs from (4) in that no linguistically significant levels of D-structure and S-structure are posited; rather, the derivation simply proceeds toward LF, and at some point in the derivation (referred to as Spell-out) the representation is sent to PF. On this approach, no linguistic conditions should refer specifically to the levels of representation known as S-structure and D-structure in earlier models; conditions are to be stated with reference only to LF and PF. I will refer to the pre-Spell-out part of the derivation as the *overt syntax*, and to the post-Spell-out part as the *covert syntax*.

The proposal defended by Huang (1982) and Lasnik and Saito (1984) is that the differences between languages illustrated in (1-3) have to do with the point in the derivation at which wh-movement takes place. In some languages, such as Bulgarian, all wh-movement takes place in the overt syntax, and therefore all wh-movement has an effect on the phonological form of the sentence, since it is the representation created by movement in the overt syntax which is sent to PF. Other languages, such as Japanese, perform all wh-movement in the covert syntax; this wh-movement is therefore invisible, since the representation sent to PF is not one to which wh-movement has applied. Finally, there are languages like English, in which some movement is overt and the rest covert.

This approach to wh-movement has a number of salient properties which will distinguish it from some of the other approaches which I will discuss shortly. One property it has is that all overt movement precedes all covert movement; that is, in a language like English, moved wh-words move before wh-words in situ. Another is that languages like Bulgarian and languages like Japanese have in common the property that

they perform all wh-movement in a single “component” of the derivation; in English, on the other hand, some movement is overt and the rest is covert.

1.2 No Covert Movement

Another theory about the right way to distinguish among the classes of languages in (1-3) has been proposed by, among others, Cole and Hermon (1994), Tsai (1994), and Reinhart (1995). This approach claims that some or all cases of wh-in-situ should be interpreted in situ, without assuming a process of covert movement which has anything in common with overt movement. On an approach of this type, we do not expect the relation between wh-in-situ and its scopal position to have any properties which are unique to movement. Moreover, these theories differ from the one in the immediately previous section in that they do not predict a similarity between languages of the Bulgarian type and languages of the Japanese type; these languages have nothing in common, apart from the negative property of not being languages of the English type. If there are syntactic properties which are crucially sensitive to the timing of wh-movement, then we expect Bulgarian and Japanese to behave differently, and neither should behave like English.

1.3 Single Output Syntax

A third type of theory about the distinction between moved wh-words and wh-in-situ has been argued for by Bobaljik (1995), Brody (1995b), Groat and O’Neil (1996), and Pesetsky (to appear). On this type of account, the architecture of the theory is very different from that given in (4) and (5). The syntax produces a single representation, which is interpreted, without further alteration, by both the semantic and the phonological components.

(6)



This account does regard both moved *wh*-words and *wh*-in-situ as being related to their scopal positions by movement, but regards the distinction between overt and covert movement as a purely phonological one. The idea is that movement chains may have either their heads or their tails pronounced; if the head of a chain is pronounced, the movement has an effect on the phonological form and is therefore “overt”, while if the tail is pronounced the chain cannot be detected from the phonology and is therefore “covert”. These approaches differ from those in 1.1 in that overt movement does not necessarily precede covert movement in the derivation; overt and covert movements may be “interleaved”. The timing of *wh*-movement in the derivation, then, is the same in English, Japanese, and Bulgarian; these languages differ only phonologically, on these approaches. If there are syntactic properties which crucially rely on the order in which *wh*-movement occurs in the derivation, these languages should all behave alike. In particular, there should be no difference between English on the one hand and Bulgarian and Japanese on the other, as there should on the theory given in 1.1.

1.4 Invisible Overt Movement

A fourth approach to the overt/covert distinction was proposed by Watanabe (1992). Operating under the theoretical architecture in (4), Watanabe argues that the timing of *wh*-movement in languages like Japanese is actually identical to that of English. He postulates a type of movement which occurs in the overt component but which has no effects on the phonological representation. His claim is that Japanese, like English, performs one *wh*-movement to every interrogative C^0 in the overt syntax, and performs all other *wh*-movement in the covert syntax. The difference between Japanese and English, on his view, is a purely phonological one; Japanese overt *wh*-movement has no effect on the phonology, and is thus indistinguishable phonologically from covert movement. Watanabe does not discuss languages of the Bulgarian type, but without further elaboration his theory would predict a contrast between Bulgarian on the one hand and Japanese and English on the other. The differences between Japanese and English are purely

phonological on his view, as we have seen, but Bulgarian differs syntactically from both, in that all of its *wh*-movement is in the overt component. If there are syntactic diagnostics which make crucial reference to the timing of *wh*-movement, we expect to find, on this theory, that Japanese and English pattern alike to the exclusion of Bulgarian.

1.5 Theoretical summary

The four theories sketched above do not by any means exhaust the range of possibilities, or even the range of theories which have been suggested and defended. Still, let us take these four as a starting point. The differences between the four theories sketched above with respect to their claims about the relative timing of *wh*-movements may be represented as follows (where the arrow represents the order in which operations take place in the derivation):

(7) 

T model (1.1):

All Bulgarian movement,
some English movement

All Japanese movement,
some English movement

No covert movement (1.2):

All Bulgarian movement,
some English movement

Single output syntax (1.3):

All Bulgarian movement,
all Japanese movement,
all English movement

Invisible overt movement (1.4):

All Bulgarian movement,
some Japanese movement,
some English movement

Some Japanese movement,
some English movement

As noted in the text above, the theory outlined in 1.1 is the only one in which Bulgarian-type languages and Japanese-type languages have something in common to the exclusion of English; they are the languages in which all wh-movement occurs in a single part of the derivation. In the theory in 1.2, the three languages have virtually nothing in common as far as the timing of wh-movement is concerned. The theory in 1.3 equates all three languages. Finally, the theory in 1.4 predicts that Japanese and English should behave similarly to the exclusion of Bulgarian.

In this thesis I will investigate the properties of multiple wh-movement in these languages and attempt to isolate phenomena which can serve as diagnostics for the timing of wh-movement. I will try to show that Bulgarian and Japanese do indeed exhibit syntactically similar behavior which differs from the behavior of English, and will offer an account of the similarities and differences which makes crucial reference to properties of the derivation. The account will thus be an argument for a “classic” theory of the type in 1.1, in which both wh-in-situ and overtly moved wh-words are related to their scopal positions via movement, and in which overt movement precedes covert movement in the derivation.

2 Assumptions

The thesis will be written in the framework of the Minimalist Program of Chomsky (1993, 1995). I will assume a strongly derivational model in which syntactic trees are built up out of lexical items via the operations Merge and Move. The derivation begins, I will assume, with a Numeration consisting of a set of lexical items to be used.

Merge is a binary operation which takes two items, either lexical items from the Numeration or more complex items created via previous syntactic operations, and makes them into a single complex item with two parts. This complex item can then again be Merged with additional structure to create more complex items, until all the lexical items in the Numeration are used up.

At any point in the derivation, the operation Move may also be employed. Move, I will assume, is an operation driven by a need to check a feature. I will be assuming

Chomsky's (1995) version of Move; on this version, features on syntactic heads act as "attractors", attracting features lower in the tree and forcing them to move into positions which are local to the attractor. For poorly-understood reasons, attracted features often "pied-pipe" material syntactically associated with them, so that entire phrasal categories move rather than single features; the eventual hope would be to show that the amount of material which must be pied-piped follows from independent properties of PF and/or of LF. Chomsky's (1995) claim is that movement in the covert syntax consists purely of feature-movement, and that pied-piping of extraneous material occurs only in the overt syntax. The truth of this claim will not be crucial to anything I will have to say.

Again following Chomsky (1995), I will assume a distinction between "interpretable" and "uninterpretable" features, where the former are features which carry some instructions for either PF or LF and the latter are purely "formal" features, with no significance to the PF or LF interfaces. Because uninterpretable features have no significance at the interfaces, they can (and arguably must, by a principle of Full Interpretation) vanish when checked. Interpretable features, by contrast, continue to exist after checking; thus, they can in principle be checked multiple times.

I will be assuming that the derivation obeys a principle of Cyclicity (which will be discussed in chapter 3) which constrains the order in which operations may occur. Cyclicity will effectively force the derivation to work "from the bottom up", allowing only those operations which make reference to a position at or near the top of the structure created at the point in the derivation at which they occur. I will also assume that Move is constrained by some kind of locality requirement of the Shortest Attract type, the nature of which will be discussed further in chapters 3 and 5. I will assume that wh-movement is (at least in principle) successive-cyclic, stopping in every intermediate Spec CP on the way to the Spec CP which is its scopal position.

Finally, I will assume the theoretical architecture in (5). That is, I will try to show that a simple account of the multiple-wh-movement facts in a number of languages can be

constructed on the assumption that *wh-in-situ* does undergo a movement operation in the covert syntax, and that all covert operations follow all overt operations. I will use “overt” and “covert” throughout to refer to parts of the derivation; overt movement is movement which precedes Spell-out, and covert movement is movement which follows Spell-out. Thus, it is in principle possible for overt movement (of a null operator, for instance) to have no effect on the phonological representation.

In the next chapter I will discuss certain properties of Subjacency. In particular, I will give evidence that Subjacency constrains both overt and covert movement. This result will have certain implications for our choice among the theoretical architectures discussed above. If Subjacency only constrains movement relations, the fact that it constrains the relation of *wh-in-situ* to C^0 suggests that this relation does indeed involve movement (hence my use of the term “covert movement” to describe this relation), and thus argues against those approaches which deny the existence of covert movement. Similarly, the approach sketched in the next chapter will undermine theories assuming invisible overt movement in languages like Japanese. The original motivation for this theory had to do with the distribution of Subjacency effects in Japanese, and we will see that the relevant facts can be captured without postulating invisible overt movement.

Chapter Two: Subjacency Forever

1. Problems: levels and taxes

Baker (1970) noted that sentences of the following form are ambiguous in English:

- (1) [Who asked [who bought what]]?

Here *what* can have scope either with the *who* in the embedded clause or with the *who* in the matrix clause. The availability of the latter reading is somewhat surprising in light of the ill-formedness of (2):

- (2) * [What did you ask [who bought]]?

That is, the process whereby *what* gets its scope in (1) is apparently immune to the constraint which rules out (2). One can imagine a number of ways in which the contrast between (2) and the wide-scope reading of (1) might be explained; I will concentrate here on two.

The first, which has been fairly well accepted in much of the literature on this problem, would be to say that the LF movement whereby *what* gets its scope in (1) is immune to Subjacency. I will refer to approaches of this kind, which make crucial reference to the level of representation on which movement takes place, as **levels**

approaches. One argument for a levels approach comes from Huang (1982), who notes that in Chinese, a language in which wh-movement is apparently not overt, no wh-island effects are observed. The same is true in Tibetan¹:

Chinese (Huang 1982, 267)

- (3) a. Ni xiang-zhidao shei mai-le sheme?

you wonder who bought what

‘What do you wonder who bought?’

Tibetan (Ngawang Jorden, p.c.)

- b. Khong-gi khyedrang-la [su -s gare gzigs-pa]

he ERG you DAT who ERG what buy that

bka’’dri- gnang- pa- red?

question do-HON PAST AGR

‘What did he ask you who bought?’

Brody (1995b) defends another possible way of accounting for the distinction between (1) and (2). This approach effectively involves saying that Subjacency need only be obeyed once per wh-comp; that is, in (1), the local movement of *who* to the matrix Comp satisfies Subjacency, rendering all subsequent wh-movements to that site (for instance, the movement of *what*) immune to Subjacency. I will refer to this as the **Subjacency tax approach**, the idea being that a well-formed movement must first pay a “Subjacency tax”, after which movement is free of Subjacency. In Chapter 5 we will see evidence for a general principle of grammar from which Subjacency tax phenomena, along with a number of other similar phenomena, may be derived.

Evidence for such a Subjacency tax approach comes from the contrasts in (4) and (5):

¹ Both of these examples, irrelevantly for our purposes, also have a reading in which both wh-words have embedded scope.

Japanese (Watanabe 1992)

- (4) a.?? John-wa [Mary -ga nani -o katta ka dooka] siritagatte-iru no?
 John TOP Mary NOM what ACC bought whether know-want Q
 ‘What does John want to know whether Mary bought?’
- b. John-wa [Mary-ga nani -o katta ka dooka] dare -ni tazuneta no?
 John TOP Mary NOM what ACC bought whether who DAT asked Q
 ‘Who did John ask whether Mary bought what?’

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c.)

- (5) a. * Koja kniga; otreče senatorăt [mălvata će iska da zabrani t_i]?
 which book denied the-senator the-rumor that wanted to ban
 ‘Which book did the senator deny the rumor that he wanted to ban?’
- b. ? Koj senator koja kniga; otreče [mălvata će iska da zabrani t_i]?
 which senator which book denied the-rumor that wanted to ban
 ‘Which senator denied the rumor that he wanted to ban which book?’

(4a) and (5a) are both Subjacency violations. In the (b) sentences we can see that adding an additional wh-word outside the Subjacency island improves the structure in both cases, as we expect on a Subjacency tax approach; the additional wh-word “pays the Subjacency tax”, permitting a second, Subjacency-violating wh-movement which would be ill-formed in isolation.

The Japanese data in (4) are the main motivation for Watanabe’s (1992) postulation of invisible overt movement. Operating under the levels approach, Watanabe took the parallelism between the data in (4) and the English facts in (1-2) as evidence that the timing of wh-movement was the same in both cases; in particular, he claimed that the wh-movement in (4a) is in fact overt, although it has no effects on the phonological representation. It is unclear, however, how this approach could be extended to deal with the Bulgarian data in (5). Here the problem for the levels approach is the opposite of that raised by the Japanese data in (4); if we persist in the assumption that Subjacency

constrains all and only overt movements, then we must conclude that movement of *koja kniga* ‘which book’ in (5b) is in fact covert, although it does have visible effects on the phonology. This move threatens to render the levels approach vacuous, however. The main virtue of the levels approach is that it makes a prediction about which movements should have to obey Subjacency; they should be all and only the *wh*-movements which have an effect on the phonology. In (4) and (5) we can see both of the logically possible types of counterexample to this prediction; there are covert movements which must obey Subjacency, like that in (4a), and overt movements which may disobey Subjacency, like that in (5b). The levels approach has nothing straightforward to say about these cases.

Neither the levels approach nor the Subjacency tax approach deals easily with the contrast between Japanese and Chinese, shown again in (6):

Japanese (Watanabe 1992)

- (6) a. ?? John-wa [Mary -ga nani -o katta ka dooka] siritagatte-iru no?
 John TOP Mary NOM what ACC bought whether know-want Q
 ‘What does John want to know whether Mary bought?’

Chinese (Huang 1982, 267)

- b. Ni xiang-zhidao shei mai-le sheme?
 you wonder who bought what
 ‘What do you wonder who bought?’

The levels approach straightforwardly deals with Chinese; *wh*-movement is covert and therefore should be immune to Subjacency. The Subjacency tax approach can handle Japanese; in (6a) there is no *wh*-movement which is well-formed with respect to Subjacency, and the Subjacency tax is therefore not paid. Neither approach, however, can deal with the other language. What we need to find, then, is some alternative explanation either for the absence of Subjacency effects in (6b) or their presence in (6a). If the

behavior of one or the other of these languages can be convincingly explained on independent grounds, we will be in a position to choose between the two theories².

Languages which do all their wh-movement covertly, then, seem to be divisible into two classes; a class with wh-island effects (including Japanese and Korean) and a class without them (including Chinese and Tibetan). What property accounts for this distinction?

Rudin (1988) notes that languages which do all their wh-movement overtly are also divisible into a class which exhibits wh-island effects (including Serbo-Croatian and Polish) and a class which lacks them (including Bulgarian and Rumanian):

Serbo-Croatian (Rudin 1988, 459)

- (7) a. * Šta si me pitao ko može da uradi?
 what AUX-2SG me asked who can to do
 ‘What have you asked me who can do?’

Bulgarian (Rudin 1988, 457)

- b. ? Koja ot tezi knigi se čudiš koj znae koj prodava?
 which of these books wonder-2SG who knows who sells
 ‘Which of these books do you wonder who knows who sells?’

The hypothesis which will be pursued in this chapter will be that the difference between the Japanese class and the Chinese class is the same as the difference between the Serbo-Croatian class and the Bulgarian class. We will see that there are independent reasons for the Chinese class of languages to lack wh-islands, and the levels approach is therefore undermined.

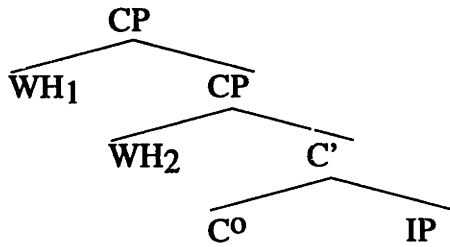
2. CP-Absorption and IP-Absorption

Adapting the proposal of Rudin (1988), I will hypothesize that multiple wh-movement can take place either by movement to multiple specifiers of CP³, as in (8a), or by movement to one or more IP projections, as in (8b).

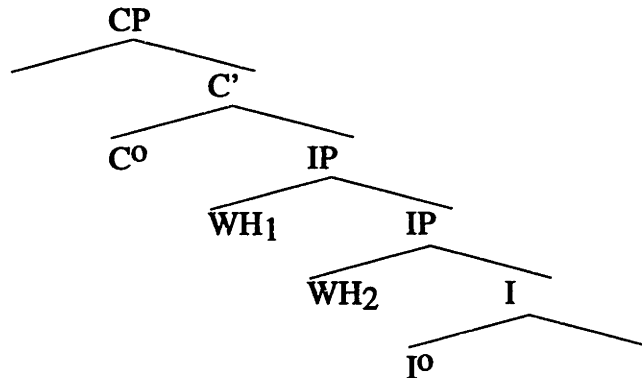
² Note that the presence of the topic marker *wa* on *John* in (6a) is not the relevant factor responsible for the ill-formedness of (6a); wh-extraction may take place across a topic, as in (4b) above.

³ Alternatively, this movement might involve multiple adjunction to CP; I have no evidence bearing on this distinction, if indeed such a distinction exists.

- (8) a. “CP-absorption” (Bulgarian, Chinese)



- b. “IP-absorption” (Serbo-Croatian, Japanese)



I will refer to the former type as “CP-absorption languages”, and to the latter as “IP-absorption languages”. Bulgarian and Chinese are CP-absorption languages; Serbo-Croatian and Japanese are IP-absorption languages.

CP-absorption languages have more or less familiar properties; wh-movement is always to a specifier of CP, and is always A-bar movement. IP-absorption languages, on the other hand, have somewhat more exotic properties. Here wh-movement most closely resembles the scrambling found in languages like Hindi (cf. Mahajan 1992) and Japanese (cf. Saito 1992); local wh-movement has certain properties of A-movement, while long-distance wh-movement uniformly acts like A-bar movement, presumably because A-chains are subject to stricter locality principles. In some IP-absorption languages, a single wh-word apparently moves obligatorily to Spec CP (Serbo-Croatian, for example, appears to be such a language, although Hungarian is not; for some further discussion of the Hungarian facts, see the Appendix). I will not speculate here on the force driving this move.

In the next section I will discuss the differences between IP-absorption languages and CP-absorption languages, and try to show that Japanese and Chinese do indeed differ in this regard.

3. Diagnostics for CP- and IP-absorption

3.1 Wh-islands

First, let us consider how the posited structures for CP- and IP-absorption languages account for the distribution of wh-island effects.

We have seen that some IP-absorption languages make use of Spec CP as a landing site for wh-movement; Serbo-Croatian apparently requires one wh-word to raise to Spec CP. In the Appendix to this chapter I will review some evidence suggesting that Spec CP is necessarily an escape hatch for long-distance wh-movement, even in IP-absorption languages such as Hungarian.

Now we are in a position to give an account of wh-islands, essentially following Reinhart (1979), Comorovski (1986), and Rudin (1988). Suppose that wh-movement past a filled Spec CP is universally barred, for familiar reasons having to do with considerations of Shortest Move. The only languages which will allow wh-movement out of a question, then, will be ones in which CP can have multiple specifiers, so that wh-movement need never skip a CP projection; there will always be a specifier of CP available as an escape hatch. In IP-absorption languages, on the other hand, it is IP which has multiple specifiers, and CP has only one. A single wh-word which has been forced to move to Spec CP, then, blocks further wh-movement past that specifier position. Thus, IP-absorption languages should exhibit wh-island effects whenever a single element occupies Spec CP, while CP-absorption languages should lack such effects.

3.2 Scrambling

The IP-absorption languages all exhibit a form of local scrambling which fixes weak crossover violations:

Serbo-Croatian (Milan Mihaljević, p.c.)

- (9) a. ?? Njegov_i susjedi ne vjeruju nijednom politicaru_i
 his neighbors not trust no politician
 ‘His neighbors trust no politician’
- b. Nijednom politicaru_i njegov_i susjedi ne vjeruju t_i
 no politician his neighbors not trust

Japanese (Saito 1992, 73)

- (10) a. ?* Soituj-no hahaoya-ga dare_i-o aisiteiru no?
 guy GEN mother NOM who ACC loves Q
 ‘Who does his mother love?’
- b. ? Dare_i-o soituj-no hahaoya-ga aisiteiru no?
 who ACC guy GEN mother NOM loves Q

Hungarian (Kiss 1994, 22)

- (11) a.* Nem szeret az pro_i anyja mindenkit_i
 not loves the mother-his everybody-ACC
 ‘His mother does not love everybody’
- b. Nem szeret mindenkit_i az pro_i anyja
 not loves everybody-ACC the mother-his

CP-absorption languages, on the other hand, apparently lack such a form of scrambling; scrambling is either absent entirely or is A-bar movement⁴:

⁴ This may well be related to the lack of overt Case morphology in Bulgarian and Chinese, as opposed to the comparatively rich Case morphology in IP-absorption languages like Japanese and Serbo-Croatian (however, Tibetan, which appears to be a CP-absorption language, has robust Case inflection).

Bulgarian (Roumyana Slabakova, p.c.)

- (12) a. * Majka mu običa vsekí čovek
 mother his love every person
 'His_i mother loves everyone_i'
- b. * Vsekí čovek običa majka mu
 every person love mother his

Chinese (Hooi Ling Soh, p.c.)

- (13) a. * Tade mama ai meigeren
 his mother love everyone
 'His_i mother loves everyone_i'
- b. * Meigeren tade mama ai
 everyone his mother love

The pattern, then, seems to be that all and only languages which allow local A-scrambling are IP-absorption languages; this is true regardless of the level on which wh-movement occurs. On the assumption that both IP-absorption and local A-scrambling involve either adjunction to IP or movement into multiple specifiers of IP, this result has a certain intuitive appeal; if a language allows this kind of movement, it uses it both for scrambling and for wh-movement, and if not, neither scrambling nor IP-absorption will be found.

3.3. Superiority

Rudin (1988) notes that in languages like Bulgarian (CP-absorption languages, in this theory), the ordering of fronted wh-phrases is subject to a rigid ordering, which she attributes to Superiority:

Bulgarian (Rudin 1988, 472-473)

- (14) a. Koj kogo vižda?
 who whom sees
 ‘Who sees whom?’
- b.* Kogo koj vižda?
 whom who sees

In an IP-absorption language like Serbo-Croatian, on the other hand, Superiority effects obtain only for long-distance movement, and not for local movement, as Bošković (1995c) points out:

Serbo-Croatian (Bošković 1995c, 5-6, 8)

- (15) a. Ko je koga vidjeo?
 who AUX whom seen
 ‘Who saw whom?’
- b. Koga je ko vidjeo?
- (16) a. Ko si koga tvrdio da je istukao?
 who AUX whom claimed that AUX beaten
 ‘Who did you claim beat whom?’
- b. * Koga si ko tvrdio da je istukao?
 whom AUX who claimed that AUX beaten

The sense in which these restrictions on ordering may be attributed to Superiority is not a straightforward one; see the next chapter for more detailed discussion. For our purposes it is sufficient to note that the differences between Bulgarian and Serbo-Croatian are accounted for by the theory developed here, assuming that Superiority constrains A-bar movement but not A-movement⁵. All Bulgarian wh-movements are A-bar movements, being movements to Spec CP; in Serbo-Croatian, on the other hand, wh-movement is adjunction to an IP-level projection, and may be an A-movement if it is sufficiently local.

⁵ In fact, as we will see in the next chapter, this is an oversimplification. However, it will remain the case that A-movement has a special status with respect to superiority effects.

The lack of Superiority effects for local movement in Serbo-Croatian (that is, in IP-absorption languages) therefore follows.

Interestingly, a similar asymmetry between local and long-distance movement seems to be present in the LF-moving IP-absorption languages. Japanese Anti-superiority, like the Superiority effects in Serbo-Croatian, is stronger (for some speakers) with long-distance movement than it is with local movement:

Japanese (Minoru Fukuda, Shigeru Miyagawa, p.c.)

- (17) a. John -ga nani -o naze katta no?

John NOM what ACC why bought Q

‘What did John buy why?’

- b. ? John-ga naze nani -o katta no?

John NOM why what ACC bought Q

- (18) a. Mary -ga [John -ga nani -o naze katta to] omotteiru no?

Mary NOM John NOM what ACC why bought that thinks Q

‘What did Mary think John bought why?’

- b. * Mary-ga [John -ga naze nani -o katta to] omotteiru no?

Mary NOM John NOM why what ACC bought that thinks Q

Thus, the Serbo-Croatian and Japanese equivalents of Superiority seem to behave similarly, in that they constrain only long-distance movement; according to the story developed here, this is because only long-distance movement has A’-properties in these languages. The prediction of this account would be that Chinese Superiority, like Bulgarian Superiority, would be equally strong locally and long-distance. Chinese word order is too rigid to test this; no alternatives parallel to those in (17-18) can be constructed. On the other hand, in Tibetan, another LF-moving CP-absorption language, word order is freer, and we find strong local Superiority effects:

Tibetan (Ngawang Jorden, p.c.)

- (19) a. Bkrashis-lags -gi gyag garebyadnas gzigs-gnang-pa -red?
 Tashi HON ERG yak why buy -HON -PAST -AGR
 ‘Why did Tashi buy a yak?’
- b. Bkrashis-lags -gi garebyadnas gyag gzigs-gnang-pa -red
 Tashi HON ERG why yak buy HON PAST AGR
- (20) a. Bkrashis-lags -gi gagi garebyadnas gzigs-gnang-pa -red
 Tashi HON ERG which why buy HON PAST AGR
 ‘Why did Tashi buy what?’
- b. * Bkrashis-lags -gi garebyadnas gagi gzigs-gnang-pa -red
 Tashi HON ERG why which buy HON PAST AGR

(19a-b) show that scrambling of *garebyadnas* ‘why’ over the direct object is possible in principle, but (20a-b) show that it is impossible if the direct object is itself a wh-word. Thus, the Tibetan equivalent of Japanese Anti-superiority strongly constrains local movement, as we expect on the hypothesis that Tibetan is like Bulgarian in that all wh-movement is A-bar movement to a Spec CP position.

3.4. Weak Crossover

Another asymmetry between local and long-distance movement in IP-absorption languages appears in the domain of weak crossover. CP-absorption languages like Bulgarian have weak crossover effects of a fairly familiar kind:

Bulgarian (Roumyana Slabakova, p.c.)

- (21) a. Koj običa majka si?
 who loves mother his
 ‘Who_i loves his_j mother?’
- b. * Kogo običa majka su?
 who loves mother his
 ‘Who_i does his_j mother love?’

In IP-absorption languages like Hungarian and Serbo-Croatian, on the other hand, weak crossover effects are found only long-distance, not locally:

Hungarian (Kiss 1987, 208, and Brody 1995a)

- (22) a. Ki szereti az anyját?
 who loves the mother-his-ACC
 ‘Who_i loves his_j mother?’
- b. Kit szeret az anyja?
 who-ACC loves the mother-his
 ‘Who_i does his_j mother love?’
- c. * Kit gondol az anyja hogy Mari szeret?
 who-ACC thinks the mother-his that Mary loves
 ‘Who_i does his_j mother think that Mary loves?’

Serbo-Croatian (Snjezana Kordić, p.c.)

- (23) a. Tko voli svoju majku?
 who loves his-ACC mother-ACC
 ‘Who_i loves his_j mother?’
- b. Koga voli njegova majka?
 who loves his-NOM mother-NOM
 ‘Who_i does his_j mother love?’
- c. * Koga njegova majka misli da Marija voli?
 who his-NOM mother-NOM thinks that Maria loves
 ‘Who_i does his_j mother think that Mary loves?’

Again, this is what we expect on a theory which likens IP-absorption wh-movement to scrambling of the type found in languages such as Japanese and Hindi. As is well known, scrambling of this kind can remedy weak crossover violations just in case it is local:

Hindi (Mahajan 1990, 26, 41)

- (24) a. sab -ko_j unkii_j bahin pyaar kartii thii
 everyone_z-ACC their sister loves do-IMP-FEM be-PAST-FEM
 ‘Everyone_j, their_j sister loves’
- b. *sab -ko_j uskii_j bahin -ne socaa [(ki) raam-ne dekhaa]
 everyone-ACC his sister-ERG thought (that) Ram-ERG saw
 ‘Everyone_j, his_j sister thought that Ram saw’

A surprising fact, given the theory developed here, is the presence of weak crossover effects in IP-absorption languages like Japanese:

Japanese (Saito 1992, 73)

- (25) ?* Soitu_j-no hahaoya-ga dare_j -o aisiteru no?
 guy GEN mother NOM who ACC love Q
 ‘Who does his mother love?’

It is not clear why *dare* cannot adjoin to IP at LF in a position higher than the pronominal variable it binds, thus obviating the weak crossover violation. One possible answer will be outlined in the next section.

3.5. Wh-movement and QR

IP-absorption, as developed here, is syntactically reminiscent of QR, in that it involves multiple adjunction to IP in order to establish scope relations. In some languages in which IP-adjunction occurs overtly, movement does indeed seem to have effects on scope relations:

Hungarian (Kiss 1994, 71)

- (26) a. Mindenki két lányt is meg táncoltatott
 everyone two girl-ACC even PREV danced
 ‘Everybody danced with two (potentially different) girls’
- b. Két lányt is mindenki meg táncoltatott
 two girl-ACC even everyone PREV danced
 ‘Two girls (the same two girls) were danced with by everybody’

Japanese (adapted from Kuroda 1971)

- (27) a. Dareka -ga daremo -o aisiteru
 someone NOM everyone ACC loves
 ‘Someone loves everyone’ ($\exists > \forall, * \forall > \exists$)
- b. Daremoj -o dareka -ga t_j aisiteru
 everyone ACC someone NOM loves
 ‘Someone loves everyone’ ($\exists > \forall, \forall > \exists$)

Furthermore, IP-adjoined wh-words in Hungarian demonstrably occupy a position which is also used as a landing site by a certain class of quantificational elements; such quantifiers have their scopes determined by movement to this position (see Kiss 1987, 1994 for discussion). In (28) we see that both wh-movement and this form of overt QR trigger inversion of the verb with a preverb, a standard test for occupying this position:

Hungarian (Kiss 1994 (37, 64))

- (28) a. Ki hívta fel Jánost?
 who called PREV John-ACC
 ‘Who called up John?’
- b. János kevés fogást kóstolt meg
 John few dish-ACC tasted PREV
 ‘John tasted few dishes’

Japanese and Chinese are both “rigid scope” languages; the scope of quantifiers is apparently entirely determined by their surface position, so that (29a-b) are both unambiguous, with the subject QP taking scope over the object QP.

Chinese (Aoun and Li 1993, 365)

- (29) a. (Yaoshi) yige ren piping meigeren...
 if one man criticize everyone
 ‘(If) someone criticized everyone...’

Japanese (Aoun and Li 1993, 365)

- b. Dareka -ga daremo -o semeta
 someone NOM everyone ACC criticized
 ‘Someone criticized everyone’

According to the theory developed here, Chinese and Japanese differ in that Japanese uses the same syntactic mechanism, IP-adjunction, to assign scope to quantifiers and to wh-words, while Chinese uses two different syntactic mechanisms: IP-adjunction and substitution to Spec CP. Interestingly, the “rigid scope” property of quantifiers is extended to wh-words in Japanese, but not in Chinese, as Aoun and Li (1993) point out; the ill-formedness of (29b) may be attributed to the inability of the wh-word to take scope over the other operators in the sentence:

Chinese (Aoun and Li 1993, 366)

- (30) a. Meigeren dou maile shenme?
 everyone all bought what
 ‘What did everyone buy?’

Japanese (Aoun and Li 1993, 366)

- b. * Daremo -ga nani -o kaimasita ka?
 everyone NOM what ACC bought Q
 ‘What did everyone buy?’

Although I have no account of “rigid scope” to offer, it seems clear that the theory developed here makes the difference between Chinese and Japanese look less surprising; the generalization, apparently, is that LF IP-adjunction in these languages cannot result in a change of scope relations⁶.

Rigid scope might also be responsible for the ill-formedness of (25), repeated as (31):

Japanese (Saito 1992, 73)

- (31) ?* Soitu_j-no hahaoya-ga dare_j-o aisiteru no?
 guy GEN mother NOM who ACC love Q
 ‘Who does his mother love?’

Whatever our eventual account of scope rigidity might be, we might expect it to entail that *dare* is unable to bind any variables at LF which it cannot bind in the overt syntax; the surprising ill-formedness of (31) would then follow.

3.6 Interacting wh-dependencies

Another interesting contrast between IP-absorption languages and CP-absorption languages has to do with the treatment of intersecting wh-dependencies. Suppose we consider a structure in which two wh-words begin in a single clause, and each could in principle move to the specifier of either of two C⁰s:

- (32) [CP [CP wh wh]]

In English, as discussed at length in Pesetsky (1982) and much subsequent literature, the only possible output of this structure, if both C⁰s bear a [+wh] feature, will involve nested paths; this is Pesetsky’s (1982) Path Containment Condition, the effects of which are exemplified in (33):

⁶ For much further discussion of this effect in Japanese, cf. Tanaka (in progress)

- (33) a. Which violin did you ask which sonata to play t on t?
-
- b. * Which sonata did you ask which violin to play t on t?
-

This is the behavior of a language in which only a single wh-phrase moves overtly and the rest move covertly. The behavior of languages which move all their wh-phrases at the same point in the derivation is quite different, as we will see. I will discuss the reasons for these differences at some length in chapter 5, section 2.6.2.2; this section will merely serve to catalog the facts, and to show that IP-absorption and CP-absorption languages pattern together in this regard as well, regardless of whether movement is overt or covert..

Let us first consider the overt IP-absorption language Serbo-Croatian. Bošković (1995c) convincingly demonstrates that the strongly preferred options in this language, given a structure like (32), are those in which all wh-words take the same scope. Thus, as we have already seen, Serbo-Croatian differs from Bulgarian in not tolerating wh-island violations:

Serbo-Croatian (Rudin 1988, 459)

- (34) * Šta si me pitao ko može da uradi?
- what AUX-2s me asked who can to do
- ‘What have you asked me who can do?’

Following Reinhart (1979), Comorovski (1986), Rudin (1988), and much subsequent work, I have analyzed the ill-formedness of (34) as diagnostic of the fact that Serbo-Croatian only allows for a single specifier of CP. In (34), then, the lower Spec CP is filled by *ko* ‘who’ at the point at which *šta* ‘what’ moves to the matrix Spec CP; because Serbo-Croatian only allows a single specifier of CP, *šta* ‘what’ cannot land in an intermediate Spec CP, and a wh-island violation results. For some speakers of Serbo-Croatian, all violations of the wh-island condition are equally ill-formed. Others find a contrast between

wh-island violations with nested wh-movement paths, like (34), and wh-island violations with crossing paths, which are slightly better:

Serbo-Croatian (Željko Bošković, Milan Mihaljević, p.c.)

(35)?* Ko si me pitao šta može da uradi?

who AUX-2s me asked what can to do

‘Who have you asked me what can do?’

On the other hand, it is clearly possible in Serbo-Croatian to move multiple wh-words long-distance, as in (36);

Serbo-Croatian (Bošković 1995c, 8)

(36) Ko si koga tvrdio da t je istukao t

who AUX whom claimed that AUX beaten

‘Who did you claim beat whom?’

Thus, we must apparently relax the claim outlined above to allow for multiple specifiers of CP in Serbo-Croatian just in case none of these specifiers will be interpreted as a scope position of a wh-word; multiple Specs of CP are licensed in Serbo-Croatian, in other words, as long as they are all used only as intermediate landing sites, as in (36)⁷.

Now let us turn to the facts in Bulgarian, a CP-absorption language. As we have seen, the availability of multiple specifiers of CP obviates wh-island effects in this language. At first sight, it would appear that Bulgarian exhibits no effect of the Path Containment Condition. The pairs in (37)-(38), for instance, consist of sentences which are equally acceptable to many speakers^{8,9}:

⁷ Recall that the matrix Spec CP in (35), by hypothesis, is occupied only by a single wh-word; *koga* ‘whom’ is adjoined to an IP-level projection.

⁸ Most Bulgarian wh-island violating cases in the literature obey the PCC, although there are a few exceptions; these include (i) (Bošković 1995c, 5) and (ii) (Legendre et al 1994, 24):

(i) čoveka kojto; ne znaeš kakvoj t; e kupil t;

the-man who not you-know what AUX bought

(ii) ?Koj student; se čudiš kakvoj t; e napisal t;?

which student self wonder-2SG what AUX written

⁹ There are also speakers who accept only the (b) examples; this pattern will be discussed immediately below.

Bulgarian (Roumyana Izvorski, Ani Petkova, Virginia Savova, Roumyana Slabakova, p.c.)

- (37) a. *Koja kniga te popita učitelja kogo ubedi Ivan t da publikuva t?*
 which book you asked teacher who convinced Ivan to publish
 ‘Which book_i did the teacher ask you who_j Ivan convinced t_j to publish t_i?’

- b. *Koj izdatel te popita učitelja kakvo ubedi Ivan t da publikuva t?*
 which publisher you asked teacher what convinced Ivan to publish
 ‘Which publisher_j did the teacher ask you what_i Ivan convinced t_j to publish t_i?’

- (38) a. *Koj kontinent te popita učitelja koj t e otkril t?*
 which continent you asked teacher who AUX discovered

- b. *Koj otkrivatel te popita učitelja kakvo t e otkril t?*
 which explorer you asked teacher what AUX discovered

However, there is a potentially confounding factor in these cases. As we saw before, Bulgarian exhibits a strict Superiority effect for both local and long-distance wh-movement. However, this effect is considerably more strict for non-D-linked wh-words than it is for D-linked wh-words:

Bulgarian (Roumyana Izvorski, p.c.)

- (39) a. Koj kakvo t e napisal t?
 who what AUX wrote
 ‘Who wrote what?’
 b. *Kakvo koj t e napisal t?
 what who AUX wrote
- (40) a. Koj aftor koja kniga t e napisal t?
 which author which book AUX wrote
 ‘Which author wrote which book?’
 b. ?Koja kniga koj aftor t e napisal t?
 which book which author AUX wrote

Although a detailed discussion of these facts will have to wait until the next chapter (see section 6.2, in particular), it seems reasonable to expect that the comparative well-formedness of (40b) might have something to do with the seeming optionality in (37)-(38), especially given that the latter sentences all involve D-linked wh-words. Suppose we consider the interaction of non-D-linked wh-words. Here speakers differ; for some speakers, D-linked and non-D-linked wh-words have identical behavior, while others get the contrast in (41):

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, Kamen Stefanov, p.c.)

- (41) a. Koj se opitvat da razberat kogo t e ubil t?
 who SELF try to find-out whom AUX killed
- b.* Kogo se opitvat da razberat koj t e ubil t?
 whom SELF try to find-out who AUX killed
-

Here we seem to get the opposite of the English PCC effect; crossing paths are preferred over nested ones. I will discuss the reasons for this effect in chapter 5, section 2.6.2.2.

For the time being, we can simply note that as far as overt movement is concerned, multiple wh-dependencies appear to interact in very different ways in CP-absorption and IP-absorption languages. In the IP-absorption language Serbo-Croatian, multiple wh-words cannot move from a single clause to specifiers of distinct CPs. In the CP-absorption language Bulgarian, on the other hand, we see an anti-PCC effect; the preferred option for multiple wh-movement, for some speakers, involves crossing paths¹⁰. Now let us move on to consider the covert-movement languages.

The claim developed here has been that Japanese, like Serbo-Croatian, is a IP-absorption language; thus, we expect that multiple wh-words which all come from the same clause will be unable to take distinct scopes in Japanese, as in Serbo-Croatian. This appears to be the case:

¹⁰ This state of affairs has an interesting parallel in the domain of long-distance anaphora. Languages with long-distance anaphors may be divided into two classes. In the first class, which includes Chinese (Hooi Ling Soh, p.c.), Japanese (Howard and Howard 1976), and Korean (Fiengo and Kim 1980), multiple clausemate long-distance anaphors must all have the same antecedent:

Japanese (Howard and Howard 1976)

- (i) Taroo-wa [Hanako -ga zibun -no heya -de zibun-no sigoto -o siteita to] itta
 Taroo TOP Hanako NOM self GEN room LOC self GEN work ACC do that said
 a. 'Taroo_i said that Hanako_j was doing his_j work in his_j room'
 b. 'Taroo_i said that Hanako_j was doing her_j work in her_j room'
 c. *'Taroo_i said that Hanako_j was doing his_j work in her_j room'
 d. *'Taroo_i said that Hanako_j was doing her_j work in his_j room'

In another class of languages, which includes Icelandic (Höskuldur Thráinsson, p.c.), German (Uli Sauerland, p.c.), Danish (Sten Vikner, p.c.), and Norwegian (Arild Hestvik, p.c.), this restriction does not apply:

Icelandic (Höskuldur Thráinsson, p.c.)

- (ii) Jón heldur að Haraldur se að skrifa bókina sína í herberginu sínu
 Jon thinks that Harold is writing book self's in room self's
 'Jon_i thinks that Harold_j is writing his_j/her's book in his_j/her's room'

In a movement-based theory of anaphora, these facts look similar to the wh-movement facts considered here; there is one class of languages in which all of the moving elements preferably land in a single place, and another in which no such preference exists. See Richards (1995a, 1995b) for a discussion of these facts (though not one which takes the parallel to wh-movement into account).

Japanese (Takako Aikawa, Satoshi Oku, Shigeru Miyagawa, p.c.)

- (42) Keesatu-wa [dare -ga dare -o korosita ka] sirabeteiru no?

police TOP who NOM who ACC killed Q are-investigating Q

a. 'Are the police trying to find out who killed who?'

b. ?'For which murderer x and which victim y are the police investigating x's

murder of y?'

- c. * 'Who are the police trying to find out who t killed t ?'

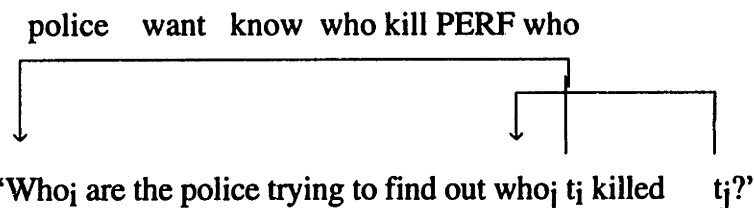
- d. ?? 'Who are the police trying to find out who t killed t?'

(42) has only two readings, not the four which would be logically possible: it has a yes-no question reading (42a), presumably derived by moving both wh-words to the lower Spec CP, and a pair-list reading (42b), which presumably involves moving both wh-words to the higher Spec CP. Single-wh-extraction readings, as in (42c) and (42d), are not as grammatical, and for some speakers are completely impossible (although of the two, (42d) is preferred for those speakers who can get a single-wh-extraction reading (Takako Aikawa, Shigeru Miyagawa, p.c.; cf. also Nishigauchi (1990, 33), Saito (1994a, 198), Grewendorf and Sabel (1996, 57) for discussion of these facts). This fact will be discussed in chapter 5, section 2.6.2.2. This parallels the situation in Serbo-Croatian, as we have seen; clausemate wh-words preferably take the same scope, although some speakers marginally allow a reading with crossing paths.

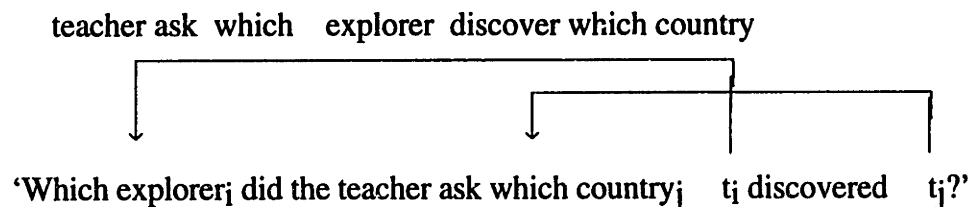
In Chinese, on the other hand, the situation is as in Bulgarian; the preferred matrix interrogative reading for sentences with this structure is a single-wh-question reading with crossing paths:

Chinese (Hooi Ling Soh, Lisa Cheng, Wei-Tien Dylan Tsai, p.c.)¹¹

- (43) a. jingcha xiang-zhidao [shei sha -le shei]



- b. laoshi wen [nayige maoxianjia faxian nayige guojia]



Some speakers, again as in Bulgarian, can get both the crossing and the nesting readings.

4. “Heterogeneous Movement” languages: German and English

The discussion thus far has centered on languages which do all their wh-movement on a single level. In this section I will discuss languages of the English type, which do one wh-movement overtly and all successive movements covertly. I will show that the division into IP-absorption and CP-absorption languages holds for these languages as well.

We have seen that the following diagnostics may be used to distinguish IP-absorption and CP-absorption languages:

¹¹ These sentences, irrelevantly for our purposes, also have a reading parallel to (42a), in which the entire sentence is a statement, with an embedded multiple-wh question. Sentences of this form are grammatical in Bulgarian as well, of course.

(44)	<u>IP-Absorption</u>	<u>CP-Absorption</u>
obeys wh-islands	yes	no
has local A-scrambling	yes	no
obeys Superiority locally	no	yes
shows local Weak Crossover effects	no	yes
wh-movement=QR	yes	no
clausemate wh-words can move to specifiers of different CPs	no ¹²	yes

Let us compare German and English, both languages in which a single wh-movement is performed overtly. The languages differ in that German, but not English, has local A-scrambling:

- (45) a. * His_i children love everyone_j
 b. * Everyone_j his_i children love t_i
- German*
- c. * ... daß seine_j Kinder jeden_i lieb haben
 that his children everyone love
- d. ...daß jeden_i seine_j Kinder t_i lieb haben
 that everyone his children love
 ‘...that his children love everyone’

According to the diagnostics in (32), then, we expect German to be an IP-absorption language, while English should be a CP-absorption language. The other diagnostics seem to confirm this. English, but not German, exhibits local Superiority effects:

¹² Again, a reading with crossing paths is marginally acceptable in both overt and covert IP-absorption languages.

- (46) a. Who bought what?
 b. *What did who buy?
German (Kim and Sternefeld 1997)
 c. Wer hat was gekauft?
 who has what bought
 d. Was hat wer gekauft?
 what has who bought

On the other hand, German does exhibit long-distance superiority effects, as expected:

German

- (47) a. Wer t glaubt, daß Hans wen gesehen hat?
 who believes that Hans whom seen has
 ‘Who believes that Hans saw who?’
 b. ??Wen glaubt wer, daß Hans t gesehen hat?
 whom believes who that Hans seen has
 ‘Whom does who believe that Hans saw?’

Similarly, English, but not German, exhibits local weak crossover effects:

- (48) a. *Who_i does his_j mother love t?

German

- b. (?) Wen_j liebt seine_j Mutter t?
 whom loves his mother
 ‘Who does his mother love?’

And again, weak crossover effects reappear in German for long-distance movement:

German

- (49) *Wen_j glaubt seine_j Mutter, daß jeder t_j liebt?
 whom believes his mother that everyone loves
 ‘Who does his mother think that everyone loves?’

Finally, as Beck (1996) points out, German exhibits restrictions on *wh*-in-situ which are reminiscent of those discussed for Japanese in section 3.5; *wh*-in-situ may not be c-commanded by quantificational operators:

German

- (50) a. *Wer hat niemanden wo angetroffen?
 who has nobody where met
 b. Wer hat wo niemanden angetroffen?
 who has where nobody met
 ‘Who didn’t meet anybody where?’

In this regard, German and Japanese appear to behave similarly; the “rigid scope” phenomenon which constrains the behavior of quantifiers in these languages is extended to *wh*-words, suggesting that *wh*-movement and QR are syntactically similar.

Thus, English and German appear to behave consistently like CP-absorption and IP-absorption languages, respectively. Of course, we do not expect to see a difference between the two languages with respect to the effect of *wh*-islands. English may have multiple specifiers of CP available to it at LF, but in the overt syntax, clearly, only a single specifier of CP can be occupied, and the escape hatch which Chinese and Bulgarian use to obviate *wh*-islands is therefore unavailable.

5. Interlude: Hungarian

Hungarian obligatorily moves all *wh*-words to a position which is preverbal but follows an overt complementizer (as well as another preverbal position, often occupied by the subject). Compare (51) and (52):

Hungarian (Horvath 1986, 54 and 67)

- (51) a. Mari az asztalra tette az edényeket
 Mary the table-onto put the dishes-ACC
 ‘Mary put the dishes on the table.’
 b. *Mari tette az asztalra az edényeket
 Mary put the table-onto the dishes-ACC
- (52) a. Nem tudtuk hogy Mari mit tett az asztalra
 not knew-1pl. that Mary what-ACC put the table-onto
 ‘We didn’t know what Mary had put on the table’
 b. *Nem tudtuk hogy Mari mit az asztalra tett
 not knew-1pl. that Mary what-ACC the table-onto put

In (51-52) we can see that the preverbal position, obligatorily occupied by *az asztalra* ‘onto the table’ in (51), is obligatorily occupied by the wh-word *mit* ‘what-ACC’ in (52); furthermore, this position is to the right of the complementizer *hogy* ‘that’. In multiple interrogation all the wh-words move to this position:

Hungarian (Kiss 1994, 38)

- (53) Mari kinek mit adott el
 Mary who-DAT what-ACC sold PREV
 ‘What did Mary sell to whom?’

Hungarian thus appears to be an IP-absorption language, although it apparently differs from the IP-absorption languages treated in Rudin (1988); in those languages, a single wh-word must apparently always occupy a [+wh] Spec CP, while the other wh-words are adjoined to IP¹³. Hungarian wh-words seem to simply adjoin to IP. I will not attempt to investigate this difference here; I assume it has to do with the strength of the wh-features on C⁰.

¹³ Persian may also be a language of this type; see Raghibdoust (1994) for discussion.

There is evidence, however, that Hungarian long-distance movement takes place via Spec CP. Local wh-movement in Hungarian triggers inversion, causing a class of particles which are usually preverbal to follow the verb:

Hungarian (adapted from Kiss 1994, 21, 37)

- (54) a. Mari felhívta Jánost

Mary PREV-called John

‘Mary called up John’

- b. Ki hívta fel Jánost?

who called PREV John

‘Who called up John?’

The mechanism driving this is unimportant for our purposes (see Horvath 1986, Kiss 1994, Brody 1995a for discussion). Interestingly, long-distance wh-movement fails to trigger inversion on verbs along the path of movement; in (55), the particles *ki* and *be* remain in their preverbal positions:

Hungarian

- (55) János melyik fiúnakj gondolta hogy Péter kijelentette

John which boy-to thought that Peter PREV-reported

hogy a házigazda már bemutatta Marit t_j?

that the host already PREV-showed Mary-ACC

‘To which boy did John think Peter declared that the host had already introduced Mary?’

This suggests that successive-cyclic wh-movement in Hungarian does not use the landing site used for local wh-movement (by hypothesis, adjunction to IP) as an intermediate landing site. Hungarian thus differs from English, for instance, which apparently uses Spec CP both for local wh-movement and as an intermediate landing site for successive-cyclic wh-movement. As Horvath (1986) suggests, Hungarian long-distance wh-movement apparently uses some landing site other than IP-adjunction as an escape hatch,

possibly Spec CP. Note that the final landing site for the long-distance move is apparently still an IP-adjoined position, as the wh-phrase *melyik fiúnak* ‘to which boy’ follows *János* ‘John’.

6. Conclusion

In this chapter I have tried to show that Rudin’s (1988) observation that multiple wh-movement languages can be classified in two types (here referred to as “CP-Absorption” and “IP-Absorption” languages) holds for languages which do wh-movement covertly as well. Briefly, the claim here has been that languages like Bulgarian and Chinese (and English) perform movement to multiple specifiers of CP, while languages like Serbo-Croatian and Japanese (and German) perform wh-movement by multiple adjunction to IP, an operation which is syntactically similar both to Japanese scrambling and to QR. Data from a variety of areas (including not only wh-island effects but also Superiority effects, interactions between wh-words and quantifiers, and the availability of local A-scrambling) seem to lend support to this claim.

To the extent that the above analysis is well-founded, it undermines the claim that Subjacency does not constrain LF movement. I have argued here that the strongest single piece of evidence for this claim, the absence of Subjacency effects in Chinese, can and should be derived in a different way. More generally, we are apparently led in the direction of a theory in which overt and covert movement are fairly similar in their distribution and properties. In particular, to the extent that Subjacency effects are taken to diagnose movement, the facts in this chapter are problematic for accounts which deny the existence of covert movement.

Many questions still remain, some of which I will attempt to answer in the following chapters. We have seen that in CP-absorption languages, and in IP-absorption languages in cases of long-distance wh-movement, wh-words must be in a particular order:

Bulgarian (Rudin 1988, 472-473)

- (56) a. Koj kogo vižda?
 who whom sees
 ‘Who sees whom?’
- b.* Kogo koj vižda?
 whom who sees

Serbo-Croatian (Bošković 1995c, 8)

- (57) a. Ko si koga tvrdio da je istukao?
 who AUX whom claimed that AUX beaten
 ‘Who did you claim beat whom?’
- b. * Koga si ko tvrdio da je istukao?
 whom AUX who claimed that is beaten

I referred to this phenomenon above as a Superiority effect, implying that it should be dealt with by the same theory responsible for the English contrast in (58):

- (58) a. Who bought what?
 b. *What did who buy?

In Chapter 3 I will investigate the nature of this phenomenon further, and will attempt to provide a natural account of the data in (56)-(58) in terms of Shortest Move.

Another question has to do with the availability of multiple long-distance wh-movement in IP-absorption languages. Following Rudin (1988), I have taken the presence of wh-island effects in languages like Serbo-Croatian to indicate that they have only a single Spec CP available as a landing site for wh-movement:

Serbo-Croatian (Rudin 1988, 459)

- (59) * Šta si me pitao ko može da uradi?
 what AUX-2SG me asked who can to do
 ‘What have you asked me who can do?’

In an example like (59), the wh-word *ko* ‘who’ occupies the single Spec CP position in the lower clause, thus triggering a wh-island effect with respect to long-distance movement of *šta* ‘what’.

On the other hand, Bošković (1995c) argues, contra Rudin (1988), that multiple long-distance wh-movement out of a non-interrogative clause is possible in Serbo-Croatian:

Serbo-Croatian (Bošković 1995c, 8)

- (60) Ko si koga tvrdio da je istukao?
 who AUX whom claimed that AUX beaten
 ‘Who did you claim beat whom?’

Given that (60) is well-formed, Serbo-Croatian must allow multiple specifiers of CP just in case none of the specifiers are occupied by wh-words at the end of the derivation; that is, multiple specifiers of a Serbo-Croatian CP can only be used as intermediate landing sites for movement, not as final landing sites. In Chapter 4 I will develop a theory of positions of this type.

Finally, we have seen evidence in this chapter for a phenomenon I referred to as the “Subjacency tax”; that is, Subjacency effects only constrain one wh-movement to a given wh-complementizer. In Chapter 5 I will discuss a number of phenomena with this character, and will propose a principle intended to deal with them.

Chapter Three: Featural Cyclicity and the Ordering of Multiple Specifiers¹

In the last chapter I sketched a number of diagnostics for distinguishing what I called IP-absorption languages from what I called CP-absorption languages. One of these diagnostics involved a family of phenomena I referred to collectively as “Superiority” effects, which were present for local movement in the CP-absorption languages but not in the IP-absorption languages. These phenomena crucially involved cases of rigid ordering of wh-words, and included the English and Bulgarian facts in (1)-(2):

- (1) a. Who t bought what?
 b.*What did who buy t?

¹ The theory developed here was independently developed in Mulders (1996, to appear), and I would like to thank Iris Mulders for much helpful discussion of the issues in this chapter.

Bulgarian (Rudin 1988, 472-473)

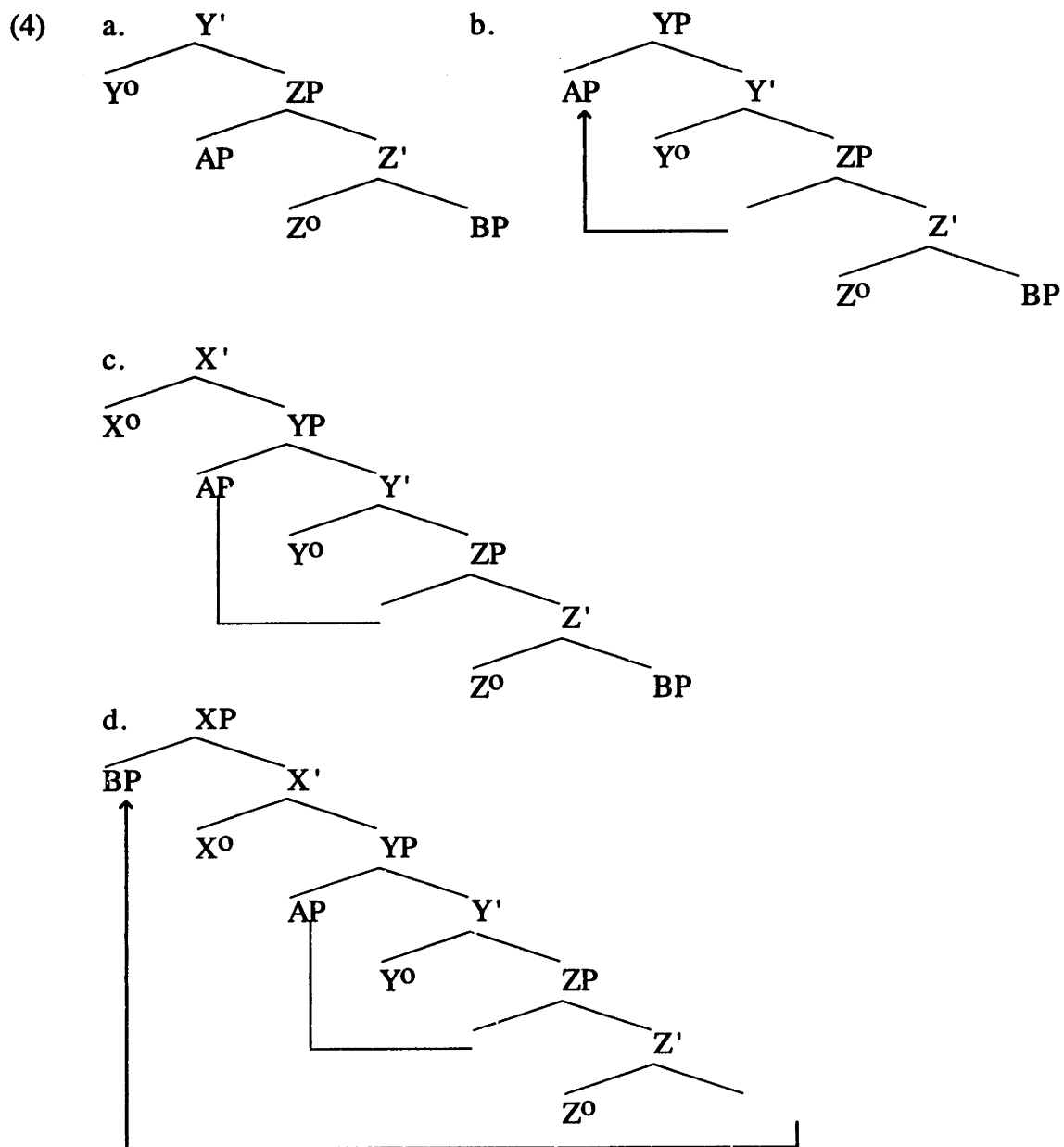
- (2) a. Koj kogo vižda?
who whom sees
'Who sees whom?'
b.* Kogo koj vižda?
whom who sees

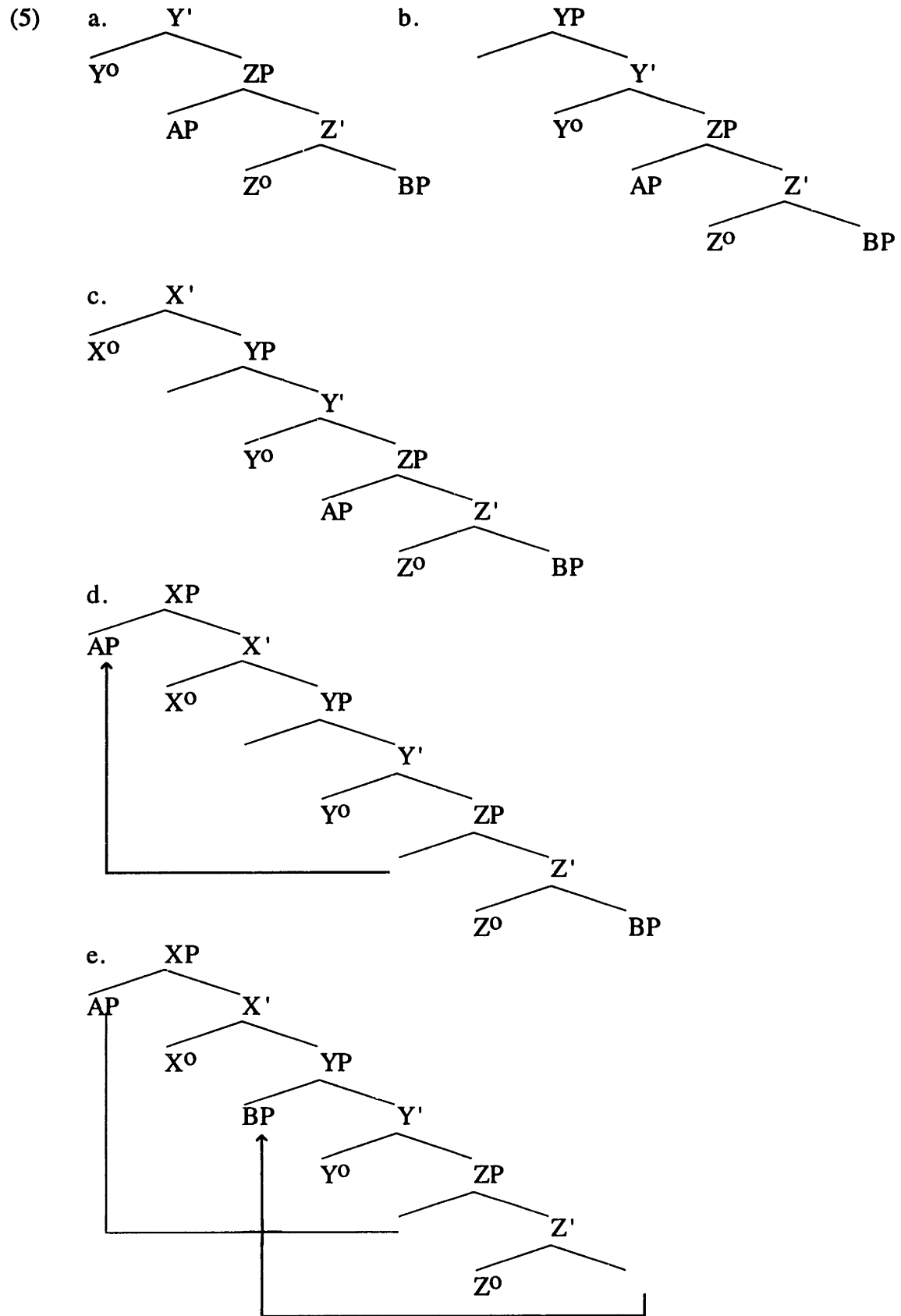
In this chapter I will discuss Superiority phenomena at greater length, defending the assumption of the previous chapter that they represent a natural class. The discussion will crucially hinge on Chomsky's (1995) notion of cyclicity, to which I now turn.

Chomsky (1995) suggests that we derive cyclicity by assuming something like the following principle:

- (3) A strong feature must be checked as soon as possible after being introduced into the derivation.

Together with the assumptions that (1) Merge always expands the tree and (2) overt movement can only take place in response to a strong feature, this principle derives cyclicity in a pleasingly minimalist way. Consider the derivations in (4)-(5):

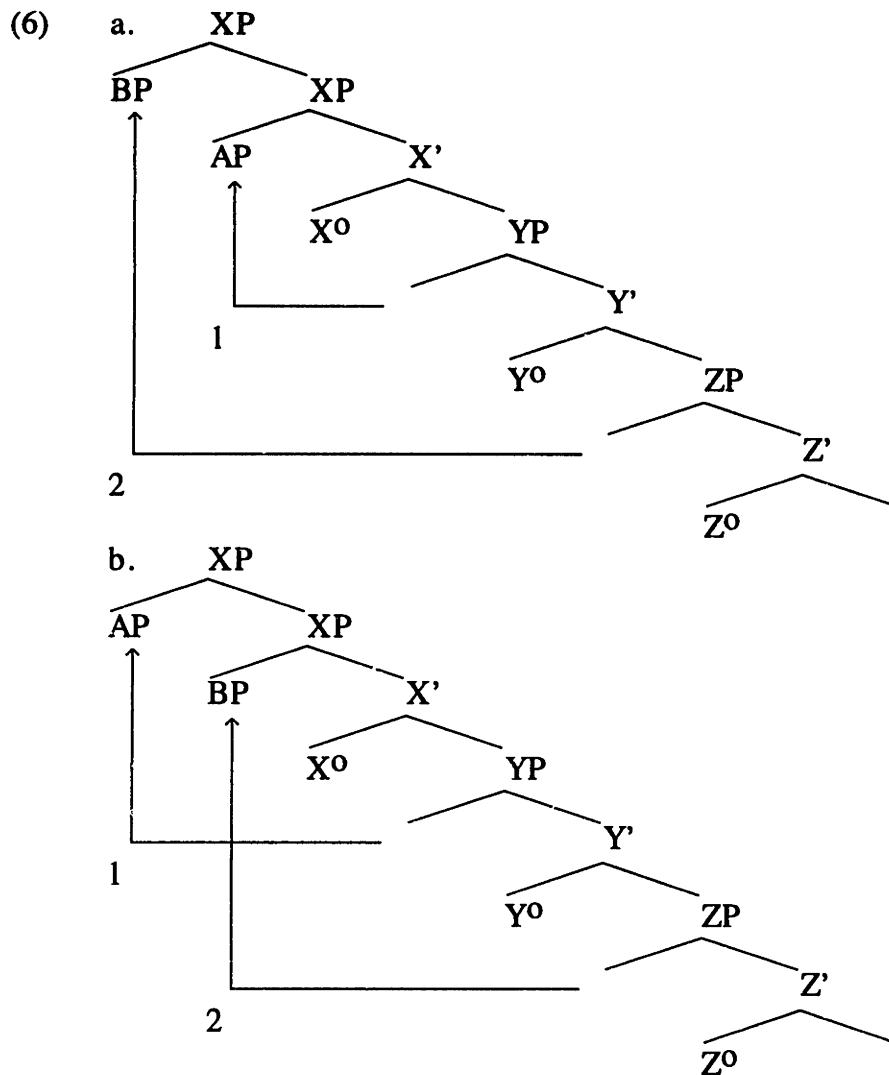




(4) and (5) both involve a ZP with specifier AP and complement BP, to which are added the heads Y^0 and X^0 , both of which bear a strong feature which might in principle attract either AP or BP. (4) is the well-formed, Cyclicity-obeying derivation; in (5), by contrast, Cyclicity is disobeyed, as BP moves to Spec YP after XP has already been projected (in step (e)). The principle in (3) correctly distinguishes between the two derivations. In the well-formed derivation in (4), the strong features introduced in the heads Y^0 and X^0 are checked off in the steps immediately after the features are introduced, as (3) requires. In (5), on the other hand, the strong feature introduced in Y^0 in step (a) is not checked off until step (e). Cyclicity thus correctly rules out the derivation. As Kitahara (1994, 1997) observes, Cyclicity, along with Shortest Move, yields the effects of Pesetsky's (1982) Path Containment Condition; intersecting paths are forced to nest, rather than cross, as we have seen.

This way of deriving cyclicity avoids problems raised by head-movement for Chomsky's (1993) definition of cyclicity. Chomsky (1993) suggested that all operations must necessarily expand the tree. This requirement successfully distinguishes between the derivations in (4)-(5), but it is always violated by head-movement, which apparently never expands the tree. The definition of cyclicity in terms of strong features, on the other hand, can be satisfied by head-movement, as long as the head-movement is checking a strong feature.

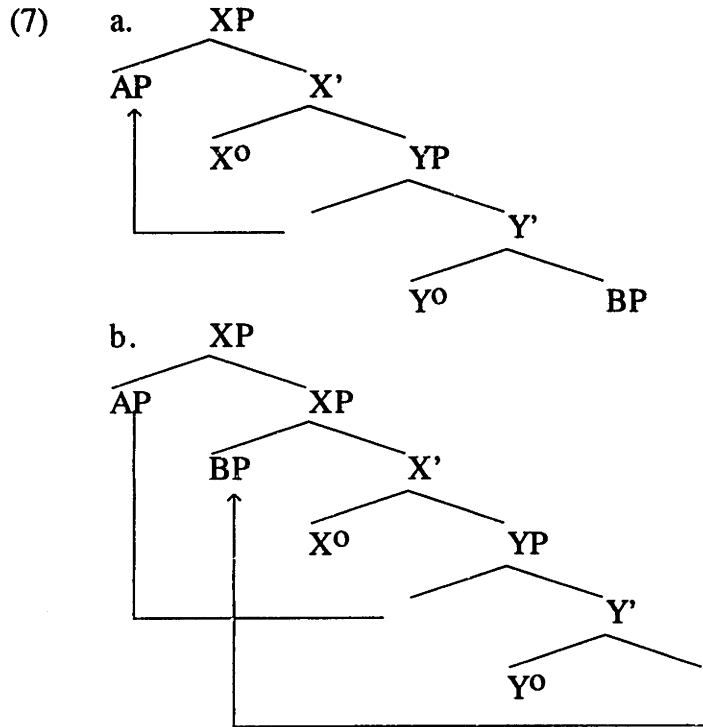
The two versions of cyclicity might also make different predictions in cases of movement to multiple specifiers of a single head. Suppose a head is generated with two strong features and attracts two XPs, as in (6):



In (6a) the two specifiers are treated just like specifiers of two separate heads; the first movement is to a specifier which is lower than the landing site of the second move. In (6b), on the other hand, the two paths cross, and the second move lands closer to the head than the first move. Chomsky's (1993) derivation of cyclicity from a requirement that every operation expand the tree would rule out the derivation in (6b); the second move here does not expand the tree. Chomsky's more recent version of cyclicity, on the other hand, fails to distinguish between the two derivations. As long as both XPs are moving to check a strong feature, either derivation ought in principle to be possible².

² The case of movement to multiple specifiers raises another potential question about this version of cyclicity. Assuming that the two movements are not simultaneous, how can a derivation in which two strong features are introduced at the same time satisfy Cyclicity at all? Whichever feature is checked first, the other feature must presumably "wait" for the first feature-checking operation to take place before it can itself be checked. One can imagine a number of ways out of this problem, which I will not try to discuss

In fact, depending on what version of Shortest Move we assume, we might expect the derivation in (6b) to be preferred over that in (6a). Consider the derivation in (6b) again, step by step:



In (7a), we first move AP into a specifier of XP. In (7b), we move BP, and must decide where to move it to; does it go to a specifier outside AP, or one inside AP? Featural cyclicity, as we have seen, allows both options. On the other hand, if the specifier inside the one containing AP is closer, in the relevant sense, to the base position of BP than the one outside AP is, then Shortest Move will require us to move to the lower specifier, as shown in (7a).

A feature-based notion of cyclicity, then, along with a certain conception of Shortest Move, predicts that multiple specifiers of a single head will be treated very differently from specifiers of multiple heads. Paths to such specifiers ought to cross, rather

here: one would be to understand the requirement that strong features be checked "immediately" as meaning that they must be checked "as soon as possible"; in particular, before any operations which do not check strong features are performed. Another approach, following Chomsky 1995 (234 ff.), would be to state the condition on strong features as one which cancels the derivation if at any point a strong feature is present on a head which is not the head of the structure in which it is contained.

than nesting, thus maintaining the base c-command relations (and, in principle, the base order) among XPs which move to them. In this paper I will argue that this is a correct result; paths to multiple specifiers of a single head do indeed cross, all other things being equal.

1. Multiple *wh*-movement

Suppose we consider the case of multiple *wh*-movement, one case in which movement to multiple specifiers is arguably involved. Multiple *wh*-movement is subject in some languages to a restriction on the order of movement; the highest *wh*-word must be moved first:

- (8) a. Who *t* bought what?
 b.?? What did who buy *t*?

This is plausibly viewed as an effect of Shortest Attract; the C^0 which attracts the *wh*-words prefers to attract *who* rather than *what*, since *who* moves a shorter distance.

Movement of *what* must then follow anyway, but on the assumption that the grammar cannot look ahead in the derivation, this is irrelevant to the choice of which *wh*-word to move first.

In certain other languages, Superiority phenomena seem to take on a rather different form. Our standard assumptions seem to make precisely the wrong predictions for *wh*-movement in certain multiple overt *wh*-movement languages. As we saw in the last chapter, Rudin (1988) shows that such languages can be divided into two groups; those which impose no ordering on multiple fronted *wh*-words (Serbo-Croatian, Polish), and those which do (Bulgarian, Romanian). For those languages which do impose such an ordering, the order essentially preserves the base c-command order:

Bulgarian (Rudin 1988, 472-473)

- (9) a. Koj kogo vižda?
 who whom sees
 ‘Who sees whom?’
- b.* Kogo koj vižda?
 whom who sees

Here we must apparently give up one of our standard assumptions. If we wish to maintain the idea that movement always expands the tree, creating a specifier higher than all the existing structure, we must apparently conclude here that a kind of "Anti-Superiority" is at work; the lower of the two wh-words (*kogo* 'whom' in (9a)) must move first. Another possibility, of course, would be to say that the order of wh-movements in this case is just as in English; *koj* 'who' moves first, followed by movement of *kogo* 'what' to a lower specifier. On this account, the paths of these multiply fronted wh-words must obligatorily cross, rather than nest.

In Japanese, we find a similar phenomenon dubbed "Anti-Superiority" by Saito (1982)³:

Japanese (Saito 1994a, 195)

- (10) a. John -ga nani -o naze katta no?
 John NOM what ACC why bought Q
 ‘Why did John buy what?’
- b. *John -ga naze nani -o katta no?
 John NOM why what ACC bought Q
 ‘What did John buy why?’

Saito points out that the contrast between (10a) and (10b) can be given an explanation based on the ECP (Chomsky 1981). On such an account, movement of *naze* 'why' must precede movement of *nani-o* 'what-ACC', in order to secure antecedent-government from

³ As we will see, speakers differ on how strong this constraint is; the judgments indicated in (10) are intended to be contrastive.

Comp; the trace of *nani-o* can then be lexically governed by the verb, and the structure is well-formed. This is referred to as “Anti-Superiority” because it apparently requires us to assume that the lower of the two wh-words is moved first; (10a), where *nani-o* ‘what’ precedes and presumably c-commands *naze* ‘why’, is the well-formed structure, in which, on Saito's theory, *naze* is able to move first. Again, this is the opposite of the order we expect on traditional versions of Cyclicity; the wh-movement paths appear to be crossing, rather than nesting, in just this case.

Thus far, the generalization appears to be that we get Anti-Superiority effects when multiple wh-words are moved to a single landing site on a single level (at LF, in Japanese, or in the overt syntax, in Bulgarian), and Superiority effects when wh-words are moved on different levels (as in English). In fact, there is some further evidence for this generalization from Japanese.

Takahashi (1993) notes that Japanese long-distance scrambling of wh-words exhibits a Superiority effect; scrambling of a wh-word over another wh-word is impossible:

Japanese (Takahashi 1993, 664)

- (11) a. John-ga [Bill -ga dare -ni [Mary -ga nani -o tabeta to]

John NOM Bill NOM who DAT Mary NOM what ACC ate that

itta to] omotteiru no?

said that thinks Q

‘Who does John think that Bill told that Mary ate what?’

- b. **Dare-ni** John-ga [Bill-ga **t** [Mary-ga nani-o tabeta to] itta to] omotteiru no?

- c. ***Nani-o** John-ga [Bill-ga dare-ni [Mary-ga **t** tabeta to] itta to] omotteiru no?

This is what our generalization leads us to expect; here, a single wh-word is being attracted in the overt syntax, so it must obey Superiority, just as in English. Multiple long-distance scrambling of wh-words, on the other hand, apparently exhibits Anti-Superiority effects again, just as we expect; the paths must cross:

Japanese (Takako Aikawa, p.c.)

- (12) a. John -ga [Tanaka-sensee -ga dare -ni nani -o yomaseta to] itta no?
 John NOM Tanaka teacher NOM who DAT what ACC read-CAUS that said Q
 ‘Who did John say Professor Tanaka made read what?’
- b. Darei -ni nanij -o John -ga [Tanaka-sensee-ga t_i t_j yomaseta to] itta no?
 who DAT what ACC John NOM Tanaka teacher NOM read-CAUS that said Q
- c. *Nanij -o darei -ni John -ga [Tanaka-sensee-ga t_i t_j yomaseta to] itta no?
 what ACC who DAT John NOM Tanaka teacher NOM read-CAUS that said Q

The examples in (13) make the same point:

Japanese (Takako Aikawa, p.c.)

- (13) a. Taroo -ga dare -ni [Hanako -ga nani -o katta to] itta no?
 Taroo NOM who DAT Hanako NOM what ACC bought that said Q
 ‘Who did Taroo tell that Hanako bought what?’
- b. **Dare-ni** Taroo-ga t [Hanako-ga nani-o katta to] itta no?
- c. * **Nani-o** Taroo-ga dare-ni [Hanako-ga t katta to] itta no?
- d. Darei-ni nanij-o Taroo-ga t_i [Hanako-ga t_j katta to] itta no?
- e. * Nanij-o darei-ni Taroo-ga t_i [Hanako-ga t_j katta to] itta no?

Here we see again that scrambling of a single wh-phrase must be of the higher of the two (13b-c), while scrambling of both must involve crossing paths (13d-e). If the theory under development here is on the right track, these examples suggest that local and long-distance scrambling must be triggered by the same attractor; in fact, a single attractor must be able to participate in both local and long-distance scrambling, as in (13d).

Thus, it is not simply the case that Japanese is an “anti-superiority language”. The choice between Superiority and Anti-Superiority is apparently a property of particular constructions, rather than of languages.

We have seen, then, that in all the cases in which the ordering of multiple wh-movements to a single landing site can be observed, the paths must apparently cross. In

cases in which a single wh-movement to a single landing site takes place, of course, the highest available wh-word must move. As was observed above, this is precisely the prediction of a featural theory of cyclicity, together with a certain definition of Shortest Move. Given these assumptions, we expect the highest wh-word to move first, followed by movement of the next highest wh-word to a specifier below the landing site of the first movement. Such a derivation does seem to give us the correct word order.

In fact, there is also evidence that the derivation proceeds in the order predicted, with the higher wh-word moving first, just as in English. Such evidence will necessarily have to involve a phenomenon which is sensitive to the order of syntactic operations. The data in (14) seem to exemplify such a phenomenon:

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c.)

- (14) a. *Koja kniga_i otreče senator_t [mălvata če pravitelstvoto iska da zabrani t_i]?
 which book denied the-senator the-rumor that the-government wanted to ban
 ‘Which book did the senator deny the rumor that the government wanted to ban?’
- b. ? Koj senator koja kniga_i otreče [mălvata če pravitelstvoto
 which senator which book denied the-rumor that the-government
 iska da zabrani t_i]?
 wanted to ban

‘Which senator denied the rumor that the government wanted to ban which book?’

(14) shows that a Complex NP Constraint violation in Bulgarian can be redeemed by moving another wh-phrase into another specifier of the same C⁰. In the last chapter, I suggested that this fact should be dealt with by the same theory that accounts for the well-known contrast between (15a) and the reading of (15b) in which *what* has wide scope:

- (15) a. *What do you wonder who bought?
 b. Who wonders who bought what?

The generalization, across both languages, seems to be that a wh-movement can disobey Subjacency if a Subjacency-obeying movement to the same position has already occurred.

Note that this phenomenon is apparently sensitive to the order of operations; the Subjacency-obeying movement must precede the Subjacency-violating movement, as (16) shows:

- (16) a. Who_i t_i persuaded the man who bought which car_j to sell the hubcaps?
 b. *Which_i car did John persuade the man who bought t_i to sell which hubcaps?

In (16a) we see another case of a well-formed movement licensing a later ill-formed movement; in this case, the later movement is that of *which car* out of the relative clause to the matrix Comp, and the well-formed move is that of *who*. In (16b), we have attempted to license an ill-formed movement with a well-formed movement again, but in this case the well-formed movement is that of *which hubcaps*, which follows the ill-formed movement⁴ of *which car*. Apparently, then, the licensing move must precede the move which it licenses. In English, of course, this will mean that the licensing move will be overt and the licensed move covert, since only the first wh-movement to a given position is overt in English. However, the Bulgarian facts, which parallel the English facts but involve only overt movements, suggest that the overt/covert distinction has nothing to do with this phenomenon.

We have seen, then, that we can use Subjacency-amelioration as a diagnostic for the order of movements; if one movement is to license another, it must occur first. Consider the Bulgarian facts again:

⁴ Here I assume, as is standard, that covert movements follow overt movements in the derivation.

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c.)

- (17) a. *Koja kniga; otreče senatorăt [mălvata če pravitelstvoto iska da zabrani tj]?
 which book denied the-senator the-rumor that the-government wanted to ban
 ‘Which book did the senator deny the rumor that the government wanted to ban?’
 b. ? Koj senator koja kniga; otreče [mălvata če pravitelstvoto
 which senator which book denied the-rumor that the-government
 iska da zabrani tj]?
 wanted to ban

‘Which senator denied the rumor that the government wanted to ban which book?’

In the well-formed (17b), the Subjacency-obeying wh-phrase *koj senator* ‘which senator’ linearly precedes the Subjacency-violating wh-phrase *koja kniga* ‘which book’. If the chain of reasoning outlined above is correct, this means that *koja kniga* must have moved to Spec CP first, followed by movement of *koja kniga* to a lower Spec CP. This is precisely what the featural notion of cyclicity predicts.

In fact, there is independent evidence from both Japanese and Bulgarian that in those languages, as in English, Subjacency-obeying movements must precede Subjacency-disobeying movements in the derivation. Consider the Subjacency violations in (18):

Bulgarian (Roumyana Izvorski, p.c.)

- (18) *Kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat [mălvata,
 what tells this official to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘What does this official tell journalists who are investigating the rumor that the
 government wants to ban that the communists have deceived their editors?’

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (19) *John -ga [Bill -ga [Mary -ga nani -o katta ka dooka] sitteita to] itta no?

John NOM Bill NOM Mary NOM what ACC bought whether knows that said Q

‘What did John say that Bill knows whether Mary bought?’

In Bulgarian, an example like (18) can be improved by overt extraction of the matrix subject:

Bulgarian (Roumyana Izvorski, p.c.)

- (20) ??Koji kakvoj kazva tj na [žurnalistite, kojto razsledvat [mālvata,

who what tells to the-journalists who investigate the-rumor

če pravitelstvoto iska da zabrani tj]],

that the-government wants to ban

če komunistite sa zabludili redaktorite im?

that the-communists AUX deceived the-editors their

‘Who tells journalists who are investigating the rumor that the government wants to ban what that the communists have deceived their editors?’

Here well-formed extraction is to a position preceding the landing site of the ill-formed extraction; by hypothesis, this indicates that well-formed extraction has preceded ill-formed extraction in the derivation.

The opposite order of operations, on the other hand, yields an ill-formed sentence:

Bulgarian (Roumyana Izvorski, p.c.)

- (21) **Kakvoj kogo_K kazva tozi služitel na [žurnalistite, kojto razsledvat [mălvata,*
what who tells this official to the-journalists who investigate the-rumor
če pravitelstvoto iska da zabrani t_j]],
that the-government wants to ban
če komunistite sa zabludili t_k?
that the-communists AUX deceived
 ‘What does this official tell journalists who are investigating the rumor that the
 government wants to ban that the communists have deceived who?’

Here ill-formed movement of *kakvo* ‘what’ is followed by well-formed movement of *kogo* ‘who’ from the object position of an embedded clause which is the complement of the matrix verb, and the result is an ill-formed sentence, as we expect.

Because the extraction sites in (21) are not in a c-command relation to each other, Superiority places no requirements on the order in which they move to the matrix Spec CP. Moving the wh-words in the opposite order improves the sentence⁵:

Bulgarian (Roumyana Izvorski, p.c.)

- (22) ??*Kogo_K kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat [mălvata,*
who what tells this official to the-journalists who investigate the-rumor
če pravitelstvoto iska da zabrani t_j]],
that the-government wants to ban
če komunistite sa zabludili t_k?
that the-communists AUX deceived
 ‘Who does this official tell journalists who are investigating the rumor that the
 government wants to ban what that the communists have deceived?’

This is what we expect, again; *kogo* is the wh-word which is not being extracted out of an island, so if it arrives in the matrix Spec CP first we expect the sentence to improve.

⁵ Judgments are apparently subtle here, but my informant says that (20) is better than (22) (Roumyana Izvorski, p.c.). I have no account of this fact to offer.

Turning to Japanese, it is clear that addition of a *wh*-word c-commanding the island from which ill-formed extraction takes place improves an example like (19), repeated here as (23):

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (23) *John -ga [Bill -ga [Mary -ga nani -o katta ka dooka] sitteita to] itta no?

John NOM Bill NOM Mary NOM what ACC bought whether knows that said Q

‘What did John say that Bill knows whether Mary bought?’

- (24) ?John -ga dare-ni [Bill -ga [Mary -ga nani -o katta ka dooka]

John NOM who DAT Bill NOM Mary NOM what ACC bought whether

sitteita to] itta no?

knows that said Q

‘Who did John tell that Bill knows whether Mary bought what?’

This is what we expect; in (24), *dare-ni* is the highest available *wh*-word, and will therefore be attracted first, paying the “Subjacency tax” and licensing later extraction of *nani-o* out of the *wh*-island.

Speakers differ on the well-formedness of sentences in which the additional *wh*-word is outside the island but does not c-command it:

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (25) %John -ga [Bill -ga [Mary -ga nani -o katta ka dooka]

John NOM Bill NOM Mary NOM what ACC bought whether

sitteita to] dare -ni itta no?

knows that who DAT said Q

‘Who did John tell that Bill knows whether Mary bought what?’

For some speakers, (25) is better than (23) but worse than (24); for others, (24) and (25) are equally good. I assume that the ambiguous status of examples like (25) has something to do with the availability of distinct derivations for such examples, corresponding to the Bulgarian examples (21) and (22). Because the *wh*-words in (25) are not in a c-command

relation to each other, they can in principle be attracted to the matrix C^0 at LF in either order. Only the order in which the Subjacency-obeying move precedes the Subjacency-disobeying move will make the sentence better than (23). In (24), by contrast, Superiority forces a derivation in which the well-formed move precedes the ill-formed move. The degraded status of (25) for some speakers might reflect a need to “search” for a well-formed derivation in this case.

Thus, it would appear that in Japanese and Bulgarian, as in English, Subjacency-amelioration effects only hold when the well-formed move precedes the ill-formed move. These data are particularly interesting in light of the fact that the opposite effect has been discussed in the literature for minimally different examples in Japanese by Watanabe (1992):

Japanese (adapted from Watanabe 1992, 270-271)

- (26) a. *John-wa [Mary -ga nani -o katta ka dooka] Tom -ni tazuneta no?

John TOP Mary NOM what ACC bought whether Tom DAT asked Q
‘What did John ask Tom whether Mary bought?’

- b. John-wa [Mary -ga nani -o katta ka dooka] dare -ni tazuneta no?

John TOP Mary NOM what ACC bought whether who DAT asked Q
‘Who did John ask whether Mary bought what?’

- c. ?John-wa dare -ni [Mary -ga nani -o katta ka dooka] tazuneta no?

John TOP who DAT Mary NOM what ACC bought whether asked Q
‘Who did John ask whether Mary bought what?’

Watanabe (1992) observes a pattern of judgments, given in (26), which is essentially the opposite of that discussed above; examples in which the additional wh-word does not c-command the island out of which ill-formed extraction is to take place (as in (26b)) are

better than those in which the additional wh-word does c-command the island (e.g., (26c))⁶.

Compare Watanabe's (1992) examples with the examples discussed above:

- (27) a. *John -ga [Bill -ga [Mary -ga nani -o katta ka dooka] sitteita to] itta no?

John NOM Bill NOM Mary NOM what ACC bought whether knows that said Q

'What did John say that Bill knows whether Mary bought?'

- b. %John -ga [Bill -ga [Mary -ga nani -o katta ka dooka]

John NOM Bill NOM Mary NOM what ACC bought whether

sitteita to] dare -ni itta no?

knows that who DAT said Q

'Who did John tell that Bill knows whether Mary bought what?'

- c. ?John -ga dare -ni [Bill -ga [Mary -ga nani -o katta ka dooka]

John NOM who DAT Bill NOM Mary NOM what ACC bought whether

sitteita to] itta no?

knows that said Q

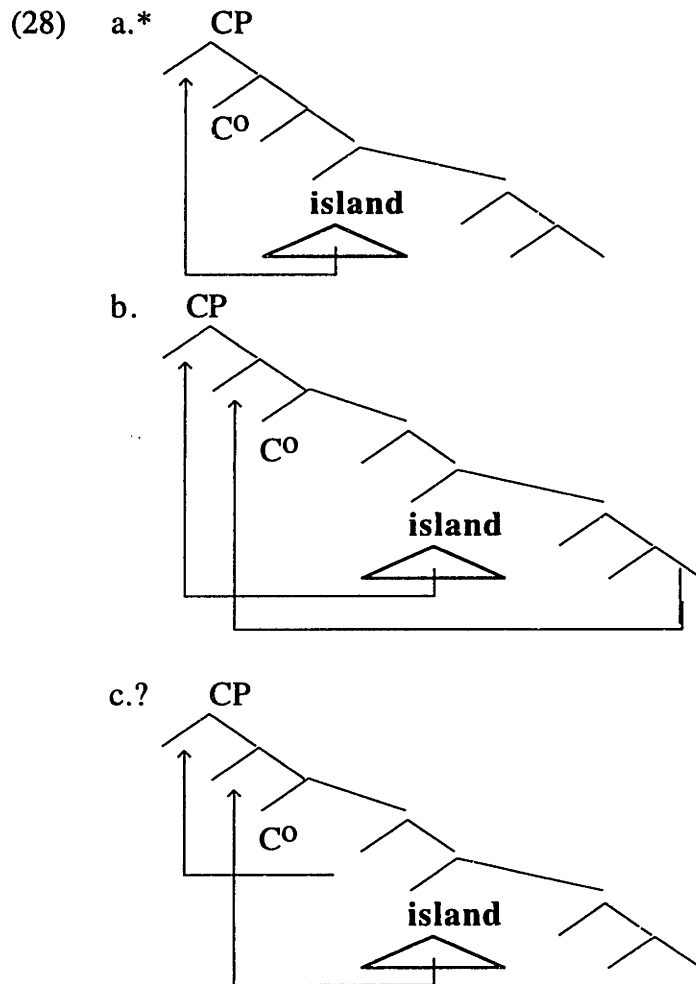
'Who did John tell that Bill knows whether Mary bought what?'

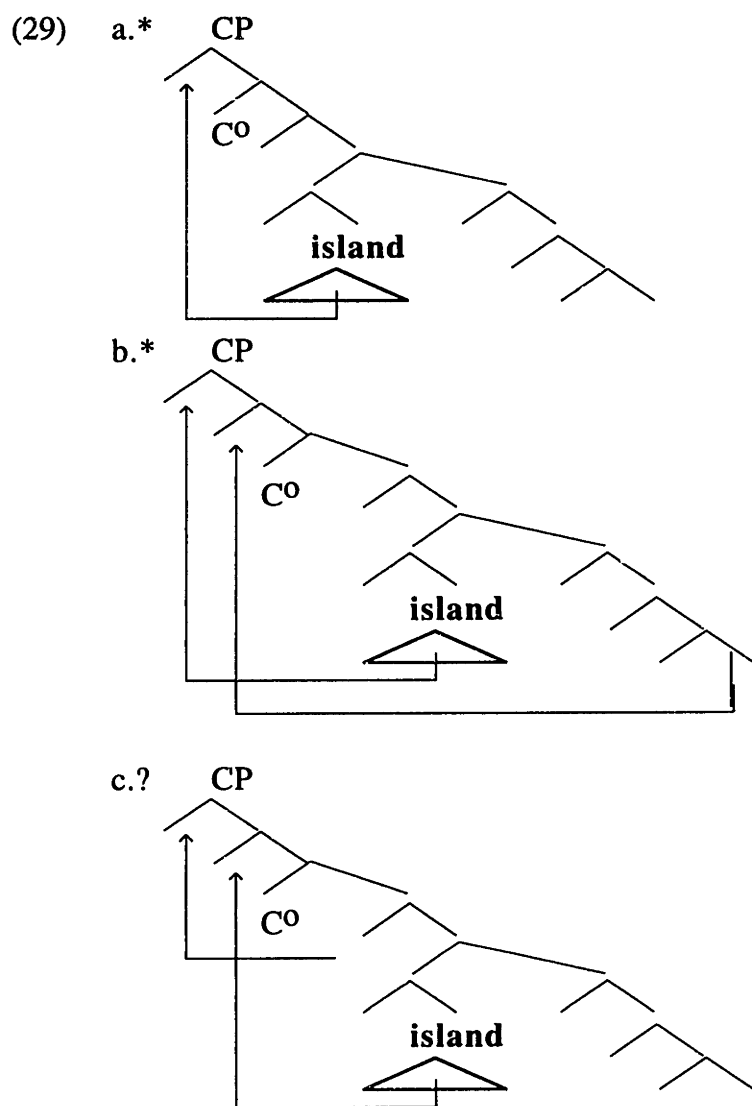
In the examples in (27), it is the (c) example, where the additional wh-word c-commands the offending island, which is uncontroversially better than the (a) example. In Watanabe's examples in (26), by contrast, the preferred example is (26b), in which the additional wh-word does not c-command the island.

There is an important structural difference between Watanabe's examples and the ones given here. In Watanabe's cases, the offending island is "along the path" of the well-formed wh-movement; that is, the island c-commands the additional wh-word's base position, and is c-commanded by the position to which the wh-word moves at LF. In my examples in (27), by contrast, the island is in an embedded clause, so that the base position

⁶ In fact, Watanabe claims that (26c) is no better than (26a), but notes that a number of speakers disagree with him on this judgment. My informants find (26c) better than (26a). See also Maki (1994, 202), who gives examples in which the "saving" wh-word c-commands the offending island.

of the additional wh-word is not in a c-command relation with the island. The facts in (26) and (27) may be diagrammed as in (28-29):





In chapter 5, section 2.6.1.1, I will argue that the unexpected well-formedness of Watanabe's example (26b) (corresponding to the structure in (28b)) has to do with a strategy for avoiding island violations which is formally identical to that employed in English parasitic gaps. As with English parasitic gaps, this strategy is only available when the offending island is along the path of the well-formed wh-movement. This is the case in (26), but not in (27). Watanabe's examples, then, illustrate the existence of a second means of dealing with islands, like that used by English parasitic gaps, which differs from the "Subjacency tax" phenomenon discussed thus far in two ways; it does not require that the well-formed wh-movement precede the ill-formed movement in the derivation, and it does impose a structural requirement that the offending island be along the path of the well-

formed wh-movement. When we consider cases in which this parasitic-gap-like strategy is unavailable because its structural requirement is not met, Japanese exhibits a “Subjacency tax” phenomenon of the familiar type, in which the well-formed move must occur first.

The parasitic-gap strategy appears in Bulgarian as well, as we can see if we consider examples in which the offending island is along the path of a well-formed wh-movement:

Bulgarian (Roumyana Izvorski, p.c.)

- (30) a. *Kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat tj],
 what tells this official to the-journalists who investigate
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘What does this official tell journalists who are investigating that the communists
 have deceived their editors?’
- b. ?Kakvoj kogož kazva tozi služitel na [žurnalistite, kojto razsledvat tj],
 what who tells this official to the-journalists who investigate
 če komunistite sa zabludili tk?
 that the-communists AUX deceived
 ‘What does this official tell journalists who are investigating that the communists
 have deceived who?’
- c. ??Koji kakvoj kazva tj na [žurnalistite, kojto razsledvat tj],
 who what tells to the-journalists who investigate
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘Who tells journalists who are investigating what that the communists have
 deceived their editors?’

As in Japanese, these facts appear to be the opposite of those illustrated by the examples discussed earlier, in which the offending islands were more deeply embedded:

Bulgarian (Roumyana Izvorski, p.c.)

- (31) a. *Kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat [malvata,
 what tells this official to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘What does this official tell journalists who are investigating the rumor that the
 government wants to ban that the communists have deceived their editors?’
- b. *Kakvoj kogož kazva tozi služitel na [žurnalistite, kojto razsledvat [malvata,
 what who tells this official to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili tk?
 that the-communists AUX deceived
 ‘What does this official tell journalists who are investigating the rumor that the
 government wants to ban that the communists have deceived who?’
- c. ??Koji kakvoj kazva tj na [žurnalistite, kojto razsledvat [malvata,
 who what tells to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘Who tells journalists who are investigating the rumor that the government wants to
 ban what that the communists have deceived their editors?’

Bulgarian, like Japanese, behaves as expected only when the offending island is deeply embedded; an island which is along the path of a well-formed wh-movement is

comparatively transparent for *wh*-movement, regardless of the relative order of the *wh*-words. As in the Japanese case, I attribute the comparative well-formedness of examples like (30b) to the availability of a mechanism like that involved in English parasitic gaps, which is only available when the offending island is along the path of a well-formed *wh*-movement but which imposes no constraints on the derivation with respect to the order of *wh*-movements. I will explore the properties of this additional mechanism further in chapter 5, section 2.6.1.1. For the time being, it suffices to point out that in Japanese and Bulgarian, as in English, Subjacency tax phenomena are apparently reliable indicators of the order of *wh*-movement, once the availability of the parasitic-gap strategy is controlled for. Furthermore, the Subjacency tax phenomena in all three languages confirm the main premise of this chapter: that *wh*-movement of the highest available *wh*-word must occur first in the derivation, before movement of lower *wh*-words, and that each *wh*-movement triggered by a given C^0 lands in a specifier of CP which is below any existing specifiers.

2. Multiple A-specifiers

We have now seen that the predictions of a theory incorporating featural cyclicity and a certain notion of Shortest Move seem to be borne out in cases of multiple *wh*-movement; we can successfully explain ordering restrictions on *wh*-movement in a variety of languages using fairly simple assumptions. Given such simple assumptions, however, we expect to find this pattern arising quite generally in all cases involving movement to multiple specifiers of a single head. Is it possible to find evidence for this theory from the domain of A-movement, for example?

(32) shows that Japanese local scrambling, which might in principle be thought to involve movement to multiple A-specifiers, does not appear to exhibit any Superiority effects. Two arguments can be scrambled over the subject in either order:

Japanese

(32) a. Taroo-ga gakusei -ni hutari hon -o ageta

Taroo NOM student DAT two book ACC gave

‘Taroo gave books to two students’

b. Gakusei- ni hutari hon -o Taroo -ga ageta

student DAT two book ACC Taroo NOM gave

c. Hon -o gakusei -ni hutari Taroo -ga ageta

book ACC student DAT two Taroo NOM gave

Of course, there may be any of a number of confounding factors preventing us from observing the effects of Superiority here; there may, for example, be more than one attracting head involved. Developing a complete theory of scrambling is well beyond the scope of this thesis. In the next sections I will simply try to show that Superiority effects do indeed hold in a certain well-defined subset of A-scrambling cases, which suggests that the general account of Superiority developed above may be on the right track. It is well-known that local scrambling interacts with such semantic properties as definiteness, specificity, and quantifier scope. In the next few sections, I will try to show that when these effects are controlled for, local scrambling does obey Superiority, as defined here. Sections 2.1 and 2.2 will show that local scrambling which has no semantic effects obeys Superiority. In section 2.3 I will give some evidence suggesting that local scrambling which violates Superiority does in fact involve multiple attractors.

2.1 Idiom chunks

One area we might want to consider is the scrambling of idiom chunks. Idiom chunks are presumably non-referential and therefore impervious to any effects scrambling might have on discourse properties. If it is these effects which are responsible for the apparent failure of A-scrambling to conform to Superiority, then, we might expect to get a clearer picture by looking at the scrambling of idiom chunks. Miyagawa (1994) notes that

scrambling of idiom chunks in Japanese is confined to short-distance scrambling; long-distance scrambling, as in (33b), is impossible:

Japanese (Miyagawa 1994, 20-21)

- (33) a. Kosi-o John -ga t orosita
 hip ACC John NOM lowered
 ‘John sat down’
- b. ?*Kosi-o Mary -ga [John -ga t orosita to] itta
 hip ACC Mary NOM John NOM lowered that said
 ‘Mary said that John sat down’

Idiom chunks, then, may apparently only be A-scrambled. It is interesting to note, therefore, that multiple idiom chunks do indeed appear to obey Superiority in their short-distance scrambling. If scrambling only affects one idiom chunk, it must raise the higher of the two:

Japanese (Shigeru Miyagawa, Kazuko Yatsushiro, p.c.)

- (34) a. Taroo -ga hi -ni abura -o sosoida
 Taroo NOM fire DAT oil ACC poured
 ‘Taroo made things worse’
- b. Hi -ni Taroo -ga t abura-o sosoida
 fire DAT Taroo NOM oil ACC poured
- c. *Abura-o Taroo-ga hi-ni t sosoida
 oil ACC Taroo NOM fire DAT poured

Both idiom chunks may also be scrambled, but the two paths must cross, in conformance with Superiority:

Japanese (Shigeru Miyagawa, Kazuko Yatsushiro, p.c.)

- (35) a. *Hi_i -ni aburaj-o Taroo -ga t_i t_j sosoida*
 fire DAT oil ACC Taroo NOM poured
 ‘Taroo made things worse’
- b. **Aburaj-o hi_i -ni Taroo -ga t_i t_j sosoida*
 oil ACC fire DAT Taroo NOM poured

Superiority, then, does not affect only A-bar movement. Furthermore, as we saw with A-bar movement, multiple A-movements to a single attractor must apparently cross and not nest, just as the account developed here predicts.

2.2 Quantifier scope

Japanese A-scrambling typically has effects on quantifier scope, creating scope ambiguities which are unavailable in the base order:

Japanese (Kuroda 1971)

- (36) a. *Dareka -ga daremo -o hihansita*
 someone NOM everyone ACC criticized
 ‘Someone criticized everyone’ ($\exists > \forall$, $*\forall > \exists$)
- b. *Daremo -o dareka -ga t hihansita*
 everyone ACC someone NOM criticized
 ‘Someone criticized everyone’ ($\exists > \forall$, $\forall > \exists$)

However, Kazuko Yatsushiro (1996, and p.c.) has noted that there is in fact a type of local scrambling of quantifiers which has no effect on scope relations. This is precisely the scrambling which obeys Superiority, as defined in this paper:

Japanese (Yatsushiro 1996 and p.c.)

- (37) a. John-ga dareka -ni daremo -o syookaisita
 John NOM someone DAT everyone ACC introduced
 ‘John introduced everyone to someone’
unambiguous: someone>everyone
- b. Dareka -ni John -ga t daremo -o syookaisita
 someone DAT John NOM everyone ACC introduced
unambiguous: someone>everyone
- c. Dareka_i -ni daremo_j -o John -ga t_i t_j syookaisita
 someone DAT everyone ACC John NOM introduced
unambiguous: someone>everyone

Thus, when scrambling affects the higher of two quantifiers, as in (37b), or affects both quantifiers but preserves the underlying c-command relation between them, as in (37c), the sentence remains unambiguous. When Superiority is disobeyed, by contrast, ambiguity is created:

Japanese (Yatsushiro 1996 and p.c.)

- (38) a. Daremo -o John -ga dareka -ni t syookaisita
 everyone ACC John NOM someone DAT introduced
ambiguous: someone>everyone, everyone>someone
- b. Daremo_j -o dareka_i -ni John -ga t_i t_j syookaisita
 everyone ACC someone DAT John NOM introduced
ambiguous: someone>everyone, everyone>someone

The contrast between (37c) and (38b), in particular, is a striking one. On the account developed here, the ordering in (38b) can only be the result of attraction by multiple attractors, a conclusion for which more evidence will be adduced in section 2.3. The ordering in (37c), on the other hand, can be produced by movement to multiple specifiers of a single head. It could also, in principle, be a result of movement to specifiers of

multiple heads (say, A-movement of *daremo-o* 'everyone' followed by A-bar-movement of *dareka-ni* 'someone' into a higher maximal projection). The lack of ambiguity in (34c) suggests that this second option is in fact excluded. A number of accounts might be given for this; one possibility is that speakers assume only as many attractors as are necessary to account for a particular word order⁷. The availability of a grammatical structure for the word order in (37c) which involves only a single attracting head, then, rules out structures with more than one such attractor.

We have seen that a certain class of A-scrambling movements do in fact appear to obey Superiority: namely, scrambling of idiom chunks and local quantifier scrambling which fails to create scope ambiguity. This lends some support to the idea that A-scrambling does in principle obey Superiority, and that what create the appearance of disobedience of Superiority are processes which are sensitive to such phenomena as referentiality and quantifier scope.

2.3 Scrambling generally

In the previous sections we have seen certain cases of semantically vacuous local scrambling which apparently must obey Superiority. In this section I will argue further that scrambling with crossing paths is (or can be) the result of multiple attraction by a single attractor, while scrambling with nesting paths must involve multiple attractors.

⁷ This is reminiscent of a general constraint on interpretation of multiple NPs with identical case marking in scrambling languages (to be discussed in section 3.4 of chapter 4), which prefers interpretations which do not assume scrambling:

- German*
- (i) Die Mutter liebt die Tochter
the mother loves the daughter
'The mother loves the daughter'
*'The daughter loves the mother'
- Tagalog* (Guilfoyle, Hung and Travis 1992)
- (ii) Kakakain ng leon ng tigre
RP-ate A lion G tiger
'The lion just ate the tiger'
*'The tiger just ate the lion'
- Japanese* (Saito 1985, 190)
- (iii) John-ga Bill -ni Mary -ni hana -o todokesaseta
John NOM Bill DAT Mary DAT flower ACC deliver-CAUS
'John made Bill deliver flowers to Mary'
*'John made Mary deliver flowers to Bill'

The argument is based on a kind of relativization in Japanese which violates islands, discussed (among others) by Kuno 1973, Hasegawa 1984, and Ochi 1996:

Japanese (Ochi 1996)

(39) [[*e_i* *e_j* kiteiru] *hukuj* -ga *yogoreteiru*] *kodomoj*

wear clothes NOM dirty child

‘the child_{*i*} that the clothes_{*j*} that *t_i* is wearing *t_j* are dirty’

Here relativization of the operator associated with *kodomo* ‘child’ seems to take place from inside a relative clause modifying *fuku* ‘clothes’, in violation of the wh-island condition and the CED. Such relativization, as Hasegawa (1984) notes, is strongly constrained, at least for some speakers. Hasegawa observes that the relative clause out of which relativization takes place (which I will call, for purposes of exposition, the “contained” relative clause) must modify the subject of the relative clause by which it is contained.

Thus, (39) contrasts with (40):

Japanese (Ochi 1996)

(40) *[[*Mary-ga* [*e_i* *e_j* kiteiru] *hukuj* -o *tukutta*] *kodomoj*

Mary NOM wear clothes ACC made child

‘the child_{*i*} that Mary made the clothes_{*j*} that *t_i* is wearing *t_j*’

In fact, the conditions on the positioning of the contained relative clause are somewhat freer than this. In particular, this kind of relativization can be affected by scrambling, as well as by constructions which have been argued to involve multiple specifiers. Thus, there is a contrast between (40a-b), as well as between (41a-b):

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (41) a. *[Taroo-ga [e_i e_j kaita] honj -o katta] sakka_i

Taroo NOM wrote book ACC bought author

'the author_i that Taroo bought the book_j that t_i wrote t_j'

- b. ?[Taroo-ga [e_i e_j kaita] honj -ga suki na] sakka_i

Taroo NOM wrote book NOM likes author

'the author_i that Taroo likes the book_j that t_i wrote t_j'

- (42) a. *[Taroo-ga [e_i e_j kaita] honj -o katta] sakka_i

Taroo NOM wrote book ACC bought author

'the author_i that Taroo bought the book_j that t_i wrote t_j'

- b. ?[[e_i e_j kaita] honj -o Taroo-ga katta] sakka_i

wrote book ACC Taroo NOM bought author

'the author_i that the book_j that t_i wrote t_j, Taroo bought '

In (41), we can see that the contained relative clause can modify the object of a double-*ga* predicate; such predicates have recently (cf. Ura 1996) been analyzed as having the multiple *ga*-marked elements in multiple specifiers of a single head. Furthermore, (42) shows that scrambling can affect the acceptability of contained relative clauses; a contained relative clause can modify an direct object which has been scrambled to the front of its clause.

I will not try to develop a full theory here of the nature of this kind of relativization, but the contrast in (41) suggests that we can use contained relative clauses as a test for structure. If a non-initial argument is capable of hosting a contained relative clause, we may conclude that it and the preceding argument(s) are in multiple specifiers, as in (41b), rather than in separate maximal projections, as in (41a).

Having established this, let us turn to the distribution of contained relatives in double object constructions. On the theory proposed here, scrambling of both double objects which preserves the base c-command relation between the scrambled arguments may involve scrambling to multiple specifiers of a single head. Scrambling which alters the

base c-command relation, on the other hand, must necessarily involve movement to distinct maximal projections. The prediction, then, is that scrambling which preserves the basic order will allow contained relative clauses to modify either object, since the objects, like the *ga*-marked nominals in (41b), are in multiple specifiers of a single head. Scrambling which alters the base order, on the other hand, will only allow a contained relative clause on the first object, since the two objects must be in specifiers of different heads, like the subject and object in (41a).

In order to test this prediction, we must first determine the base order of double object constructions in Japanese, which is not a straightforward task. Miyagawa 1997a argues convincingly that both IO-DO and DO-IO word orders can be base-generated in Japanese. One of his arguments has to do with the effects in Japanese of Rizzi's (1986) Chain Condition (or whatever its successor might be; the relevant condition will need to forbid A-movement of an antecedent over the anaphor it binds. See Snyder 1992, McGinnis to appear for further discussion). Scrambling of a potential antecedent over an anaphor yields a Chain Condition violation:

Japanese (Miyagawa 1997a, 4)

- (43) *[John-to Mary]_i -o otagai_i -ga t_i mita
 John and Mary ACC each-other NOM saw
 'John and Mary, each other saw'

(43) is apparently ruled out by the Chain Condition. Note that scrambling in Japanese can remedy Condition A violations, which shows that (43) is not a violation of Condition A:

Japanese (Miyagawa 1997a, 5)

- (44) [John-to Mary]_i-o [otagai_i -no sensei] -ga t mita
 John and Mary ACC each-other GEN teacher NOM saw
 'John and Mary, each other's teachers saw'

Interestingly, the first object may bind the second object in either the IO-DO or the DO-IO order:

Japanese (Miyagawa 1997a, 5)

- (45) a. John -ga [Hanako -to Mary]_i-ni (paati-de) otagai_i -o syookaisita
 John NOM Hanako and Mary DAT party at each-other ACC introduced
 'John introduced, to Hanako and Mary, each other at the party'
- b. John -ga [Hanako -to Mary]_i-o (paati-de) otagai_i -ni syookaisita
 John NOM Hanako and Mary ACC party at each-other DAT introduced
 'John introduced Hanako and Mary to each other at the party'

This suggests that the orders in (45) are not derived from one another by movement; if they were, we would expect one of them to violate the Chain Condition, like (43).

Miyagawa argues that the distinct base-generated orders correspond to different structural realizations of the indirect object; in the IO-DO order, he says, the IO is a dative-marked NP, while in the DO-IO order it is a PP. One argument for this conclusion is the distribution of floated numeral quantifiers which, as he argues in Miyagawa 1989, can modify NPs but not PPs. These quantifiers can appear on the IO in the IO-DO order, but not in the DO-IO order⁸:

Japanese (Miyagawa 1989)

- (46) a. Mary -ga tomodati-ni hutari CD -o okutta
 Mary NOM friend DAT two CD ACC sent
 'Mary sent two friends a CD'
- b. *Mary -ga CD -o tomodati-ni hutari okutta
 Mary NOM CD ACC friend DAT two sent

Now we are in a position to test the predictions of the theory being developed here. We can see that a floated numeral quantifier on the dative argument signals an IO-DO base order. Thus, in cases in which the double objects are scrambled to the left of the subject and the dative argument is marked by a floated numeral quantifier, we expect to find that a contained relative clause can modify the second object only when the base IO-DO order is

⁸ (43b) can be improved by focussing the accusative argument.

maintained. If the order is DO-IO, then scrambling must have involved specifiers of distinct attractors, and a contained relative clause on the second object should not longer be possible.

The sentences in (47), then, form a minimal pair⁹:

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (47) a.?[seijika -ni hutari [t_i t_j kyoonen osieta] Tanaka -san -no kodomoj-o
 politician DAT two last-year taught Tanaka HON GEN child ACC
 Hanako -ga syookaisita] sensee_j
 Hanako NOM introduced teacher
 ‘the teacher_j that Hanako introduced the children_j of Mr. Tanaka that t_i taught t_j last
 year to two politicians’
- b.*[seijika -o [t_i t_j kyoonen osieta] Tanaka-san -no kodomoj-ni hutari
 politician ACC last-year taught Tanaka HON GEN child DAT two
 Hanako -ga syookaisita] sensee_j
 Hanako NOM introduced teacher
 ‘the teacher_j that Hanako introduced the politician to two of the children_j of Mr.
 Tanaka that t_i taught t_j last year’

In both cases, the contained relative clause modifies the second NP in the relative clause. In the first case, however, the base order of the scrambled double objects is preserved, and the word order is therefore compatible with movement to multiple specifiers of a single attractor; relativization into the second nominal is therefore possible. In the second case, the base order of the scrambled elements is not preserved, and scrambling must, on this account, involve movement to specifiers of distinct maximal projections. The contrast between (47a) and (47b) thus falls out from this theory, given a theory of locality which distinguishes between multiple specifiers of a single head and specifiers of distinct heads.

⁹ In (47) the possessor *Tanaka-san-no* on the NP out of which relativization takes place guards against the possibility that the relativized element is actually a possessor of the NP, rather than a position inside the contained relative clause.

It is worth noting that the ill-formed example (47b) improves considerably if the numeral quantifier *hutari* 'two' is dropped:

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

(48) ? [seijika -o [t_i t_j kyoonen osieta] Tanaka-san -no kodomoj-ni

politician ACC last-year taught Tanaka HON GEN child DAT

Hanako -ga syookaisita] sensee_j

Hanako NOM introduced teacher

'the teacher_j that Hanako introduced the politician to the children_j of Mr. Tanaka

that t_i taught t_j last year'

This is consistent with Miyagawa's (1997a) approach to local scrambling in Japanese. Recall that floated numeral quantifiers crucially force a base IO-DO order, since this, by hypothesis, is the order in which IO is an NP and floated numeral quantifiers can only modify NPs. In (48), however, there is no floated numeral quantifier; thus, a base DO-IO order is possible in this case. In other words, the scrambled objects can reflect the base order, and relativization out of the second object should be possible, which appears to be the case¹⁰.

In the last section, we saw that a single attractor is apparently necessarily involved whenever multiple scrambling preserves the c-command relation between scrambled elements; that is, no more attractors are posited than necessary. This was taken to account for the lack of quantifier scope ambiguity in cases in which double objects are scrambled in a way which preserves their base order. In this section, we have seen that scrambling

¹⁰ The difference between (47b) and (48) is apparently not simply based on length; replacing the numeral quantifier in (47b) with another word leads to a similar improvement:

Japanese (Shigeru Miyagawa, p.c.)

(i) ? [seijika -o [t_i t_j kyoonen osieta] Tanaka-san -no kodomoj-ni kesa

politician ACC last-year taught Tanaka HON GEN child DAT yesterday

Hanako -ga syookaisita] sensee_j

Hanako NOM introduced teacher

'the teacher_j that Hanako yesterday introduced the politician to the children_j of Mr. Tanaka

that t_i taught t_j last year'

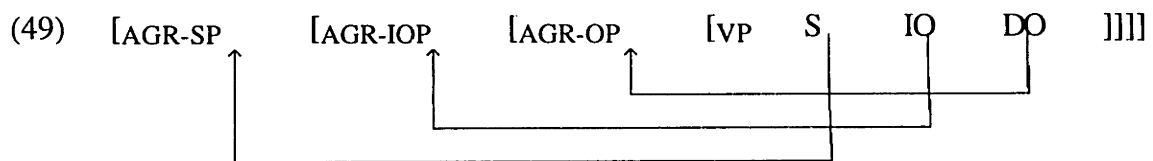
(i) is just as long as (47b), but is like (48) in lacking a numeral quantifier modifying the dative argument, and is just as good as (48). The most straightforward parsing account of these facts therefore does not appear to go through.

which does not preserve the base order must necessarily involve multiple attractors. The most straightforward theory about scrambling, then, can apparently be maintained; scrambling is subject to the general conditions on movement discussed in this paper. The apparent violations of Superiority found in scrambling are a result of the availability of multiple attractors. Speakers apparently may and must posit exactly as many attractors as are necessary to account for the word order of a particular sentence.

3. Object Shift

In the preceding sections we have seen evidence that nested paths are a result of multiple attraction by multiple attractors, while crossing paths are a result of multiple attraction by a single attractor. To the extent that we find this generalization accurate, we are entitled to suspect the work of a single attractor in cases where paths must obligatorily cross.

One case in which crossing paths are standardly invoked is in the movement of arguments from their base positions to their case-checking positions. Such movement has traditionally been thought to involve obligatorily crossing movement to the specifiers of three distinct heads¹¹:



This assumption is taken to account, for instance, for the data in (50)-(51):

¹¹ Here I abstract away from questions about the base-position of the subject; this is assumed by Koizumi 1993 and Collins and Thráinsson 1993, for instance, to be base-generated below AgrIOP but above AgrOP.

Dutch (Neeleman 1994, 419)

- (50) a. ...dat Jan **gisteren** de mannen de foto toonde
 that Jan yesterday the men the picture showed
- b. ...dat Jan de mannen **gisteren** de foto toonde
 that Jan the men yesterday the picture showed
- c. ...dat Jan de mannen de foto **gisteren** toonde
 that Jan the men the picture yesterday showed
 ‘that Jan showed the men the photo yesterday’
- d. *...dat Jan de foto **gisteren** de mannen toonde
- d.*...dat Jan de foto de mannen **gisteren** toonde

Icelandic (Collins and Thráinsson 1993, 143, 149, 154)

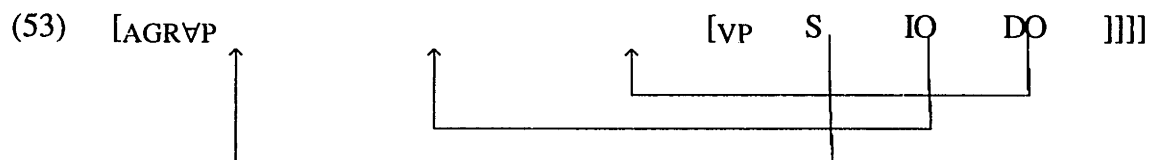
- (51) a. Ég lána **ekki** Maríu bækurnar
 I lend not Maria books
- b. Ég lána Maríu **ekki** bækurnar
 I lend Maria not books
- c. Ég lána Maríu bækurnar **ekki**
 I lend Maria books not
 ‘I do not lend Maria the books’
- d. * Ég lána bækurnar **ekki** Maríu
- e. * Ég lána bækurnar Maríu **ekki**

Examples such as these are typically taken as involving shift of the NPs in question over the relevant adverbs. Crucially, in these cases no reordering of the arguments with respect to other arguments is possible; (52), for instance, is ungrammatical, indicating that the paths of the indirect and direct objects cannot nest:

Icelandic (Collins and Thráinsson 1993, 145)

- (52) * Ég lána bækurnar Maríu **ekki**
 I lend books Maria not

Given the approach developed here, we might take the obligatory crossing of the paths in this case to indicate that only a single attractor is at work. That is, we might alter (49) to (53):

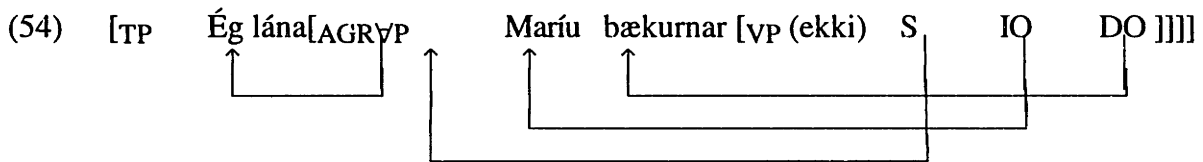


We might assume, then, that argument shift is always triggered by a single AgrP head, which obligatorily triggers overt raising of some non-null set of arguments (that is, at least the subject, and possibly some of the other arguments), the case features of which are checked by Agr. There are a number of other assumptions which are consistent with these data. For instance, I have represented the head responsible for Case-checking as having no other function. Another possibility would be to say that this head also plays some other role. For instance, it might be the head which assigns a theta-role to the subject; we would then need to assume that the subject can have its case checked in situ¹². Nothing crucial will hinge on this in what follows, as far as I can see. Note that as long as Merge is preferred to Move (see Chomsky 1995), the subject will have to be merged before the other arguments are moved to Spec AgrVP; thus, we still predict that the paths of the indirect object and the direct object to Spec AgrVP will obligatorily cross and will be forced by Shortest Move to land under the specifier occupied by the subject, even if the subject is base-generated in Spec AgrVP¹³.

I assume that there is at least one other head which attracts a single XP to its specifier and hosts the finite verb in the overt syntax, which we can think of as the head responsible for the EPP; I will refer to this head, again not crucially, as T. The structure of a sentence like (51c), then, would be as in (54):

¹² See section 4 of this chapter for another possible role for this head.

¹³ See Mulders (1996, to appear) for an analysis of Transitive Expletive Constructions which involves movement to multiple specifiers of which one is filled by Merge.



One potential advantage of this approach over some of its predecessors is that of simplicity. The movements ascribed to arguments in this picture follow completely straightforwardly from the same mechanisms as those used to regulate A-bar movement; no additional stipulations are necessary, and the number of maximal projections assumed diminishes considerably. Of course, this is not an advantage if independent motivation can be mustered for either the stipulations or the maximal projections.

A second potential advantage has to do with the apparent “flattening” of the clause in certain Germanic languages. Consider two groups of languages, one of which I will refer to as “object shift languages”, and the other as “non-object-shift languages”. The groups will crucially differ from those established by the $[\pm\text{Spec TP}]$ parameter of Jonas and Bobaljik (1993), Bobaljik (1995), Bobaljik and Jonas (1996) in that I will include among the object shift languages languages such as Danish, Swedish, and Norwegian, in which only pronominal objects may shift. This group will therefore include all the Germanic languages other than English, while the non-object shift languages will include English and presumably some non-Germanic languages.

The claim made here is that object shift languages may potentially move subjects, objects, and indirect objects into multiple specifiers of a single head in the overt syntax. Suppose we accept the popular claim (Ura 1996, Rudin 1988, Comorovski 1986, Reinhart 1979) that in such a configuration all the multiple specifiers are equally accessible to attraction; none is structurally higher than the others, at least as far as c-commanding attractors are concerned. Then we expect to find evidence that arguments other than the subject are more accessible to attraction in object shift languages than in non-object shift languages.

There does seem to be further evidence that object shift must sometimes be used as an escape hatch for further movement. Haegeman (1993, 1996) gives some evidence along these lines from West Flemish. In West Flemish, as in Dutch, the word order is rigidly S-IO-DO:

West Flemish (Haegeman 1996, 150)

- (55) a. da Valère Marie dienen boek verzekest nie toogt
 that Valere Marie that book probably not shows
 b. da Valère Marie verzekest dienen boek nie toogt
 that Valere Marie probably this book not shows
 c. da Valère verzekest Marie dienen boek nie toogt
 that Valere probably Marie this book not shows

Thus, object shift of the direct object is apparently only possible if the indirect object also shifts. This follows straightforwardly from the approach given here, in which object shift involves multiple attraction by a single attractor; as always, the highest available mover must be attracted first.

Interestingly, cliticization of the direct object is also only possible if the indirect object shifts:

West Flemish (Haegeman 1996, 160):

- (56) a. da Valère Marie **ze** misschien gegeven eet
 that Valere Marie them perhaps given has
 b. da Valère **ze** Marie misschien gegeven eet
 c. da **ze** Valère Marie misschien gegeven eet
- (57) a. *da Valère **ze** misschien Marie gegeven eet
 that Valere them perhaps Marie given has
 b. *da **ze** Valère misschien Marie gegeven eet

That is, for the direct object clitic to undergo further movement to a higher position, it must first enter into a multiple-specifier relation with the indirect object. Similarly, in Dutch, shift of the indirect object is necessary for the direct object to move to the pre-V2 slot¹⁴:

Dutch (Sjef Barbiers, Iris Mulders, p.c.)

- (58) a. De foto heeft Jan de mannen gisteren getoond
 the photo has Jan the men yesterday shown
 b. ??De foto heeft Jan gisteren de mannen getoond

Again, movement of the direct and indirect objects to multiple specifiers of AgrP allows the syntactically lower object to move to a higher position, a familiar situation to devotees of multiple specifiers.

Path Containment Condition effects are another case in which object shift languages might be claimed to show a more “flat” structure than non-object shift languages. PCC effects are claimed to be absent for at least some speakers of Norwegian, Swedish, and Danish (Christensen 1982, Engdahl 1982, 1984, 1985, Erteschik-Shir 1982, Taraldsen 1986):

- (59) *Norwegian* (Øystein Vangsnes, p.c.)

- a. ? Hvilken oppdagelsesreisende; spurte læreren deg
 which explorer asked teacher you
 hvilket kontinent; tj oppdaget tj?
 which continent discovered

Swedish (Engdahl 1982, 169)

- b. Sina föräldrar; är det lätt att glömma hur mycket; man är skyldig tj tj
 self's parents is it easy to forget how much one owes

¹⁴ There is apparently a similar effect with wh-movement, but the contrast is not as strong, for reasons I do not understand:

Dutch (Sjef Barbiers, p.c.)

- (i) Wat heeft Jan de mannen gisteren getoond?
 what has Jan the men yesterday shown
 (ii) ? Wat heeft Jan gisteren de mannen getoond?

Danish (Erteschik-Shir 1982, 186)

c. Marie_j ved jeg ikke hvem_j Peter tror t_j elsker t_j

Marie know I not whom Peter thinks loves

Furthermore, it has been claimed (Kirsti Koch Christensen, p.c. to David Pesetsky, p.c.) that such crossing paths are only available for D-linked phrases; that is, for elements which might in principle undergo object shift. The theory sketched here makes a number of predictions. One is that the sentences in (56) will become bad again if a clause boundary intervenes between the extraction sites; such a boundary would presumably prevent the arguments from becoming specifiers of a single head. This prediction is borne out in Norwegian:

Norwegian (Øystein Vangsnes, p.c.)

(60) a. ? Hvilken oppdagelsesreisende_j spurte læreren deg

which explorer asked teacher you

hvilket kontinent_j t_j oppdaget t_j?

which continent discovered

b. * Hvilken oppdagelsesreisende_j spurte læreren deg

which explorer asked teacher you

hvilket kontinent_j t_j trodde at han kunne oppdage t_j?

which continent thought that he could discover

One potential problem with this account will be discussed in section 1.2 of the next chapter.

I have claimed that the absence of PCC effects for clausemate wh-words in the Scandinavian languages is related to the availability of object shift in these languages; wh-words can move into the multiple specifiers of AgrVP on their way to their eventual landing sites, thus becoming equidistant. However, object shift can typically only target pronominals in Mainland Scandinavian languages:

Swedish (Holmberg and Platzack 1995, 141)

- (61) a. Johan känner henne_i inte t_j
 Johan knows her not
 b.*Läste studenterna artikeln_i inte alla t_j
 read the-students the-article not all
 'Didn't the students all read the article?'

Thus, if this account is on the right track, object shift in Scandinavian will have to be possible not only for pronominals, but for wh-words on their way to higher positions in the clause. I will try to explain what unifies this set of elements in the next chapter, where I will liken the conditions on object shift in Scandinavian to those on participial agreement in French.

Finally, the AgrVP approach has the advantage that it would make it technically and conceptually more feasible to implement a class of highly successful theories of Case which I will refer to here collectively as "Dependent Case theories"; theories of this type have been proposed and developed by, for instance, Massam (1985), Yip et al. (1987), Marantz (1991), Bobaljik (1993), and Harley (1995). Dependent Case theories deny the premise that particular morphological cases are linked to particular AgrPs. Rather, the case that appears on a given NP is determined by which other structural cases have been checked in that clause. In a nominative-accusative language, for instance, nominative case must always be assigned to some nominal, ideally (but not always) the subject; once nominative is assigned, accusative is assigned to the next structurally case-marked nominal, and so forth. Harley's (1995, 161) Mechanical Case Parameter is a typical example of case-assignment algorithms of this type:

(62) *The Mechanical Case Parameter*

- (a) If one case feature is checked structurally in a clause, it is realized as
Nominative. (*mandatory case*)
- (b) If two case features are checked structurally in a clause the second is realized as
Accusative.
- (c) If three case features are checked structurally in a clause, the second is realized
as Dative and the third as Accusative.
- (d) The mandatory case in a multiple-case clause is assigned in the top/bottom
AgrP.

Here the parameter in (d) distinguishes between nominative and absolutive languages; nominative-accusative languages assign nominative to the “first” case-bearing nominal in the clause, while ergative-absolutive languages assign absolutive to the “last” case-bearing nominal.

Dependent Case theories are highly successful, for example, in predicting the distribution of morphological cases in a language like Icelandic. As is well known, Icelandic has certain constructions in which dative case is assigned to the subject, and the object receives nominative case:

Icelandic (Harley 1995, 144)

- (63) Calvini liki verkið
Calvin-DAT likes job-NOM
'Calvin likes the job'

The dative argument in (63) has been convincingly argued to be a subject by, among others, Thráinsson (1979), Zaenen et al (1985) and Harley (1995), and the nominative argument can equally convincingly be shown to be an object. In an account in which nominative is always assigned in AgrSP, the object in (63) must be moved to AgrSP; however, there is no reason, apart from the case facts, to posit such a move. In the Dependent Case theories, nominative is assigned to the object because the subject has

received a non-structural case; nominative must be assigned, and it is assigned to the first available argument. The location in which case assignment takes place is irrelevant, and no unmotivated movements of the object need be posited.

On the other hand, Dependent Case theories are conceptually somewhat suspect in that they involve some notion of “communication” between different parts of the clause. A Dependent Case theory based on the standard clause structure must assume that AgrOP can somehow “know” whether nominative case has already been assigned, so that it can assign nominative case only if it remains unassigned by other AgrP heads. It seems reasonable to want to exclude this kind of communication between heads.

In a theory that posits only a single AgrP which is responsible for all case assignment, a Dependent Case theory can be trivially stated; in a nominative-accusative language, for instance, AgrVP assigns nominative case to the first non-structurally-case-marked argument it attracts, accusative case to the second, and so forth¹⁵.

More generally, this theory of case assignment makes the strong prediction that if only a single argument has its Case feature checked in the overt syntax, it will be the subject; AgrVP must first attract the highest available mover. This prediction seems to be borne out.

4. Spec NegP

Another case in which obligatorily crossing paths have been reported in the literature is the phenomenon of “negative fronting”. Izvorski (1995) reports, for instance, that negative elements in Bulgarian undergo a kind of movement which is subject to the same ordering restrictions as wh-movement:

¹⁵ In control infinitivals, AgrVP will presumably also be responsible for assigning Null Case to the first argument it attracts.

Bulgarian (Izvorski 1995, 66)

- (64) a. Nikoj na nikogo ništo ne beše kazal
 nobody to nobody nothing not AUX said
 'No one had said anything to anyone'
- b. *Ništo na nikogo nikoj ne beše kazal
 nothing to nobody nobody not AUX said
- (65) a. Koj na kogo kakvo beše kazal?
 who to whom what AUX said
 'Who said what to whom?'
- b. *Kakvo na kogo koj beše kazal?
 what to whom who AUX said

Similar facts are noted by Haegeman (1995) for West Flemish. West Flemish negative elements may be fronted, in which case they are interpreted as negative indefinites:

West Flemish (Haegeman 1995)

- (66) da Valère an niemand; nietsj nie t_j t_j gezeid en -eet
 that Valere to noone nothing not said NEG has
 'that Valere did not tell anyone anything'

If negative elements are left in situ, the sentence receives a double negation reading:

West Flemish (Haegeman 1995)

- (67) da Valère an niemand; nie t_j niets gezeid en -eet
 that Valere to noone not nothing said NEG has
 'that Valere did not tell anyone nothing'

Haegeman (1995) interprets these facts as indicating that negative elements in West Flemish may optionally undergo movement to the specifier of a Neg⁰ head, in which case they undergo a semantic operation comparable to Absorption of wh-words.

Haegeman (1995) notes that multiple fronted negative NPs are subject to the same restrictions as NPs which have undergone Object Shift in West Flemish:

West Flemish (Haegeman 1995)

- (68) a. da Valère niemand; nietsj nie t_i t_j getoogd en -eet
 that Valere noone nothing not shown NEG has
 ‘that Valere did not show anyone anything’
- b. *da Valère nietsj niemand; nie t_i t_j getoogd en -eet
 that Valere nothing noone not shown NEG has
- (69) a. da Valère Jan; dienen boekj nie t_i t_j getoogd en -eet
 that Valere Jan this book not shown NEG has
 ‘that Valere did not show Jan that book’
- b. *da Valère dienen boekj Jan; nie t_i t_j getoogd en -eet
 that Valere this book Jan not shown NEG has

Thus, negative fronting, like object shift, should be analyzed in this theory as involving movement to multiple specifiers of a single head.

In fact, however, Haegeman (1995) also points out that the word order restrictions in (68)-(69) hold when only one of the two fronted NPs is a negative element:

West Flemish (Haegeman 1995)

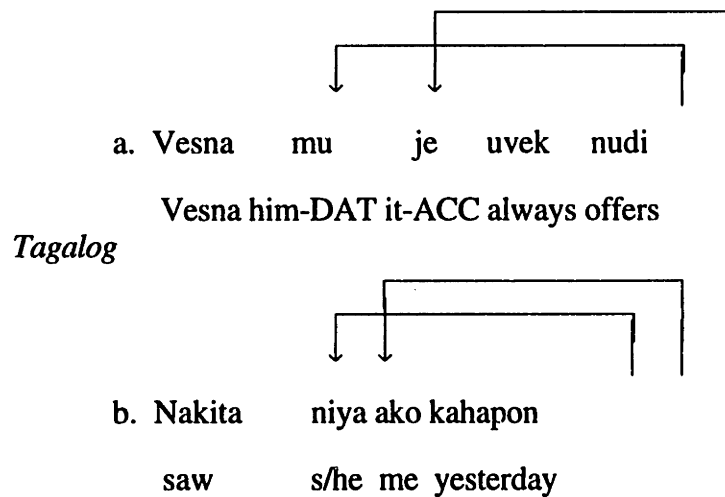
- (70) a. da Valère Jan; nietsj nie t_i t_j getoogd en -eet
 that Valere Jan nothing not shown NEG has
 ‘that Valere did not show Jan anything’
- b. *da Valère nietsj Jan; nie t_i t_j getoogd en -eet
 that Valere nothing Jan not shown NEG has
- (71) a. da Valère niemand; dienen boekj nie t_i t_j getoogd en -eet
 that Valere noone this book not shown NEG has
 ‘that Valere did not show anyone this book’
- b. *da Valère dienen boekj niemand; nie t_i t_j getoogd en -eet
 that Valere this book noone not shown NEG has

Regardless of which NP is negative, then, the shifted indirect object must always precede the shifted direct object. Adapting a suggestion of Cho (1996), we might interpret these facts as indicating that there is a single head which is responsible both for negative fronting and for object shift in West Flemish; that is, that the head referred to in the previous section as AgrV^0 is in fact Laka's (1992) Σ^0 , at least in West Flemish.

5. Cliticization

Another case in which multiple movements appear to obligatorily cross is that of cliticization in many languages:

(72) *Serbo-Croatian* (Franks 1997)



In the theory developed here, this would indicate that clitics of this type are all syntactically moving to check their features against a single head. Some evidence that something like this may in fact give the correct structure for the clitic cluster is given in work in progress by Sandra Stjepanović, who notes the following ellipsis possibilities for the clitic cluster in Serbo-Croatian:

Serbo-Croatian (Franks 1997)

(73) Ona mu ga je dala,...

she him-DAT it-ACC AUX-3sg gave

a. ...a i ja sam mu ga [dala]

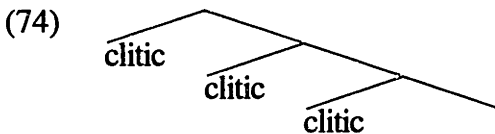
and also I AUX-1sg him-DAT it-ACC gave

b. ...a i ja sam mu [ga-dala]

c. ...a i ja sam [mu-ga-dala]

d.* ...a i ja sam [mu] ga [dala]

On the assumption that ellipsis targets constituents, these ellipsis data argue for a structure of the clitic cluster like that in (74):



This, of course, is the structure predicted by the approach given here. Many interesting questions now arise, which I am not in a position to address here. Are the clitics in (72-73) in multiple specifiers? What happens in cases in which clitics attach to a head with specifiers? In part, this will depend on the content of the notion “specifier”, a question which is not crucial to the theory developed in this chapter (despite its title). In a bare theory of phrase structure, we presumably expect the syntactic position of moved elements to follow from independent principles of syntax. We would not be surprised to find, for instance, that clitics are forced for morphosyntactic reasons to “tuck in” to positions below the specifiers in which fully phrasal categories check features. The only requirement of the theory developed here would be that each movement be to as low a landing site as possible.

6. Freedom of ordering

The discussion thus far has centered on a range of cases in which paths obligatorily cross, leading to a strict word order which is determined by the syntax. There appear to be cases in which the strict word order can be disrupted by other factors, some of them

possibly non-syntactic. For instance, although it is the case, as Rudin (1988) points out, that fronted wh-words in Bulgarian are typically subject to a strict ordering, as shown in (75), this ordering is relaxed somewhat for D-linked wh-words, as we see in (76):

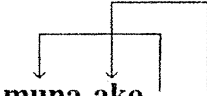

Bulgarian (Roumyana Izvorski, p.c.)

- (75) a. Koj kogo e vidjal?
 who whom AUX seen
 b. *Kogo koj e vidjal?
- (76) a. Koj profesor koja kniga e vidjal?
 which professor which book AUX seen
 ‘Which professor saw which book?’
 b. ?Koja kniga koj profesor e vidjal?

For D-linked wh-words, the order predicted by the theory developed here is still preferred, but the grammaticality of other word order is improved.

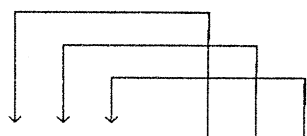
Similarly, in the case of cliticization, the order of clitics predicted by this theory can sometimes be affected by what appear to be morphological factors. In Tagalog, for instance, the general pattern of crossing paths is overridden by a requirement that monosyllabic clitics must precede polysyllabic clitics:

Tagalog

- (77) a. Umuwi **muna ako** 
 went-home first I
 ‘I went home first’
- b. Umuwi **ka muna** 
 went-home you first
 ‘You went home first’

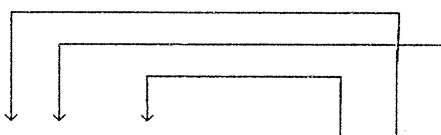
Similarly, in Serbo-Croatian clitic placement can be affected by morphophonological factors which disrupt the basic crossing-paths pattern:

Serbo-Croatian (Steven Franks, p.c.)



(78) Ja **sam mu ga** dala,

I AUX-1sg him it gave,



a i ona **mu ga** **je** dala

but and she him it AUX-3sg gave

'I gave it to him, and she also gave it to him'

The first clause of (78) exhibits the clitic-ordering predicted by this theory, in which all the paths cross. In the second clause of (78), however, this order is disrupted by an independent requirement on the 3rd person auxiliary clitic *je*, which must follow the other clitics (see Spencer 1991, Cavar 1996 for some discussion).

One can imagine a number of approaches to these phenomena. I will be unable to address the problem at any length here, but I will briefly consider a few possible explanations. The questions are interesting ones, and would seem to bear on the nature and extent of interactions between the syntax, on the one hand, and the semantics and morphology on the other.

One class of accounts would posit a fairly close relation between the syntax and the other components of grammar, claiming that the syntactic properties of attraction to multiple specifiers may be influenced by non-syntactic factors. We might, for instance, have a theory in which the requirement that all movements land in the closest specifier to the head can sometimes be overridden by other factors. Or we could claim that heads can in some cases attract a more distant XP rather than the closest one. For example, we could develop

a theory of feature strength in which some strong features are stronger than others and must be checked first, even if they are not the closest XPs with strong features that can be checked. Or we might have a theory that allowed Merge of attractable features during a derivation to XPs embedded in already Merged structure; such a theory would predict that the movement driven by such features would allow free word order, since the features driving movement could be merged at any point in the derivation.

An alternative class of accounts would involve a more distant relation between the syntax and the rest of grammar. Accounts of this type would claim that the alternations in order described above are the result of additional factors operating on a syntactic structure of the type predicted by this theory. This might, for example, involve additional syntactic attractors, or manipulations of the syntactic structure by the morphological component.

These accounts are fairly difficult to distinguish from one another. Of course, there may well be different accounts for different cases, and we should in principle examine each case individually. In what follows I will briefly consider some of the relevant data.

6.1 Morphological effects on syntax: clitic ordering

In the case of the Serbo-Croatian clitics, there appears to be evidence in favor of an account in which a syntactic structure of the type predicted by this theory is altered by the morphological component. Let us consider again the ellipsis data which motivated a particular syntactic structure for the clitic cluster ((73), repeated as (79)):

Serbo-Croatian (Franks 1997)

(79) Ona mu ga je dala,...

she him-DAT it-ACC AUX-3sg gave

a. ...a i ja sam mu ga [dala]

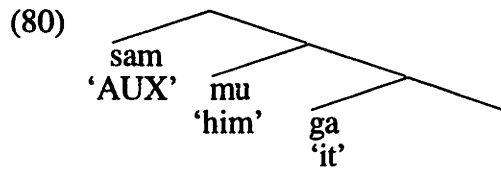
and also I AUX-1sg him-DAT it-ACC gave

b. ...a i ja sam mu [~~ga~~-dala]

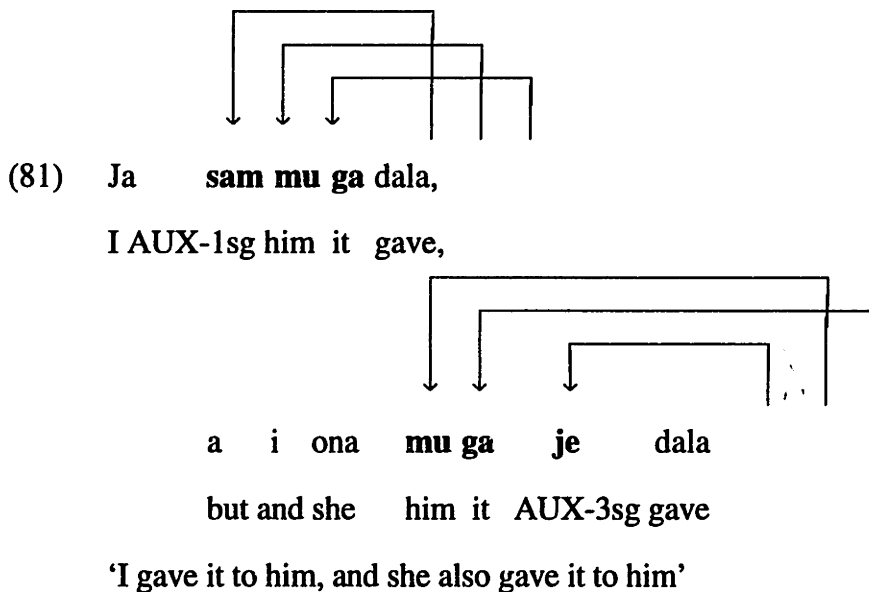
c. ...a i ja sam [~~mu~~-ga-dala]

d.* ...a i ja sam [~~mu~~] ga [dala]

On the assumption that ellipsis can only target constituents, these data suggest a structure for the clitic cluster along the lines given in (80):



Now let us consider again the case in which the order of clitics is influenced by morphological factors. This was the case in (78), repeated as (81):



In the second conjunct of (81), the general pattern of obligatorily crossing paths is disrupted by a morphological requirement that the auxiliary *je* follow the other clitics. Interestingly, the ellipsis facts appear to reflect a structure in which this disruption has not taken place; thus, (82a) is grammatical, and (82b) is impossible¹⁶:

¹⁶ Note that this is not simply the result of a ban on ellipsis of an auxiliary; thus, i. is well-formed:
Serbo-Croatian (Damir Cavar, p.c.)

(i) Pitam se sta li mu je Ivan dao, a sta li [~~mu~~—~~je~~—~~dao~~] Marija
 I-ask SELF what Q him AUX-3sg Ivan gave and what Q him AUX-3sg gave Marija
 'I wonder what Ivan gave him, and what Marija (gave him)'
 Thanks to Damir Cavar for much helpful discussion of these facts.

Serbo-Croatian (Steven Franks, Damir Ćavar, p.c.)

- (82) a. Ja sam mu ga dala, a i ona [~~mu-ga~~] je [~~dala~~]

I AUX-1sg him it gave, but and she him it AUX-3sg gave

- b. *Ja sam mu ga dala, a i ona mu-~~[ga—je—dala]~~

I AUX-1sg him it gave, but and she him it AUX-3sg gave

That is, ellipsis appears to behave as though *je*, like its more well-behaved counterpart *sam*, were at the beginning of the clitic cluster rather than the end. The most straightforward analysis of these facts would seem to involve a syntactic structure of the type predicted by the theory developed here, to which ellipsis is sensitive, and which can then be altered by morphophonological requirements (for instance, by the requirement that *je* be the last clitic), presumably in the morphological component.

6.2 Syntactic effects on syntax: Bulgarian wh-words

We have seen that D-linked wh-words in Bulgarian have more freedom of ordering than non-D-linked wh-words:

Bulgarian (Roumyana Izvorski, p.c.)

- (83) a. Koj kogo e vidjal?

who whom AUX seen

'Who saw whom?'

- b. *Kogo koj e vidjal?

- (84) a. Koj profesor koja kniga e vidjal?

which professor which book AUX seen

'Which professor saw which book?'

- b. ?Koja kniga koj profesor e vidjal?

Again, we have at least two options for dealing with the relative freedom of ordering in (84). We might claim that in this case, the strict locality requirements which force the wh-words in (80) to move to particular specifiers in a particular order are relaxed for D-linked

wh-words. Alternatively, we might claim that an additional attractor or attractors is capable of attracting D-linked wh-words, thereby altering their order.

One fact which bears on this discussion is the mildness of the contrast in (84):

Bulgarian (Roumyana Izvorski, p.c.)

- (85) a. Koj profesor_i koj vāpros_j t_i iska da kaže molitva [predi da obsādim t_j]
 which professor which question wanted to say prayer before we-discuss
 'Which professor wanted to say a prayer before we discuss which issue?'
- b. ?Koj vāpros_j koj profesor_i t_i iska da kaže molitva [predi da obsādim t_j]
 which question which professor wanted to say prayer before we-discuss

These are cases like those discussed in chapter 2 and in section 1 of this chapter, in which one wh-dependency obviates the effects of an island on another dependency. In this case, the wh-phrase *koj vāpros* 'which issue' has been extracted out of an adjunct island. Such extraction would be ill-formed in isolation:

Bulgarian (Roumyana Izvorski, p.c.)

- (86) * Koj vāpros_j iska Ivan da kaže molitva [predi da obsādim t_j]
 which question wanted Ivan to say prayer before we-discuss
 'Which issue did Ivan want to say a prayer before we discuss?'

In (85), however, the ill-formed dependency is apparently redeemed by the presence of a well-formed dependency, namely that involving *koj profesor* 'which professor', which is simply extracted out of the matrix subject position.

We saw in section 1 that this phenomenon of island obviation is sensitive to the order of operations. This can be seen, for instance, by the English contrast in (87):

- (87) a. Who_i t_i persuaded [the man who bought which car_j] to sell the hubcaps?
 b. *Which_j car did John persuade the man who bought t_i to sell which hubcaps?

In (87a), well-formed overt movement of *who* apparently licenses covert movement of *which car* out of a complex NP; this is parallel to the facts in (85). In (87b), it is the overt movement which is ill-formed, and the covert movement which would be well-formed in

isolation. The contrast between (87a) and (87b) suggests that in order for a well-formed dependency to improve the status of an ill-formed dependency, the well-formed dependency must come first in the derivation. Dependencies cannot "redeem" other dependencies retroactively, it seems.

Now let us return to the Bulgarian contrast in (85), repeated as (88):

Bulgarian (Roumyana Izvorski, p.c.)

(88) a. Koj profesor_i koj vāprosj t_i iska da kaže molitva [predi da obsādim t_j]

which professor which question wanted to say prayer before we-discuss

'Which professor wanted to say a prayer before we discuss which issue?'

b. ?Koj vāprosj koj profesor_i t_i iska da kaže molitva [predi da obsādim t_j]

which question which professor wanted to say prayer before we-discuss

(88) displays the relative freedom of ordering which is characteristic of D-linked wh-phrases in Bulgarian. One possibility, in principle, would be to say that this freedom represents an optionality in the order of operations; we might say, for instance, that in the case of D-linked wh-words, the object can be moved before the subject is moved, yielding the order in (88b). If this were the case, however, we would expect a sharp distinction in grammaticality between (88a) and (88b); (88b) would involve an ill-formed dependency which precedes a well-formed dependency in the derivation. In fact, the contrast in grammaticality is not especially sharp; (88b) is no worse than any other sentence in which a D-linked wh-object precedes a D-linked wh-subject. We can therefore apparently rule out optionality in the order of operations, at least as the sole cause of freedom of ordering in this case.

In fact, there is reason to believe that an additional attractor might be responsible for the freedom of ordering in this case. Bulgarian wh-phrases may be preceded in their clause by one or more "topics"¹⁷ (Rudin 1985), which can be in any order:

¹⁷ For some discussion of the semantic properties of these elements, cf. Rudin 1985.

Bulgarian (Roumyana Izvorski, p.c.)

- (89) a. [TopP Ivan včera [CP kakvo kupi]]?
 Ivan yesterday what bought
 'What did Ivan buy yesterday?'
 b. [TopP Včera Ivan [CP kakvo kupi]]?
 yesterday Ivan what bought
 'What did Ivan buy yesterday?'

We might analyze these topics, following Rudin 1985, as being in specifiers of one or more XPs dominating the CP into which the wh-words move.

Interestingly, it appears that only D-linked wh-words may move into the Topic field. A topicalized adverb may marginally intervene between D-linked wh-words, but not between non-D-linked wh-words:

Bulgarian (Roumyana Izvorski, p.c.)

- (90) a. *[TopP Koj včera [CP kakvo kupi]]?
 who yesterday what bought
 'Who bought what yesterday?'
 b. ?[TopP Koja žena včera [CP koja kniga kupi]]?
 which woman yesterday which book bought
 'Which woman bought which book yesterday?'

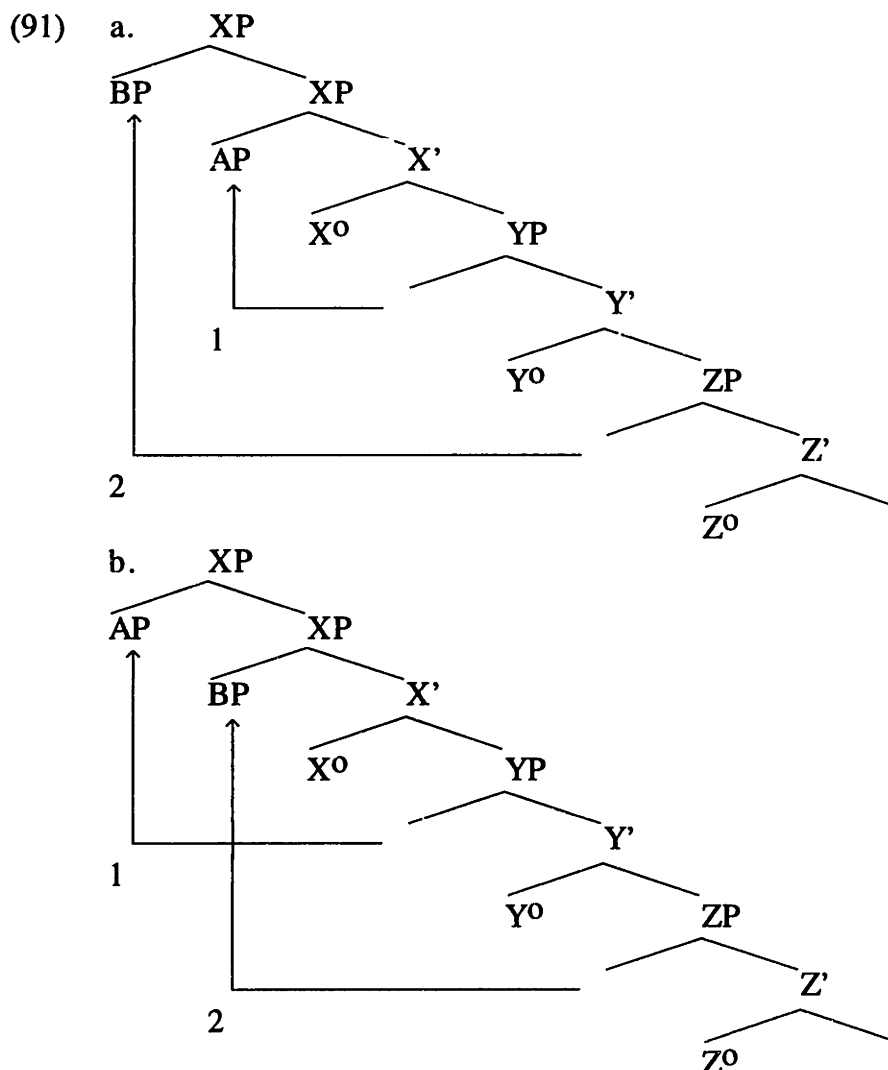
Neither sentence in (90) is especially good, but there is a clear contrast between them. This suggests that D-linked wh-phrases may move into a Topic position, which would account for their relative freedom of ordering¹⁸.

7. Shortest Move

I began this chapter by pointing out that the strict ordering of multiple specifiers might be due to Shortest Move. The basic idea was that in cases of movement to multiple

¹⁸ This still leaves open the question of what accounts for the typical absence cross-linguistically of Superiority effects for D-linked wh-words (see Pesetsky 1987, in particular, for discussion). It is not at all clear that the account developed here of the Bulgarian facts will generalize to languages which do not perform all wh-movement overtly. I will have to leave this issue for further research.

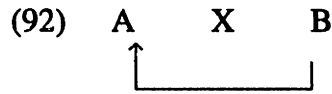
specifiers of a single head, Featural Cyclicity makes no distinction between crossing and nesting paths; both involve multiple checking of a strong feature. On a certain version of Shortest Move, we expect Shortest Move to prefer movement to a lower specifier to movement to a higher specifier. That is, we should be able to employ a version of Shortest Move to choose (91b) over (91a), which is the correct result:



To get this result, we will need a version of Shortest Move which differs in two crucial respects from a version which is currently commonly assumed.

The first is that Shortest Move will have to be sensitive both to potential moveable elements and to potential landing sites; that is, that (92) is a violation of Shortest Move whether X is a movable element that could have moved to A or a possible landing site for

the kind of movement involved in movement from B to A (cf. Murasugi 1992 for a theory of Shortest Move of this kind, and Ochi 1997 for a similar proposal).



Let us define a version of Shortest Move that has the desirable properties. Consider the operation *Attract*, illustrated in (93):

- (93)
- | | | | | | | | |
|----|-------------|----|------------|---|------------|---|---|
| a. | | K | [α | F |] | | |
| b. | [α' | F' |] | K | [α | F |] |

The attractor K attracts the feature F, causing the creation of a copy of F and the minimal element α containing F that allows convergence; at a minimum, α is the formal features of F, but may also be forced by well-formedness conditions imposed by the PF and LF interfaces to be some constituent containing the formal features of F. In the case of overt wh-movement, for instance, α might be an entire NP, although the feature F is simply the wh-feature; the NP is “pied-piped” for reasons having to do with the requirements on well-formed PF objects. The copy α' of α then Merges with K, entering K’s checking domain.

Attract, then, may be stated as in (94):

- (94) *Attract*

An attractor K attracts a feature F, creating a copy α' of an element α containing F, and Merging α' with K. The relations between α' , K, and F must all obey

Shortest.

Shortest is defined in (95):

- (95) *Shortest*

A pair P of elements $\{\alpha, \beta\}$ obeys Shortest iff there is no well-formed pair P' which can be created by substituting γ for either α or β , and the set of nodes c-commanded by one element of P' and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominating the other.

Shortest will crucially constrain the relations between K and F, and between α' and F. The effect of Shortest on the relation between K and F will be to force the attractor to attract the nearest available mover. The effect on the relation between α' and F will be to force movement to be to the closest available landing site; in the case of movement to multiple specifiers, for instance, movement will have to be to the closest available specifier, as desired. Furthermore, Shortest's effect on the relation between α' and F will prevent movement of F past an attractor which could attract F, since the intervening attractor could have an element γ in its specifier which would enter into a well-formed dependency with F. As Attract is defined in (94) it also requires the relation between α' and K to obey Shortest. This is purely for the sake of simplicity; this requirement will play no role in the theory. Requiring this relation to obey Shortest will have no effect other than to force every movement to multiple specifiers to be to the closest available specifier, a result already derived by the effects of Shortest on the relation between α' and F.

In principle, we might split the constraint Shortest into two parts, referring to the effect of Shortest on the relation between K and F as *Shortest Attract*, and to the effect of Shortest on the relation between α' and F as *Shortest Move*. As the theory has been developed thus far, these are simply labels for the effects of Shortest on different types of elements; the constraints Shortest Move and Shortest Attract are not formally distinct. In section 2.6.2.1 of chapter 5 I will discuss the properties of Shortest further. One issue which will be addressed is the question of whether the distinction between Shortest Move and Shortest Attract should in fact be collapsed in this way.

Another conclusion to which we are driven, on this theory, is that the specifiers in (91) are not "equidistant", at least not from the perspective of elements which are moving into them; the lower specifier will have to be closer to a moving element than a higher specifier would be. I will consider the validity of this second assumption in the next section.

8. **Equidistance**

A standard claim in the literature on multiple specifiers has been that multiple specifiers can be used as an escape hatch for movement to higher positions; all the multiple specifiers are typically assumed to be equally accessible to attraction by higher heads. This assumption underlies the claims of Reinhart (1979), Comorovski (1986), and Rudin (1988) that languages with multiple overt wh-movement allow wh-island violations, and the parallel claim by Ura (1996) that languages with multiple subject constructions allow Super-raising.

This assumption of equidistant multiple specifiers is somewhat at odds with the theory developed here, according to which Shortest Move is responsible for the "tucking in" nature of movement to multiple specifiers. For this account to be tenable, a lower specifier will have to be a closer landing site than a higher specifier. Of course, there are a number of technical ways of distinguishing between the Equidistance cases discussed in the literature and the facts discussed here. For instance, the Equidistance cases crucially involve attraction by a higher head, while the discussion here has largely centered on multiple specifiers as landing sites. In principle, we might invoke some asymmetry between the head and tail of a movement chain.

However, we have seen empirical evidence suggesting that for wh-movement, at least, multiple specifiers are not in fact equidistant even from higher attractors. Recall from section 3.6 of chapter 2 that CP-absorption languages--by hypothesis, the languages which allow multiple specifiers of CP--show a preference for extracting the higher of two wh-phrases in Spec CP:

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova,

Kamen Stefanov, p.c.)

- (96) a. ? Koj se opitvat da razberat kogo t e ubil t ?
 who SELF try to find-out whom AUX killed
- b. * Kogo se opitvat da razberat koj t e ubil t ?
 whom SELF try to find-out who AUX killed
-

Chinese (Lisa Cheng, Hooi Ling Soh, Wei-tien Dylan Tsai, p.c.)

- (97) jingcha xiang-zhidao [shei sha -le shei]
 police want know who kill PERF who
- a. 'Who_i are the police trying to find out who_j t_i killed t_j?'
 b. *'Who_j are the police trying to find out who_i t_j killed t_i?'
-

The contrast in (96-97) follows straightforwardly from the approach developed here.

Consider the derivation of the well-formed (a) sentences. In (98a), the embedded C⁰ is Merged and attracts the wh-words to its multiple specifiers; the paths cross, for reasons already discussed. Eventually, in (98b), the matrix C⁰ is merged, and must attract one of the wh-words in the lower CP. Apparently it must attract the higher of the two specifiers:

- (98) a. [CP wh wh [t t]]
 b. [CP [CP wh wh [t t]]]
-

Thus, we have evidence that multiple specifiers are not in fact equidistant to attraction by higher heads, in this case, at least. The cases in which it is possible to move a wh-word other than the highest one in Bulgarian, as we saw in section 3.6 of the last chapter, are those involving D-linked wh-words, in which there is some freedom of ordering of the wh-words:

Bulgarian (Roumyana Izvorski, p.c.)

- (99) a. *Koja kniga te popita učitelja kogo ubedi Ivan t da publikuva t?*
 which book you asked teacher who convinced Ivan to publish
 ‘Which book_i did the teacher ask you who_j Ivan convinced t_j to publish t_i?’

- b. *Koj izdatel te popita učitelja kakvo ubedi Ivan t da publikuva t?*
 which publisher you asked teacher what convinced Ivan to publish
 ‘Which publisher_j did the teacher ask you what_i Ivan convinced t_j to publish t_i?’

- (100) a. *Koj kontinent te popita učitelja koj t e otkril t?*
 which continent you asked teacher who AUX discovered

- b. *Koj otkrivatel te popita učitelja kakvo t e otkril t?*
 which explorer you asked teacher what AUX discovered

- (101) a. *Koj after koja kniga t e napisal t?*
 which author which book AUX wrote

‘Which author wrote which book?’

- b. *?Koja kniga koj after t e napisal t?*

which book which author AUX wrote

As we saw in section 5.2 of this chapter, there is some reason to believe that this freedom of ordering for D-linked wh-words is caused by an additional attractor (referred to above as Topic⁰).

Multiple A-bar specifiers, then, would appear not to be Equidistant to attraction by higher heads. On the other hand, we have no evidence for this conclusion for cases of A-movement. In fact, I crucially assumed in section 3 of this chapter that multiple A-specifiers are equidistant for attraction by higher heads, and Ura's (1996) account assumes the same. If these accounts are correct, then, there is apparently a fundamental difference between multiple A-bar specifiers and multiple A-specifiers. From the point of view of higher, attracting heads, A-specifiers are apparently equidistant, unlike A-bar specifiers. I will have to leave this issue for further research; the number of clear cases of movement to multiple specifiers is too small to draw any definite conclusions. For the time being, I have only a few speculations to offer.

One possibility, suggested to me by Molly Diesing (p.c.), would be to exploit the distinction drawn in Chomsky 1995 between interpretable and uninterpretable features. The cases of A-movement discussed here arguably involve checking of uninterpretable features, which vanish under checking, while the cases of A-bar movement appear to involve interpretable features. If the features which drive movement are somehow responsible not only for attraction but also for distinguishing between multiple specifiers, then it might be that the erasure of uninterpretable features under checking somehow erases the hierarchical relations between the specifiers as well, making them "equidistant" after checking.

Another possible move would take advantage of the idea, suggested above in section 4, that the head responsible for object shift is in fact Laka's Σ^0 , and that shifted nominals undergo a process similar to Absorption of wh-phrases. If this process has syntactic effects, we might expect it to do away with the hierarchical relations between the attracted specifiers. Note that no such process is being undergone by the wh-words in

multiple specifiers in the cases under discussion; such wh-words are crucially not undergoing Absorption with the other wh-words in the relevant cases, since they are undergoing further movement to a higher CP and thus do not have the same scope as the other specifiers.

9. Conclusion

In this chapter I have tried to show that Chomsky's (1995) notion of featural cyclicity is to be preferred over earlier versions of cyclicity. Featural cyclicity predicts--correctly, I have argued--that multiple movements will cross rather than nesting just in case their destinations are multiple specifiers of a single head. I have claimed that this is true for both A- and A-bar movement, which allows us to derive these facts from a straightforward theory based on featural cyclicity and Shortest Move.

Chapter Four: In Full Pursuit of the Unspeakable

In Chapter 2 we encountered a paradox having to do with the behavior of multiple wh-movement in Serbo-Croatian. Recall that Serbo-Croatian, unlike Bulgarian, forbids wh-movement out of wh-islands. Following Reinhart (1979), Comorovski (1986), and Rudin (1988), I took this to be diagnostic of the lack of availability of multiple specifiers of CP in Serbo-Croatian:

(1) *Serbo-Croatian* (from Rudin 1988, 459)

a. * Šta si me pitao ko može da uradi?

what AUX-2s me asked who can to do

‘What have you asked me who can do?’

Bulgarian (from Rudin 1988, 457)

b. ? Koja ot tezi knigi se čudiš koj znae koj prodava?

which of these books wonder-2s who knows who sells

‘Which of these books do you wonder who knows who sells?’

On the other hand, Bošković (1995c) points out that multiple wh-movement out of a non-wh clause is possible in Serbo-Croatian:

Serbo-Croatian (Bošković 1995c, 8)

- (2) Ko si koga tvrdio da t je istukao t
 who AUX whom claimed that AUX beaten
 ‘Who did you claim beat whom?’

On the assumption that long-distance wh-movement must be successive-cyclic, this indicates that Serbo-Croatian may have multiple specifiers of CP, just in case none of the moving wh-words remain in any of the specifiers. In this chapter I will develop a theory of positions of this type, which can be occupied only during a derivation but cannot be final landing sites for movement.

The theory will be based on the principle in (3):

- (3) PF must receive unambiguous instructions about which part of a chain to pronounce¹.

(3) must be understood in conjunction with the principle in (4):

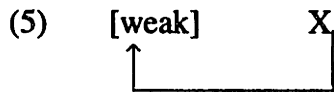
- (4) A strong feature instructs PF to pronounce the copy in a chain with which it is in a feature-checking relation.

(3) and (4) impose certain restrictions on the possible operations which can be performed in the “overt syntax”. I use that term here, as elsewhere, to refer to the part of the derivation which precedes Spell-Out and which is therefore subject to a requirement that it produce well-formed PF objects. I assume a version of the copy theory of movement, according to which movement involves the creation of a copy of the moved element in the position to which it moves. By “chain” I mean the total set of copies of a single object in the representation given to PF. This notion of chain may differ importantly from the notion of chain needed in the syntactic and semantic components; for instance, A-movement followed by A-bar movement, which is typically assumed to create two syntactic chains, only creates one chain for purposes of the theory developed here. The claim being made is that all of

¹ Here I assume, for the sake of simplicity, that only a single element in a chain will be pronounced. This assumption has been denied in a number of works (e.g., Koopman 1984, Pesetsky to appear). It seems to me that approaches of this kind are also compatible with the theory developed here, although they will clearly complicate it somewhat.

the copies of a single object are treated by the phonology as a unit of a certain type, to which the constraints in (3) and (4) apply.

One example of a chain ruled out by (3) and (4) is an overt movement of the type in (5):

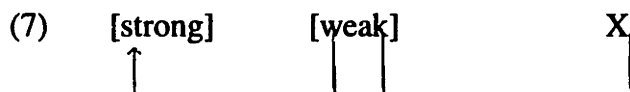


(5) contains two positions which could in principle be pronounced, and there is no reason to choose one of them over the other. (3) is therefore violated in this case. Two examples of chains which would be allowed by (3) and (4) are given in (6):



The chain in (6a) contains two elements, but one of them is associated with a strong feature which instructs PF to pronounce the head of the chain, and (3) is therefore satisfied. The trivial chain in (6b) has no positions associated with strong features, but there is only a single candidate for pronunciation, so there is no question of PF having to choose between positions. (3) and (4) thus provide us with a representational version of Chomsky's (1993) Procrastinate, ruling out overt movement to check weak features, but not overt movement to check strong features. However, the theory consisting of (3) and (4) is both weaker and stronger than Procrastinate. I will try to show that the differences between Procrastinate and the theory sketched above are in fact desirable ones.

(3) leaves open at least three classes of cases in which a weak feature drives overt movement but movement is well-formed. One would be a case in which an element moves to check a weak feature and then moves further to check a strong feature.



On this theory, then, we expect to find a class of movements which are only licensed if the moved element then undergoes a further move, but not if it remains in situ. A number of

movements of this type have been discussed in the literature, one of them being French participial agreement (for discussion of these facts, see Kayne 1989, Branigan 1992, Ura 1993, among many others):

French

- (8) a. *Josèphe a écrite cette lettre
Joseph has written-FEM this-FEM letter
- b. Quelle lettre Josèphe a-t-il écrite?
what-FEM letter Joseph has-he written-FEM
- c. la lettre que Josèphe a écrite hier
the-FEM letter that Joseph has written-FEM yesterday
- d. Cette lettre, Josèphe l'a écrite hier
this letter Joseph her-has written-FEM yesterday
- e. Les lettres ont tous été écrites
the-PL letters have all been written-PL

On the assumption that triggering participial agreement involves movement to the specifier of some kind of AgrP projection, these seem to be the kind of cases we expect to see; movement to the agreement-triggering projection is impossible unless the moved element undergoes further movement.

Another case of a well-formed chain which is not associated with any strong features would be one which contains only a single copy which could be pronounced. One instance of this, of course, depending on our theories about the distribution of strong features, would be a trivial, single-membered chain. Elements in their base positions are not standardly assumed to be associated with strong features, yet PF has no difficulty pronouncing them; presumably, (3) only rules out cases in which PF is unable to determine which of several copies to pronounce.

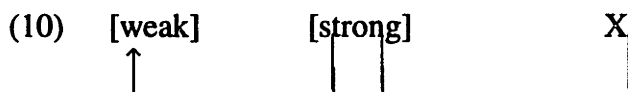
Another, more surprising case in which PF only has to consider a single position for pronunciation is one in which a weak feature drives overt movement out of an ellipsis

site. In this case, PF presumably receives instructions not to pronounce anything within the ellipsis site. A chain of this kind, then, like a trivial chain, has only a single member which is a candidate for pronunciation. The fact that this member is not associated with a strong feature should be irrelevant.

I will argue that this second case is the one responsible for such phenomena as gapping and (multiple) sluicing (for discussion of examples like (9b), cf. Bolinger 1978, Merchant 1996, Nishigauchi to appear; compare the ill-formed (9c)):

- (9) a. Some prefer syntax, and others phonology
 b. ??I know somebody bought something, but I can't remember who what
 c. *I know somebody bought something, and I'm pretty sure (that) John a car

A third possible type of well-formed chain, according to the theory developed here, would be one in which movement to check a strong feature is followed by movement to check a weak feature:



If the theory developed here is right, such chains might exist but would be rather difficult to detect; the phonology would receive instructions to pronounce this chain at the intermediate position, the one associated with a strong feature. It will therefore be difficult to distinguish a chain of the type in (10), involving overt movement to check a weak feature, from a chain in which overt movement stops at the position associated with the strong feature, and movement to the position associated with the weak feature is covert. I will discuss some possible examples of chains of this type in section 3 of this chapter. The answer to the question of whether such chains are allowed will depend partly on our theory of the nature of the derivation. As was noted in chapter 1, there is a class of theories, including those of Bobaljik (1995), Brody (1995b), Groat and O'Neil (1996), and Pesetsky (to appear), in which overt and covert movements are "interleaved", rather than being segregated as they are in the model assumed here into pre-Spell-out and post-Spell-

out components. The only difference between an overt and a covert movement, on these theories, is the decision as to which copy in the movement chain to pronounce. Such a theory would therefore allow chains of the type in (10), as cases of covert movement to the position associated with the weak feature. Note that the theory developed here allows “bottom-heavy” chains of this type only in cases in which movement to a weak feature is preceded by movement to a strong feature. A chain created by a single overt movement to a weak feature would be ill-formed (unlike in the Single Output Syntax models alluded to above, in which movement to a weak feature simply produces a chain whose tail is pronounced).

In addition, there is one class of cases ruled out by (3) and (4) which Procrastinate allows. This would be a case in which successive-cyclic movement takes place to check two strong features:



In the theory developed here, each strong feature instructs PF to pronounce the copy of the chain associated with it, and PF again fails to receive unambiguous instructions about which copy in the chain to pronounce. (11) is thus no better, in the account developed here, than most cases of overt movement to check a weak feature.

One case which is standardly taken to be of the type in (11) is overt *wh*-movement of a subject in a language like English:

(12) Who do you think *t* ate all the zucchini?

A standard theory about examples like (12) is that they involve overt raising of the subject from an internal subject position to an external position to check Case or an EPP feature, followed by overt *wh*-movement to Spec CP. If the theory developed here is correct, this cannot be the case. There is much cross-linguistic evidence that subject extraction is special in various ways; I will try to show that the properties of subject extraction follow from the theory developed here.

Finally, we expect, on this theory, that phonologically null elements (e.g., PRO and *pro*) should have greater freedom of movement than phonologically overt elements. Here there is presumably no question of PF deciding which part of the chain to pronounce, since the entire chain is phonologically null. We should therefore find cases, for instance, in which phonologically null elements can undergo overt movement to satisfy a weak feature. We will see some evidence that this is the case.

1 Overt movement to “weak” positions

The literature mentions a number of cases in which movement to a position is only possible when the moving element is phonologically null or undergoes further movement out of the position in question. In this section I will discuss some cases of this and sketch how the theory developed here might account for them.

1.1 French Participles

Consider again the French participial agreement facts in (8), repeated as (13):

French

- (13) a. *Josèphe a écrite cette lettre
 Joseph has written-FEM this-FEM letter
- b. Quelle lettre Josèphe a-t-il écrite?
 what-FEM letter Joseph has-he written-FEM
- c. la lettre que Josèphe a écrite hier
 the-FEM letter that Joseph has written-FEM yesterday
- d. Cette lettre, Josèphe l’a écrite hier
 this letter Joseph her-has written-FEM yesterday
- e. Les lettres ont tous été écrites
 the-PL letters have all been written-PL

French participial agreement is only possible if the object has undergone further movement; it is impossible if the object remains in situ, as in (13a), but can be licensed by A'-movement, as in (13b-c), by cliticization, as in (13d), or by A-movement, as in (13e).

Participial agreement in French thus appears to be a phenomenon of the type predicted by this theory. In our terms, the feature responsible for participle agreement is weak and thus cannot trigger overt raising itself, as in (13a), as the resulting chain would be an illegitimate PF object. On the other hand, a movement triggered by another strong feature can use the specifier of the projection headed by the participle agreement feature as an intermediate landing site; the resulting chain has a strong feature at its head, and is therefore well-formed. The morphology of participle agreement, then, appears in all and only those cases in which the specifier of the relevant head is occupied at some point in the overt syntax. We will see other cases in which the presence of overt morphology has a different significance, indicating the presence of a strong rather than a weak feature.

There are some subtleties in the facts about participle agreement which do not fall out of this theory without extra assumptions. For instance, participial agreement in Italian, like that in French, must be licensed by movement, but only A-movement and cliticization can license it. Compare (13) with (14):

Italian

- (14) a. *Ho creduti gli Australiani
 have-1sg believed-PL the Australians
 ‘I have believed the Australians’
- b. Li ho creduti
 them have-1sg believed-PL
 ‘I have believed them’
- c. Gli Australiani sono stati creduti
 the Australians are been-PL believed-PL
 ‘The Australians have been believed’
- d. *Quanti Australiani hai creduti?
 how-many Australians have-2sg believed-PL
 ‘How many Australians have you believed?’

A perhaps related fact is that participial agreement in French is optional with *wh*-movement, but obligatory with passivization:

French (adapted from Branigan 1992, 33-34)

- (15) a. Les livres de Jules Verne ont tous été imprimés /*imprimé
 the books of Jules Verne have all been printed-PL printed
 ‘Jules Verne’s books have all been printed’
- b. Quelle lettre Josèphe a-t-il écrite / écrit?
 what-FEM letter Joseph has-he written-FEM written
 ‘Which letter did Joseph write?’

I will discuss these matters further in section 1.9.

1.2 Object Shift in Mainland Scandinavian

Icelandic and the Mainland Scandinavian languages differ in that object shift may only apply to pronominal objects in the latter, while in Icelandic any definite NP may shift:

Icelandic (Holmberg and Platzack 1995, 141)

- (16) a. Jón þekkir hana; ekki t_j
 Jon knows her not
 ‘Jon knows her’
- b. Lásu stúdentarnir greinina; ekki allir t_j?
 read the-students the-article not all
 ‘Didn’t the students all read the article?’

Swedish (Holmberg and Platzack 1995, 141)

- (17) a. Johan känner henne; inte t_j
 Johan knows her not
 ‘Johan knows her’
- b. *Läste studenterna artikeln; inte alla t_j
 read the-students the-article not all
 ‘Didn’t the students all read the article?’

Holmberg (1986) and Holmberg and Platzack (1995) point out that the pronouns which undergo object shift in Mainland Scandinavian languages are clitic-like in many respects; they must be unstressed and morphologically simple.

In section 3 of chapter 3, I offered an account of the absence of Path Containment Condition effects in the Mainland Scandinavian languages which made crucial reference to the availability of overt object shift in those languages (Christensen 1982, Engdahl 1982, 1984, 1985, Erteschik-Shir 1982, Taraldsen 1986):

(18) *Swedish* (Engdahl 1982, 169)

- a. Sina föräldrar_i är det lätt att glömma hur mycket_j man är skyldig t_j t_j
 self's parents is it easy to forget how much one owes

Danish (Erteschik-Shir 1982, 186)

- b. Marie_i ved jeg ikke hvem_j Peter tror t_j elsker t_j
 Marie know I not whom Peter thinks loves

The idea was that object shift involves movement of shifted arguments into multiple specifiers of a single head (which I called, for ease of reference, Agr \bar{V} P). In an example like (18b), for instance, the subject and the object of the most deeply embedded clause are multiple specifiers of Agr \bar{V} P at a certain point in the derivation. For reasons which are still unclear, multiple A-specifiers (though not multiple A-bar specifiers; cf. chapter 3, section 8) are apparently equidistant to higher attraction, and the arguments are therefore both equally accessible to higher attractors; the lack of Path Containment Condition effects follows.

If this account is right, then object shift must be available in Mainland Scandinavian not only for the clitic-like object pronouns but also for elements which undergo further wh-movement out of their shifted positions, as in (18). The set of things which can undergo object shift, then, is similar to the set of things which can trigger participial agreement in French, and the account should be the same; the feature which is responsible for object

shift in Mainland Scandinavian is weak, and object shift is therefore only possible for elements which undergo further movement to a higher position².

1.3 V-to-I in Mainland Scandinavian

The Mainland Scandinavian languages, like many of their Germanic cousins, exhibit a V2 effect in matrix clauses, standardly analyzed as involving movement of V through one or more Infl projections to C:

Danish (Vikner 1995, 142)

- (19) a. Helge vil gerne læse den her bog
 Helge will readily read this here book
 b. Den her bog vil Helge gerne læse
 this here book will Helge readily read
 c. *Den her bog Helge vil gerne læse
 this here book Helge will readily read

In non-V2 clauses, however, Vikner (1995) argues convincingly that the verb remains in VP, without raising to I:

Danish (Vikner 1995, 145)

- (20) a. *Jeg spurgte hvorfor Peter havde ikke læst den
 I asked why Peter had not read it
 b. Jeg spurgte hvorfor Peter ikke havde læst den
 I asked why Peter not had read it

The Mainland Scandinavian languages seem to differ in this regard from, for instance, Icelandic, which has an independent process of raising V to I even in non-V2 clauses.

Thus, the verb must always raise over negation:

²Note that we do not predict that French participial agreement and Icelandic object shift will be syntactically identical in other respects; the only similarity which is necessary, on this theory, is that they are both driven by a weak feature. The phenomena clearly differ, for instance, with respect to Holmberg's Generalization.

Icelandic (Vikner 1995, 145)

- (21) a. Ég spurði af hverju Pétur hafði ekki lesið hana
 I asked why Peter had not read it
 b.* Ég spurði af hverju Pétur ekki hafði lesið hana
 I asked why Peter not had read it

Movement of the verb into I, then, is typically impossible in Mainland Scandinavian. On the other hand, in cases in which the verb moves into C, our standard assumptions about locality and head movement force us to the conclusion that the verb moves into I as an intermediate landing site. Thus, V-to-I movement is possible just in case the verb continues to move, but not if it remains in I. In our terms, the feature forcing V-to-I movement in Mainland Scandinavian is weak, and simple movement of the verb into I therefore creates an illegitimate PF object; further movement to a head with a strong feature must take place.

1.4 Japanese *wh*-movement

Japanese *wh*-movement, as we have seen, appears to be driven by a weak feature:

Japanese

- (22) Taroo-wa nani -o katta no?
 Taroo TOP what ACC bought Q
 ‘What did Taroo buy?’

One prediction made by this theory is that null operators in Japanese should be able to move in the overt syntax to check weak features on C^0 , since they are immune to the PF-imposed ban on overt movement to weak features. There is some evidence that this is the case. Japanese relativization arguably involves movement of a null operator:

Japanese

- (23) [OP John-ga t yonda] hon
 John NOM read book
 ‘the book John read’

As a diagnostic for the position of this null operator in the overt syntax, we can use the distribution of nominals marked with the emphatic postposition *koso*. *Koso*-NPs cannot be c-commanded by a wh-word in the overt syntax:

Japanese (Tanaka (in preparation))

- (24) a. John-koso nani -o yonda no?
 John EMPH what ACC read Q
 ‘What did *John* read?’
 b. *Nani-o_i John-koso t_j yonda no?
 what ACC John EMPH read Q
 (25) a. *Dare-ga LGB-koso yonda no?
 who NOM LGB EMPH read Q
 ‘Who read *LGB*?’
 b. LGB-koso_i dare-ga t_j yonda no?
 LGB EMPH who NOM read Q

Thus, we may use *koso*-NPs to determine whether the null relative operator has moved or not. *Koso*-NPs may appear in complex NPs:

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (26) a. [John-koso LGB -o yonda to iu] uwasa

John EMPH LGB ACC read that rumor

‘the rumor that *John* read LGB’

- b. [John-ga LGB -koso yonda to iu] uwasa

John NOM LGB EMPH read that rumor

‘the rumor that John read LGB’

However, a *koso*-NP cannot occur in a relative clause:

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (27) a. *[John-koso yonda] hon

John EMPH read book

‘the book that *John* read’

- b. *[LGB -koso yonda] hito

LGB EMPH read person

‘the person that read LGB’

The ill-formedness of (27) receives a natural account if we assume that the relative operator in Japanese always undergoes overt movement to Spec CP--that is, to a position c-commanding any *koso*-NP in the relative clause. (27) would then be ill-formed for the same reason that (24b) and (25a) are.

Moreover, if this line of reasoning is correct, overt movement of the null relative operator in Japanese is not only possible but obligatory. If it were possible for the operator to remain in situ, (27a) at least should be well-formed.

1.5 Tagalog vs. Chamorro

The Austronesian language Tagalog exhibits a phenomenon which I will refer to here as topicalization (the nominals I will refer to as “topics”, following Schachter 1976, are also known in the literature as “subjects”, “foci”, “Nominative NPs” (Kroeger 1993,

Schachter 1996) or “triggers”, among other names). A Tagalog sentence often has a number of rough paraphrases like the ones in (28)³:

Tagalog

(28) a. Bumili si Maria ng kalabaw sa tindahan

AT-bought T Maria G water-buffalo Loc store

‘Maria bought a water buffalo at the store’

b. Binili ni Maria ang kalabaw sa tindahan

GT-bought A Maria T water-buffalo Loc store

‘Maria bought the water buffalo at the store’

In most Tagalog sentences, one argument must be chosen as the topic (the topics above are underlined). Topics are marked by nominal morphology, and morphology on the verb roughly indicates the theta-role of the topic. The exact form of the verbal morphology varies with the conjugation class of the verb; in the sentences above, the relevant morphemes are the infixes *-um-* and *-in-*. The sentences in (28) only represent a small fraction of the possible topics in Tagalog; other topicalizable elements include locations, instruments, causees, and reasons.

Topicalization in Tagalog is standardly viewed as involving movement to a fairly high position in the clause, the properties of which are much-disputed (cf. Schachter 1976, 1996, Richards 1990, 1993, to appear a, Guilfoyle, Hung, and Travis 1992, Kroeger 1993, MacLachlan and Nakamura 1994, MacLachlan 1995, Nakamura 1996, Voskuil 1996

³In this section I generally use the abbreviations of Schachter (1976) in Tagalog glosses, and those of Chung (1982) in glossing Chamorro. I will, however, use the same glosses for the Tagalog and Chamorro verbal morphology which I refer to here as topicalization morphology, thus departing from Chung’s (1982) glosses in this regard.

The abbreviations used for Tagalog are:

T=Topic

A=Actor (roughly, logical subject)

G=Goal (roughly, logical object)

Loc=Locative

AT=Actor-Topic (verbal morphology signalling topicalization of the actor)

GT=Goal-Topic (verbal morphology signalling topicalization of the goal)

The abbreviations used for Chamorro, in addition to AT, GT, and Loc, are:

Unm=Unmarked case

Erg=Ergative

sg=singular

and references cited there) and need not concern us greatly here. One intriguing property of the topic is its interaction with various forms of extraction. To form an argumental wh-question in Tagalog, for instance, the form of the verb which topicalizes the extracted argument must be used; for instance, extraction of the direct object in (28) forces the use of the form of the verb in (28b):

Tagalog

- (29) a. Ano ang **binili** ni Maria sa tindahan?
 what T GT-bought A Maria Loc store
 ‘What did Maria buy at the store?’
 b.*Ano ang **bumili** si Maria sa tindahan?
 what T AT-bought A Maria Loc store

Again, the reasons for this property of the Tagalog topic are irrelevant to the discussion here. What we have seen is that Tagalog exhibits a type of verbal morphology the form of which is sensitive to the presence of a nominal with special syntactic properties, here referred to as the “topic”.

The data in (29) are strongly reminiscent of the phenomenon of “wh-agreement” in the related language Chamorro (in fact, some of the relevant morphemes seem to be etymologically related). Some Chamorro examples of this phenomenon are given in (30) :

Chamorro (adapted from Chung 1982, 49-50)

- (30) a. Hafa **finahan-ña** si Maria gi tenda
 what GT-buy-3sg Unm Maria Loc store
 ‘What did Maria buy at the store?’
 b. Hayi **fuma’gasi** i kareta?
 who AT-wash the car
 ‘Who washed the car?’

In Chamorro, as in Tagalog, verbs in clauses from which certain types of extraction has taken place bear special morphology, the form of which is determined by the theta-role

borne by the extracted nominal. The Chamorro facts differ from the Tagalog facts in one crucial respect, however, which has led to the use of the term “wh-agreement” for the Chamorro data; this morphology is only used in Chamorro in cases of extraction :

Chamorro (Aoun 1985a, xiv)

(31) a. Ha -fa'tinas i patgun i sena

3sgErg make the child the dinner

‘The child made dinner’

b. *Fuma'tinas i patgun i sena

AT-make the child the dinner

On the other hand, as we have seen, a typical Tagalog sentence containing a verb will always involve topicalization morphology, regardless of whether extraction takes place.

One way of looking at these facts would be as follows. It appears that there is a syntactic position fairly high in the clause, the occupant of which triggers special verbal morphology in both Tagalog and Chamorro; I have referred to this here as the “topic” position. In Tagalog, this topic position can always be occupied (in fact, it typically must be). In Chamorro, on the other hand, the topic position can only be moved through in the process of extraction; Chamorro nominals, unlike Tagalog nominals, do not land in the topic position and stop⁴. The Chamorro topic position, then, appears to be of the type under discussion here.

As Chung (1982), Aoun (1985a, 1985b), Dukes (1993), and Nakamura (to appear) have noted, Chamorro wh-agreement can also appear in the complements of control verbs:

⁴In Richards (1993) I argued that Tagalog topicalization is triggered by a weak feature as well. If I am right, then the difference between Chamorro and Tagalog is a purely morphological one; the relevant verbal morphology is present in Chamorro only when the appropriate specifier is occupied at some point in the overt syntax, while in Tagalog it is always present, even if the feature associated with it is not checked until the covert syntax. Chamorro topicalization morphology, on this account, would be not unlike French participial agreement; see section 1.1.

Chamorro (Chung 1982, 49)

- (32) *Malägu' gui' bumisita si Rita*
 want he AT-visit Unm Rita
 'He wants to visit Rita'

As we mentioned above, this is the kind of case the theory developed here leads us to expect. An empty category (in this case, PRO) may freely move into a position which cannot be overtly occupied by lexical elements. Here there is no question of PF being unable to choose between positions in a chain to pronounce, since the moving element is phonologically empty. PRO is therefore free to undergo overt movement to check off a weak feature, thus triggering wh-agreement⁵.

1.6 Agreement in Mohawk, Chichewa, and Irish

There are several other cases in the literature of agreement morphology which can only be triggered by elements which are either extracted or phonologically null. The account under development here should extend unproblematically to cover these.

1.6.1 Mohawk

Baker (1996) argues persuasively that overt NPs may not occupy specifiers of verbal agreement projections in Mohawk. Apart from wh-phrases, to which we will shortly return, each overt argument appears in a dislocated position, coindexed with a pro in the specifier of an agreement projection. Here I will simply summarize one of his arguments, which has to do with the distribution of Condition C effects in Mohawk.

Condition C effects do appear in Mohawk, as can be seen in the contrast in (33):

⁵Aoun (1985a, 1985b) points out that Chamorro pro-drop does not license the use of *-um-*:
Chamorro (Aoun 1985a, xiv)

(i) *ha-taitai i lepblu*
 3sgErg-read the book
 'He read the book'

(ii) **tumaitai i lepblu*
 AT-read the book

If the account developed here is on the right track, some independent factor will have to rule out (ii); one possibility is that this has something to do with the conditions on identifying pro.

Mohawk (Baker 1996, 43–44)

- (33) a. Wa- hi- hróri- ' tsi Sak ruwa- núhwe'-s
 FACT-1sS/MsO-tell-PUNC that Sak FsS/MsO-like- HAB
 'I told him_i/*_j that she likes Sak_j'
- b. Wa- hí- 'nha' -ne' ne tsi Sak ra- yo'ta- hser- íyo
 FACT-1sS/MsO-hire-PUNC because Sak MsS-work- NM-be.good
 'I hired him_i/_j because Sak_j is a good worker'

The facts in (33) are reminiscent of standard Condition C facts in English; a pronoun in the main clause cannot be coindexed with a name in a complement clause, which it presumably c-commands, but can be coindexed with a name in an adjunct clause.

Interestingly, however, nominal arguments containing names behave for purposes of Condition C like the adjunct clause in (33b), rather than like the argument clause in (33a):

Mohawk (Baker 1996, 45)

- (34) a. Ro- ya'takéhnha- s Sak raó- a'share'
 MsO- help- HAB Sak MsP-knife
 'Sak_j's knife helps him_i/_j'
- b. Wa'- t- há- ya'k- e' Sak raó- a'share'
 FACT-DUP-MsS-break-PUNC Sak MsP-knife
 'He_i/_j broke Sak_i's knife'

The fact that coreference is possible in (34b) suggests that the overt direct object *Sak raó'a'share'* 'Sak's knife', like the adjunct clause in (33b), is adjoined in a position above that occupied by the null pronominal subject. Baker argues that this is generally the case for overt nominals; with one exception, only *pro* can occupy the specifier of an agreement projection in Mohawk.

Wh-words are the exception to this rule, according to Baker. Mohawk wh-phrases can in fact occupy Spec AgrP in Mohawk, but must undergo overt wh-movement out of

AgrP. One of the few restrictions on word order in Mohawk is that all wh-words must be overtly fronted to the position where they take scope:

Mohawk (Baker 1996, 72)

- (35) a. Úhka oh nahóta í- hr- ehr- e' wá- hse- k- e'?
- who what Ø-MsS-think-IMP FACT-2sS-eat-PUNC

'Who thinks you ate what?'

- b. *Úhka í- hr- ehr- e' oh nahóta wá- hse- k- e'?
- who Ø-MsS-think-IMP what FACT-2sS-eat-PUNC

'Who thinks you ate what?'

Furthermore, there is evidence from weak crossover that wh-phrases, unlike other overt arguments, are base-generated in argumental positions and undergo movement out of them, rather than being base-generated in adjoined positions. Mohawk exhibits weak crossover effects of a familiar type:

Mohawk (Baker 1996, 79)

- (36) a. Úhka wa- ha- at- hróri- ' tsi raúha raó- skare' yako-nuhwáktani?
- who FACT-MsS-SRFL- tell- PUNC that him MsP-friend FsO- sick/STAT

'Who_i told that his_{i/j} girlfriend was sick?'

- b. Úhka akaúha akó- skare' í- hr- ehr- e' yako-ya't- á- hskats?
- who her FsP-friend Ø-MsS-think-IMP FsO-body-Ø-beautiful

'Who_j does her*_{i/j} boyfriend think is beautiful?'

Thus, it appears that in Mohawk, as in English, wh-words can only bind pronominals if a trace of the wh-word in an A-position c-commands the pronominal. The cases in (36) all involve long-distance wh-movement. Now consider cases of local wh-movement:

Mohawk (Baker 1996, 79-80)

- (37) a. Úhka yako- ya'takéhnha- s ne akaúha ako- núhkwa?
 who NsS/FsO- help- HAB NE her FsP-medicine
 'Who_i did her*_{i/j} medicine help?'
 b. Úhka wa'- akó- ati- ' ne akaúha ako-núhkwa?
 who FACT-FsO-lose-PUNC NE her FsP-medicine
 'Who_i lost her*_{i/j} medicine?'

Binding of a pronominal by a *wh*-word is impossible when the *wh*-word and the NP in which the pronominal is contained are clausemates. Recall that the NP in which the pronominal is contained is in an adjoined position, above the base position of the *wh*-word. There will therefore never be an A-position associated with the *wh*-word which c-commands a clausemate NP's base position. We should expect, then, that a *wh*-word in Mohawk will never be able to bind a pronominal variable in a clausemate NP, as this will always involve a weak crossover violation.

To summarize, then, it appears that in order to occupy the specifier of an agreement projection in Mohawk, nominals must either be phonologically null or must move out of the specifier in the overt syntax. Mohawk AgrPs are therefore cases of the type of phenomenon this theory seeks to explain; on the account being developed here, Agr features are weak in Mohawk, and can therefore only trigger movement if PF either will not need to decide which part of the movement chain to pronounce (as in the case of phonologically null nominals) or if later movement triggered by a strong feature will give PF clear instructions about which part of the chain to pronounce.

A remaining question, on the account developed here, is why AgrPs must necessarily have their specifiers filled in the overt syntax in Mohawk; why can nominals not simply remain in situ? Leaving potentially moveable elements in situ is standardly taken to be the reaction to weak features, and something will therefore need to be said about

why this course is not taken in the Mohawk case. We will see another case with the same property in section 1.5, where I will discuss the matter further.

1.6.2 Chicheŵa

Bresnan and Mchombo (1987) discuss a phenomenon in Chicheŵa which bears a strong resemblance to the Mohawk facts described above. The phenomenon in question is the distribution of object agreement in Chicheŵa. Chicheŵa verbs always agree with their subjects in person, number, and gender, and may optionally agree with their objects as well. If no object agreement is present, the object must immediately follow the verb⁶, but word order is otherwise free:

Chicheŵa (Bresnan and Mchombo 1987, 744)

- (38) a. Njũchi zi- ná- lúm- a alenje (SVO)
 bees SM-PAST-bite-INDIC hunters
 ‘The bees bit the hunters’
- b. Zínálúma alenje njũchi (VOS)
- c. *Alenje zínálúma njũchi (OVS)
- d. *Zínálúma njũchi alenje (VSO)
- e. *Njũchi alenje zínáluma (SOV)
- f. *Alenje njũchi zínáluma (OSV)

Object agreement may appear when the object is a null pronominal, but not with overt pronominals:

⁶ The object does not appear to have incorporated into the verb in these examples; inflectional morphology intervenes between the verb stem and the direct object.

Chicheŵa (adapted from Bresnan and Mchombo 1987, 769)

- (39) a. Fîsi a- ná- ú- dy- a
hyena SM-PAST-OM-eat-INDIC
'The hyena ate it'
- b. Fîsi a- ná- dy- á ʼtwo
hyena SM-PAST-eat-INDIC it
'The hyena ate it'

Similarly, object agreement appears in relative clauses and cleft constructions, cases presumably involving movement of a null operator:

Chicheŵa (Bresnan and Mchombo 1987, 769)

- (40) a. Ndi- ku- lír- ír- a mkángó uméné fîsi á- ná- ú- dy- a
SM-PRES-cry-APPL-INDIC lion REL hyena SM-PAST-OM-eat-INDIC
'I'm crying for the lion that the hyena ate'
- b. Sí mkángó uwu uméné fîsi á- ná- ú- dy- a
NEG.COP. lion this REL hyena SM-PAST-OM-eat-INDIC
'It's not this lion that the hyena ate'

On the other hand, wh-in-situ cannot trigger object agreement:

Chicheŵa (Bresnan and Mchombo 1987, 759)

- (41) a. (Kodí) mu- ku- fún- á chiyâni?
Q SM-PRES-want-INDIC what
'What do you want?'
- b. *(Kodí) mu- ku- chí- fún- á chíyâni?
Q SM-PRES-OM want-INDIC what
'What do you want?'

Finally, object agreement may appear with ordinary objects, which then exhibit a greater freedom of word order than objects with which agreement has not taken place. Bresnan and Mchombo analyze the direct object in these cases as a dislocated topic:

Chicheŵa (Bresnan and Mchombo 1987, 745)

- (42) a. Njũchi zi- ná- wá-lum- a alenje (SVO)

bees SM-PAST-OM bite-INDIC hunters

‘The bees bit them, the hunters’

- b. Zináwáluma alenje njũchi (VOS)

- c. Alenje zináwáluma njũchi (OVS)

- d. Zináwáluma njũchi alenje (VSO)

- e. Njũchi alenje zináwáluma (SOV)

- f. Alenje njũchi zináwáluma (OSV)

Bresnan and Mchombo give evidence from the behavior of tones in Chicheŵa that the object in examples like (42a) and (42b) has undergone string-vacuous rightward extraposition. Phrase-final high tones in Chicheŵa retract to penultimate low tones, yielding a rising tone. The distribution of this effect therefore provides evidence for the right boundary of the Chicheŵa VP. For instance, the subjunctive ending *-é* has a high tone when it is followed by VP-internal material, such as an object which does not trigger object agreement or a manner adverb:

Chicheŵa (Bresnan and Mchombo 1987, 750)

- (43) a. Ndikufúná kutí áná ánga [a- pitiriz -é phúnziro]

I-want that children my SM-continue-SUBJN lesson

‘I want my children to continue the lesson’

- b. Ndikufúná kutí áná ánga [a- pitiriz -é páng’ónó pang’ôno]

I-want that children my SM-continue-SUBJN a.little a.little

‘I want my children to continue slowly’

On the other hand, when pronounced in isolation or followed by non-VP-internal material, such as a postposed subject, the high tone on the subjunctive ending retracts to the penultimate syllable:

Chicheŵa (Bresnan and Mchombo 1987, 750)

- (44) Ndikufúná kutí [a- pitírĩz -e] aná ánga
 I-want that SM-continue-SUBJN children my
 'I want my children to continue'

A direct object which triggers object agreement also triggers retraction of the high tone, thus patterning with the postposed subject in (44). This suggests that such objects are in fact VP-external⁷:

Chicheŵa (Bresnan and Mchombo 1987, 750)

- (45) Ndikufúná kutí aná ánga [a- li- pítírĩz -e] phunziro
 I-want that children my SM-OM-continue-SUBJN lesson
 'I want my children to continue it, the lesson'

If additional VP-internal material is added to an example like (45), the high tone on the subjunctive suffix reappears:

Chicheŵa (Bresnan and Mchombo 1987, 751)

- (46) Ndikufúná kutí aná ánga [a- li- pítírĩz -é ndí inu] phunziro
 I-want that children my SM-OM-continue-SUBJN with you lesson
 'I want my children to continue it with you, the lesson'

Thus, *Chicheŵa* object agreement, like Mohawk agreement generally, seems to be triggered only by phonologically null elements and by elements which undergo some kind of movement. *Chicheŵa* object agreement does appear to differ from Mohawk agreement in that it is optional; not all objects need trigger object agreement.

1.6.3 Irish

McCloskey and Hale (1984) examine a form of agreement in Irish which is reminiscent of the two forms of agreement discussed above in that it is incompatible with the presence of an overt nominal:

⁷ The high tones on the first two syllables of the verb in this example are caused by the presence of the object agreement morpheme.

Irish (adapted from McCloskey and Hale 1984, 491, 507)

- (47) Chuirfeadh (*mé) isteach ar an phost sin
 put-CONDIT-1sg I in on that job
 'I would apply for that job'

- (48) Bhí mé ag caint leofa (*iad) inné
 was I talk-PROG with-3pl them yesterday
 'I was talking to them yesterday'

Uninflected verbs and prepositions are compatible with lexically overt nominals:

Irish (adapted from McCloskey and Hale 1984, 490, 507)

- (49) Chuirfeadh Eoghan isteach ar an phost sin
 put-CONDIT Owen in on that job
 'Owen would apply for that job'

- (50) gan iad
 without them
 'without them'

Whether Irish agreement is compatible with overt wh-movement is less clear. There is some controversy about the nature of wh-extraction in Irish; both sides of the controversy are compatible with the approach developed here, as far as I can see.

Irish wh-extraction may optionally involve a resumptive pronoun, as shown in

(51):

Irish (adapted from McCloskey and Hale 1984, 497)

- (51) a. an rud aL choinníonn tú ceilte orthu
 the thing that keep-PRES you concealed on-them
 b. an rud aN gcoinníonn tú ceilte orthu é
 the thing that keep-PRES you concealed on-them it

Irish agreement patterns with resumptive pronouns in a number of respects; I will cite only one here. Note that the examples in (51) involve different complementizers,

distinguishable by their phonological effects on the following verb. The complementizer used without resumptive pronouns triggers lenition in the following verb (and is therefore referred to as *aL* in the literature), while the complementizer associated with resumptive pronouns triggers nasalization (and is therefore known as *aN*). In cases of long-distance extraction, extractions which trigger the use of *aL* must trigger it in all intervening clauses:

Irish (McCloskey and Hale 1984, 497)

- (52) an rud aL shíl mé aL dúirt tú aL choinneofá ceilte orthu
 the thing that thought I that said you that keep-CONDIT-2sg concealed on-them
 ‘the thing that I thought you said you would keep hidden from them’

The complementizer *aN*, on the other hand, typically appears in only the highest clause. The lower clauses use the particle *goN*, which is the complementizer for ordinary complement clauses:

Irish (McCloskey and Hale 1984, 498)

- (53) an rud aN gceapaim goN gcoinníonn siad ceilte orainn é
 the thing that think-PRES-1sg that keep-PRES they concealed on-us it

The pattern of complementizers used with extraction from a site that controls agreement is exactly parallel to that used for extraction with a resumptive pronoun:

Irish (McCloskey and Hale 1984, 498)

- (54) na daoine aN raibh mé ag dúil goN gcuirfidís isteach ar an phost sin
 the people COMP was I expect-PROG that put-CONDIT-3pl in on that job
 ‘the people that I expected that they would apply for that job’

Whether we believe that wh-extraction can license Irish extraction, then, depends on our theories about the properties of resumptive pronouns, which seem to pattern syntactically with agreement as far as extraction is concerned. If resumptive pronouns indicate the absence of a movement relation between a wh-word and the extraction site, then agreement presumably also indicates such a relation; thus, wh-movement in Irish cannot license agreement. If, on the other hand, wh-movement can leave a resumptive

pronoun behind, then agreement is consistent with *wh*-movement in Irish (although it need not be present). Both of these positions have been defended in the literature on Irish syntax.

McCloskey (1990) develops an account of these facts in which the presence of a resumptive pronoun, and the use of the complementizers which accompany resumptive pronouns, signals the use of a non-movement strategy for extraction. On this account, when an operator is base-generated in an argument position and extracted directly from that position, no resumptive pronoun appears and a distinct set of complementizers is used. If this account is correct, the fact that the complementizers used with resumptive pronouns are also used when extraction is from a position controlling agreement shows that *wh*-extraction cannot license Irish agreement; on this account, in other words, Irish differs from Mohawk in that agreement may be licensed by a phonologically null NP but not by *wh*-extraction. Irish and Mohawk would then differ in much the way Italian and French do; as we have seen, *wh*-extraction can trigger participial agreement in French, but not in Italian.

Noonan (to appear) develops an alternative approach to these facts, in which both resumptive pronouns and their absence involve a movement relation between the extracted operator and the extraction site. On this approach, Irish verbal and prepositional agreement is more like French participial agreement; agreement may optionally be triggered by moved *wh*-words.

1.7 Derived Object Condition

Postal (1974) notes the existence of a class of verbs in English which are capable of behaving like ECM verbs only when the subject of the lower clause undergoes some kind of movement:

- (55) a. *John wagered Mary to have won the race
 b. Mary_i was wagered t_j to have won the race
 c. Mary, who_j John wagered t_j to have won the race...

Here, again, we have a somewhat familiar pattern. On the assumption that case assignment to the object involves movement to a position where Case features are checked, this movement is apparently only possible if the moved object then undergoes further movement out of the position. Similar facts have been observed in a number of other languages (see Kayne 1983b for a discussion of a similar phenomenon in French, and Rizzi 1982 for discussion of an Italian equivalent). The difference between *wager*-type verbs and true ECM verbs like *believe*, on this account, would be that *wager* verbs have an obligatorily weak object Case-checking feature, while verbs like *believe* can cooccur with a strong feature.

As in the case of Mohawk agreement, movement to the case position for these verbs appears to be obligatory; failure to use one of the class of elements which can land in this position and then move on yields ungrammaticality, rather than simply resulting in failure to move, as is standardly taken to be the case with weak features. What we need, then, is some principle forcing raising in these cases despite the weakness of the feature to which raising takes place.

In this case, there is some independent evidence that some form of obligatory overt raising takes place in English ECM constructions, and that ECM subjects and ordinary direct objects differ in this regard. The word order alternation in (27) has been seen (for instance, by Johnson (1991) and Koizumi (1993), (1995)) as a case of optional overt Object Shift in English:

- (56) a. I can't make the US out on this map
 b. I can't make out the US on this map

On this account, (56a) is derived from (56b) by shift of the object over the particle *out*.

Interestingly, particle verbs in ECM contexts show a different pattern, suggesting that this shift is obligatory for ECM subjects⁸:

⁸ Several previous works have described sentences like (57b) as being relatively acceptable (e.g., Kayne (1985), Johnson (1991), den Dikken (1992)), but for me and all the native English speakers I have talked to such examples are quite bad. I have no real explanation for this contrast in judgments; it is perhaps worth

- (57) a. He made the US out to be one of the major sponsors of international terrorism
 b. *He made out the US to be one of the major sponsors of international terrorism
- See Postal (1974), Lasnik and Saito (1991), Branigan (1992), Koizumi (1993), (1995), Bošković (1995b) for further discussion of object shift in ECM constructions.

The question, then, is what the relevant syntactic difference between (56) and (57) is. The null assumption would be that the Case properties of the higher verb are identical in each case; in principle, the attractor involved should be optionally strong in (57), as in (56). If this is correct, the principle which forces raising, ruling out (57b), will have to be something unrelated to the feature strength of the attracting feature.

One other case in which shift over a particle is obligatory is that of double-object verbs (Koizumi 1993, 125):

- (58) a. *The secretary sent out the stockholders a schedule
 b. The secretary sent the stockholders out a schedule

It seems reasonable to try to unify this fact with the similar ECM facts. We might, for instance, say that the ECM subject and the clause from which it raises are licensed in the same way as double objects like those in (58). This type of licensing is apparently subject to a requirement, not present for simple transitives, which forces raising of the higher of the two internal arguments.

Another, similar phenomenon is the distribution of Stylistic Inversion in French (see Kayne 1972, Kayne and Pollock 1978, Deprez 1988, Valois and Dupuis 1992, and references cited there for discussion). French subjects may be postposed in *wh*-extraction contexts:

noting that the discussion in the above works centers on contrasts between sentences like (57b) and those like (i) and (ii):

- (i) *He made out it to be one of the major sponsors of international terrorism
 (ii) *He made out the US a terrorist nation

I agree that (i) and (ii) are even worse than (57b), but this is consistent with an approach in which (i)-(ii), have defects over and above that responsible for the ungrammaticality of (57b). In the case of (i), this would presumably have to do with whatever forces even ordinary direct object pronouns to shift over particles. I do not know what rules out (ii), but I find even the shifted version of it quite bad (in fact, I am not sure that it is appreciably better, although den Dikken (1992) disagrees).

French (Kayne and Pollock 1978, 595)

- (59) a. Quand partira ton ami?
 when will-leave your friend
 ‘When will your friend leave?’
- b. Je me demande quand partira ton ami
 I me ask when will-leave your friend
 ‘I wonder when your friend will leave’

However, Stylistic Inversion is impossible when the verb is transitive:

French (Valois and Dupuis 1992, 327)

- (60) a. *Je me demande quand mangera sa pomme Marie
 I me ask when will-eat her apple Marie
 ‘I wonder when Marie will eat her apple’
- b. *Je me demande quand mangera Marie sa pomme
 I me ask when will-eat Marie her apple
 ‘I wonder when Marie will eat her apple’

In general, then, there seems to be something outlawing “overcrowding”; if there are too many arguments in a particular position, one of them must move. The exact nature of this factor is unimportant for our concerns here; see de Wind 1997, Alexiadou and Anagnostopoulou 1997, and references cited there for some further discussion. For our purposes, it is enough to note that something other than the feature strength of the Case feature in the higher clause apparently forces overt movement of an ECM subject. We have seen that in a number of cases raising is apparently forced regardless of feature strength.

Given this, we are in a position to account for the peculiar behavior of the DOC cases ((55), repeated as (61)):

- (61) a. *John wagered Mary to have won the race
 b. Mary_i was wagered t_j to have won the race
 c. Mary, who_i John wagered t_j to have won the race...

The hypothesis developed here has been that the factor distinguishing between verbs like *wager* and regular ECM verbs like *believe* is the fact that verbs like *wager* cannot be associated with a strong Case feature for the object. This accounts for the contrast between (61a) and (61b-c). In (61a), movement of *Mary* has occurred to check a weak feature, and PF is therefore unable to determine which member of the movement chain to pronounce. In (61b-c), on the other hand, the ECM subject has undergone further movement for a strong feature, and PF is therefore provided with instructions to pronounce the head of the movement chain. The question which remains, on this account, is why (61a) cannot be made well-formed by leaving *Mary* in its in-situ position, the usual response to a weak feature on an attractor. We have seen, however, that this is never possible with ECM verbs; ECM subjects, unlike ordinary direct objects, must always undergo overt movement. We thus have good reason to believe that some independent factor rules out this possibility, and thus makes (61a) ill-formed.

In section 1.4.1 we saw another case in which leaving an argument in situ was ruled out as a response to a putatively weak feature. Mohawk arguments must apparently agree with the verb, and this agreement is weak. The question arises in that case, as in the case just discussed, why an argument cannot simply remain in situ and fail to interact with agreement at all in the overt syntax. In this section we have seen that factors other than feature strength can sometimes force overt raising, and I will have to hope that one of these can be found for the Mohawk case as well.

1.8 Bulgarian vs. Serbo-Croatian

Now we are in a position to resolve a paradox that arose in chapter 2 regarding the behavior of multiple wh-movement in Serbo-Croatian. Recall that Serbo-Croatian differs from Bulgarian in that it exhibits wh-island effects. On the theory developed in chapter 2, this indicates that Serbo-Croatian CP cannot have multiple specifiers:

(62) *Serbo-Croatian* (from Rudin 1988, 459)

a. * Šta si me pitao ko može da uradi?

what AUX-2s me asked who can to do

‘What have you asked me who can do?’

Bulgarian (from Rudin 1988, 457)

b. ? Koja ot tezi knigi se čudiš koj znae koj prodava?

which of these books wonder-2s who knows who sells

‘Which of these books do you wonder who knows who sells?’

However, Bošković (1995c) points out that multiple wh-movement out of a non-wh clause is possible in Serbo-Croatian:

Serbo-Croatian (Bošković 1995c, 8)

(63) Ko si koga tvrdio da t je istukao t

who AUX whom claimed that AUX beaten

‘Who did you claim beat whom?’

On the assumption that long-distance wh-movement must be successive-cyclic, this indicates that Serbo-Croatian may have multiple specifiers of CP, just in case none of the moving wh-words remain in any of the specifiers. We have now seen that this is a common phenomenon. A position associated with a weak feature may quite generally be checked in the overt syntax by an element which undergoes further movement, driven by a strong feature, to a higher position.

We are entitled to wonder at this point why the escape hatch which allows long-distance wh-movement in (63) does not allow it in (62a); that is, why one of the wh-words cannot check the strong [+wh] feature on the lower C⁰ while the other checks the weak feature involved in successive-cyclic movement and then raises to the higher C⁰. I will discuss this question in section 2.6.2.2.1 of chapter 5.

1.9 Conclusion

We have seen that a number of syntactic positions have the property that they cannot serve as final landing sites, although they may be landed in in the course of movement to higher positions. One question which came up in the previous discussion had to do with the different properties of different types of movement. For instance, participial agreement in French is obligatory for A-movement, but optional for A'-movement:

French (adapted from Branigan 1992, 33-34)

- (64) a. Les livres de Jules Verne ont tous été imprimés /*imprimé
 the books of Jules Verne have all been printed-PL printed
 ‘Jules Verne’s books have all been printed’
- b. Quelle lettre Josèphe a-t-il écrite / écrit?
 what-FEM letter Joseph has-he written-FEM written
 ‘Which letter did Joseph write?’

Similarly, Italian participles must agree with A-moved elements but cannot agree with A'-moved elements⁹:

Italian

- (65) a. *Ho creduti gli Australiani
 have-1sg believed-PL the Australians
 ‘I have believed the Australians’
- b. Gli Australiani sono stati creduti
 the Australians are been-PL believed-PL
 ‘The Australians have been believed’
- c. *Quanti Australiani hai creduti?
 how-many Australians have-2sg believed-PL
 ‘How many Australians have you believed?’

⁹ I will omit discussion of cliticization here, as its syntactic nature is quite unclear; for purposes of the phenomena discussed here it appears to pattern with A-movement (but see Deprez 1994 for cases of cliticization which fail to trigger participial agreement).

This paradigm raises a number of questions. One is why A-movement and A'-movement should differ as they do; another is what is responsible for the different behavior of French and Italian.

In Richards (in press) I suggested that we might exploit Chomsky's (1995) notion of feature-movement to deal with these facts. On this approach to movement, particular features are attracted by other features in order to enter into checking relations. For poorly understood reasons having to do with well-formed relations between features of a given element, however, a single feature often is unable to move by itself, and must "pied-pipe" the rest of the category with it. Thus, for instance, in the case of overt wh-movement, only the [+wh] feature of the wh-word is attracted, but the rest of the features typically follow the [+wh] feature in order to satisfy additional well-formedness requirements.

Returning to the participial agreement facts, then, we see that movement that involves checking of ϕ -features--that is, A-movement--is the type of movement that forces participial agreement. This state of affairs can be made sense of under the feature-movement approach, if we assume that overt pied-piping of ϕ -features is in principle optional. For movements which involve checking of ϕ -features, then, the ϕ -features must move and agreement must take place, but for movements which do not, ϕ -feature pied-piping is optional (in French) or impossible (in Italian).

On this account, French wh-movement may involve either movement of the [+wh] feature alone or movement of both the [+wh] feature and its ϕ -features. The idea that a wh-word and its trace may be connected in either of two ways is not a new one; notable examples of this idea in the literature include Rizzi's (1990) and Cinque's (1990) idea of a contrast between referential and non-referential wh-phrases, and Manzini's (1991) address-based and categorial dependencies. Accounts of this type attempt to deal with the distinct effects of islands on arguments and adjuncts by positing a type of wh-dependency which is immune (or less susceptible) to islands and which is available only to arguments.

Frampton (1990) and Cresti (1995) suggest that distinctions of this type should be linked to the semantic notion of scope reconstruction (see also Kroch 1989, Dobrovie-Sorin 1993, Hornstein 1995, Dekydtspotter 1996 for discussion). Examples like (66) have long been recognized as having at least two readings, given here as (66a) and (66b) (Cresti 1995, 83):

- (66) How many people should I talk to?
- a. For what (number) *n*: there are *n*-many people x_i such that I should talk to x_i .
 - b. For what *n*: it should be the case that there are *n*-many people that I talk to.

(66a) would be appropriate in a context in which there are a number of particular people who are experts on the subject I want to learn about, and I am trying to find out how many of these people there are. (66b), by contrast, might be used in a situation in which I must talk to a certain number of people (to make my results statistically valid, for instance), but in which the identity of these people is unimportant. The ambiguity, then, has to do with the relative scope of the modal *should* and an indefinite or existential associated with the *wh*-phrase. I will refer to this as the “existential”; its exact semantic properties will not be very important to the theory under development here (though we will crucially need to assume that some analog of this is present in every NP). The *wh*-movement may be interpreted either as having brought this existential along with it, as in (66a), or as having left it in situ, as in (66b). As Longobardi (1987) notes, that these two semantically distinct types of dependencies behave differently with respect to weak islands; only the first type, but not the second, may cross a weak island (Cresti 1995, 84):

- (67) How many people do you think I should talk to?
- a. For what n : there are n -many people x_i such that you think I should talk to x_i .
 - b. For what n : you think that it should be the case that there be n -many people that I talk to.
- (68) How many people do you wonder whether I should talk to?
- a. For what n : there are n -many people x_i such that you wonder whether I should talk to x_i .
 - b. *For what n : you wonder whether it should be the case that there be n -many people that I talk to.

The data in (67) are parallel to those in (66): long-distance wh-extraction, like local wh-extraction, can optionally pied-pipe the existential component of the wh-word. On the other hand, (68) shows that long-distance wh-extraction across a wh-island must obligatorily pied-pipe the existential; the reading on which the existential is left in situ is unavailable.

This semantic distinction is reminiscent of the syntactic distinction posited above as a possible way of dealing with the optionality of French participial agreement with wh-moved objects. Suppose we proceed on the hypothesis that they are in fact the same. The ϕ -features associated with a wh-feature, then, may optionally raise as free riders during wh-movement; if they raise, they must trigger participial agreement, and if they do not they cannot. Raising the ϕ -features yields the reading in which the existential quantifier associated with the wh-word takes wide scope; leaving the ϕ -features in situ gives a reading in which the existential takes narrow scope. More generally, we can assume that ϕ -features can in principle be given an indefinite or existential reading in any NP of which they are a part. This hypothesis makes a number of predictions about the interaction of participial agreement with the semantic properties of moved wh-words. I will try to show that these predictions are correct.

To begin with, the basic semantic contrast in question does appear to be sensitive to participial agreement in French. Obenauer (1992) points out the following contrast:

French (Deprez 1994, 8)

- (69) a. Combien de fautes a-t-elle faites?

how-many of mistakes has she made-AGR

‘How many (amongst a known set of) mistakes has she made?’

- b. Combien de fautes a-t-elle fait?

how-many of mistakes has she made

‘What is the number of things that are mistakes she has made?’

Agreement-triggering wh-movement must have a specific reading, while wh-movement which does not trigger agreement can have a non-specific, cardinal reading. Following Deprez (1994), we can understand the reading in (69a) as reflecting pied-piping of the existential into Diesing’s (1992) Restrictive Clause, thus giving a specific reading; the reading in (69b), then, involves leaving the restriction in the Nuclear Scope, giving a non-specific reading. As predicted, the reading which is associated with pied-piping of the restriction is also associated with participial agreement.

A similar contrast is found with long-distance wh-movement:

French (Marie Claude Boivin, p.c.)

- (70) a. Combien de nouvelles Jean croit -il que Marie a écrites?

how-many of novels Jean believes-he that Marie has written-AGR

‘For what n: there are n-many novels x_j such that Jean believes that Marie has written x_j ?’

- b. Combien de nouvelles Jean croit -il que Marie a écrit?

how-many of novels Jean believes-he that Marie has written

‘For what n: Jean believes that it is the case that there are n-many novels that Marie has written?’

Again, participial agreement creates a preference for the reading in which the existential associated with the *wh*-word has wide scope, while lack of agreement appears to indicate failure to pied-pipe the existential.

Cresti (1995) notes, following Frampton (1990), that leaving the existential quantifier in situ is necessary in existential constructions (Cresti 1995, 86):

- (71) How many people do you need there to be at the meeting?
- a. *For what *n*: there are *n* people x_i such that you need x_i to be at the meeting.
 - b. For what *n*: you need there to be *n* people at the meeting.

That is, (71) can only be interpreted as a question about how many people are necessary to make a quorum, not about the number of (specific) people that you have in mind as people you need to have at the meeting. The effect in (71) is presumably a special case of the well-known Specificity Effect associated with existential constructions.

According to the hypothesis being pursued here, then, pied-piping of the ϕ -features associated with a *wh*-word is ruled out in existential constructions. We expect to find that *wh*-extraction from an existential construction will be unable to trigger participial agreement in French, as the *wh*-word has no ϕ -features with which to check agreement. This is the case, as Kayne (1989) notes:

French

- (72) Qui sait [combien d'erreurs] il sera fait(*es) t?
 who knows how-many of-errors there were made AGR

Finally, the hypothesis under consideration here makes a prediction about the interaction of participial agreement with weak islands. We have seen that pied-piping of the existential component of a *wh*-word is obligatory if the *wh*-word crosses a weak island ((68), repeated as (73)):

- (73) How many people do you wonder whether I should talk to?
- a. For what *n*: there are *n*-many people x_i such that you wonder whether I should talk to x_i .
 - b. *For what *n*: you wonder whether it should be the case that there be *n*-many people that I talk to.

We have seen that there are semantic reasons to believe that participial agreement in French signals pied-piping of the existential. If this is correct, then we expect participial agreement to improve the status of extraction out of a weak island. Here the judgments are apparently quite subtle, but there is a slight improvement of the kind we expect, at least with some weak islands:

French (Marie Claude Boivin, p.c.)

- (74) a. *Quelle nouvelle Jean a-t-il nié le fait que Balzac a écrit t?
 which-FEM novel Jean has-he denied the fact that Balzac has written
 ‘Which novel did Jean deny the fact that Balzac wrote?’
- b. ??Quelle nouvelle Jean a-t-il nié le fait que Balzac a écrite t?
 which-FEM novel Jean has-he denied the fact that Balzac has written-FEM

Thus, we seem to have good semantic evidence, and somewhat less sturdy syntactic evidence, that the optionality of participial agreement with *wh*-words in French reflects optional pied-piping of that part of the *wh*-phrase which would trigger agreement. We have seen that this optional pied-piping has semantic effects of the expected kind. In the case of movement triggered by a need to check ϕ -features (for instance, passivization), of course, no optionality is expected, or found.

One question raised by the above discussion, of course, is why Italian consistently fails to show participial agreement with *wh*-words:

Italian

- (75) Quanti Australiani hai creduto/*creduti?
 how-many Australians have-2sg believed/believed-AGR

Assuming that Italian *wh*-movement exhibits the same semantic ambiguities as those available in French and English, the question arises of why those ambiguities are not signalled in Italian as they are in French by participial agreement. Italian participles appear to only agree with elements undergoing movement to check ϕ -features.

In Richards (in press) I suggested that this contrast between French and Italian might reflect different levels of “autonomy” of ϕ -features in the two languages. Italian ϕ -features, on this account, do not move as free riders in the overt syntax, but move overtly only in order to enter into a feature-checking relation. On this account, we must assume that Italian ϕ -features are capable of autonomous covert movement independently of the *wh*-phrase of which they are a part, giving rise to the semantic ambiguities present in French. French ϕ -features, by contrast, cannot undergo such autonomous covert movement; in order to be interpreted in a moved position, they must undergo movement as free riders on the attracted *wh*-feature.

Another possible ramification of the approach developed here would be a new way of approaching the properties of infinitival complements. We have seen that for ECM subjects, at least, some factor forces overt raising of the subject into the matrix clause. I hypothesized that this was a constraint on “overcrowding”, preventing both the ECM subject and the infinitival clause from remaining in situ. If this is on the right track, there is apparently a motivating force for raising which has nothing to do with feature strength. We have seen that this force can drive raising out of the clausal complements of *wager*-class verbs, even though the feature that would drive such movement is weak.

Clearly, many questions remain; we need to develop a theory of the forces other than feature strength that can drive raising. Suppose it is generally true, however, that raising to subject position (in English, at least) is always forced, by some factor other than feature strength. Suppose further that the feature that drives subject raising in infinitivals is weak. There would then be two ways in which the subject of an infinitive could be licensed; it could be phonologically null, which would render its chain irrelevant to PF, or

it could raise further out of the embedded subject position, perhaps into the matrix clause. The first of these options, on this account, would correspond to control infinitives, while the second option would be that used by ECM infinitives. For this approach to work, of course, we would need an account of why some infinitival complements must use one of these options and not the other, and at this point I have none to offer.

2 Destroying the evidence: ellipsis

Another case in which the theory under development here allows overt movement for a weak feature is that in which the tail of the potentially offending chain is within an ellipsis site:



If α is an ellipsis site, then PF presumably receives instructions not to pronounce anything within α . Thus, the chain created by movement in this case has only a single position which is a candidate for pronunciation. The chain is therefore a legitimate PF object. In the following sections I will discuss several phenomena which receive natural accounts on the assumption that the movement in (76) is in fact well-formed.

2.1 (Pseudo)gapping

Examples of what have classically been termed Gapping and Pseudogapping, respectively, are given in (77) and (78):

(77) Some bought books and others ~~bought~~ records

(78) Some have read books and others will ~~buy~~ magazines

In what follows I will concentrate on Gapping, although it seems to me that a similar account may be possible for Pseudogapping. I will follow Johnson (1994, 1997) in referring to the NPs in the second conjunct of (77) as the *remnants*, and to the material which is missing from this conjunct as *gapped material*.

A number of accounts of gapping have assumed that it involves movement of the remnants out of a constituent which then undergoes ellipsis (Pesetsky 1982, Jayaseelan

1990, and Lasnik 1995 are all accounts of this type; see Johnson 1994, 1997 for arguments against such an account, although he also assumes movement of gapping remnants in certain cases). There is a fair amount of evidence that gapping remnants must undergo some type of movement.

For instance, as Neijt (1979) argues extensively, the relation between the remnants is subject to such familiar conditions on movement as the *wh*-island condition:

(79) John tried to cook dinner today, and Peter ~~tried to cook dinner~~ yesterday

(Neijt 1979, 138)

(80) *John wondered what to cook today and Peter ~~wondered what to cook~~ tomorrow

In fact, it seems to me that (80) is worse than (81):

(81) ??John wondered when to serve the chicken and Peter ~~wondered when to serve the~~
tofu

The contrast between (80) and (81) is reminiscent of a familiar argument-adjunct asymmetry with respect to extraction out of islands:

(82) *When is John wondering whether to serve the tofu *t*?

(83) ??What is John wondering whether to serve *t* tomorrow?

This is to be expected, if gapping in (80) and (81) involves extraction of a remnant out of a *wh*-island. In (80), the extracted remnant *tomorrow* is an adjunct, and is thus more sensitive to the effects of a *wh*-island than the argumental extracted remnant *the tofu* in (81).

Similarly, Pesetsky (1982) notes that gapping across a tensed clause boundary exhibits a subject-object asymmetry of a kind which is familiar from the domain of *wh*-movement:

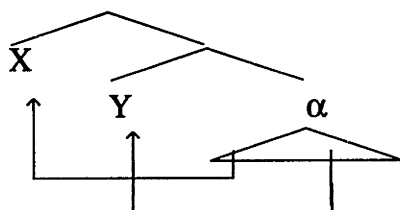
(Pesetsky 1982, 645)

(84) This doctor thinks that I should buy tunafish, and that doctor ~~thinks that I should~~
buy salmon

(85) *This doctor thinks that tunafish will harm me, and that doctor ~~thinks that salmon~~
will harm me

Suppose we take the above evidence for a movement relation between the remnants as conclusive, and assign to sentences involving gapping a structure like the one in (86), where X and Y are remnants and α is a constituent which undergoes ellipsis:

(86)



A number of questions now arise. One is why the paths in a structure like (86) must cross; a well-formed sentence like (87a), for instance, contrasts sharply in grammaticality with (87b), where the base order of the remnants has been reversed:

- (87) a. Some bought books and others_i records_j [α t_i bought t_j]
b. *Some bought books and records_j others_i [α t_i bought t_j]

Another question is why movement of the type undergone by the remnants is possible only when the VP is elided; (87a) contrasts with (88), where no ellipsis takes place:

- (88) *Some bought books and others_i records_j [α t_i bought t_j]

The first of these questions can be answered in terms of the theory developed in chapter 3; remnant movement in gapping involves movement to multiple specifiers of a single head, and therefore cannot alter the base order of the moving elements. The answer to the second question is that the features on this head which are responsible for attracting the remnants are weak in English, and thus cannot ordinarily be active in the overt syntax. VP ellipsis, however, makes these weak features capable of driving overt movement, as predicted by the theory developed here. The chains headed by the remnants have only a

single copy outside the ellipsis site, and are therefore legitimate PF objects, since they give PF unambiguous instructions as to which part of the chain to pronounce.

2.2 Multiple Sluicing

Bolinger (1978), Merchant (1996), and Nishigauchi (to appear) discuss English sentences of the following form:

(adapted from Bolinger 1978, 109)

(89) I know that in each instance one of the girls got something from one of the boys.

But they didn't tell me which from which

The second sentence of (89) appears to be a case of sluicing involving more than a single wh-word. Such cases of “multiple sluicing” are also attested in Japanese (Takahashi 1993, 1994b, Nishigauchi to appear), Bulgarian (Merchant 1996), and German (Sauerland 1995):

Japanese (Takahashi 1994b, 284-285)

(90) John -ga [dareka -ga nanika -o katta to] itta.

John NOM someone NOM something ACC bought that said

Mary-wa [dare -ga nani -o ka] siritagatteiru.

Mary TOP who NOM what ACC Q wants-to-know

‘John said someone bought something. Mary wants to know who what.’

Bulgarian (Merchant 1996, 4)

(91) Njakoј vidja njakogo, no ne znam [koј kogo]

someone saw someone but not I-know who whom

‘Someone saw someone, but I don't know who whom’

German (Sauerland 1995, 34)

(92) Ich habe jedem Freund ein Buch gegeben,

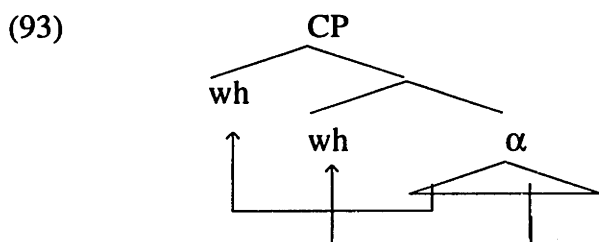
I have every friend a book given

aber ich weiß nicht mehr wem welches.

but I know not anymore whom which

‘I gave every friend a book, but I don't remember any more whom which’

The theory developed here provides us with a natural account of multiple sluicing which is very similar to the account of gapping developed in the previous section. Apart from Bulgarian, none of the languages above allow multiple overt wh-movement; that is to say, some or all of the features on C^0 which drive wh-movement are weak. However, we expect weak features to be able to drive overt wh-movement when the IP is elided; the chains headed by the wh-words will be legitimate PF objects, since they will only contain a single candidate for pronunciation:



The best English examples of multiple sluicing are like Bolinger's example (89) in that they involve one nominal wh-word and one wh-word which is a prepositional phrase. Still, it seems clear that there is a contrast in grammaticality between even a multiple sluicing case with multiple nominal wh-words and a parallel example in which the remnants are not wh-words:

(94) I know that everybody bought something.

a. ??I just don't know who what.

b. *I'm pretty sure John a car.

As Sauerland (1995) points out, this restriction on English multiple sluicing is reminiscent of a similar contrast noted by Moltmann (1995) concerning exceptives and *even*:

(Moltmann 1995)

(95) a. *Every man met every woman, except/even John Mary.

b. ?Every man danced with every woman in every room, except/even John with
Mary in the kitchen.

I have no account of the restriction to offer, however.

Nishigauchi (to appear) and Merchant (1996) both argue that multiple sluicing is related in some way to gapping, at least in English. It seems clear, however, that multiple sluicing is not simply a case of gapping. For one thing, as we have seen, multiple sluicing, like sluicing, is not subject to the strict locality restrictions which require a gapped clause to be conjoined with the corresponding non-gapped clause:

- (96) a. Mary bought a record, and John a car
 b. *Mary bought a record, and I'm pretty sure John a car.
- (97) ??I know that everybody bought something. I just don't know who what.
- (98) I know that somebody bought a car. I just don't know who.

Similarly, multiple sluicing is like sluicing in that it allows *wh*-word remnants to precede prepositions of which they are the objects:

- (99) I know John was talking with somebody, but I don't know who with
- (100) I know John was talking with somebody about something, but I don't know
 who with about what.

In this regard, too, multiple sluicing and sluicing both differ from gapping:

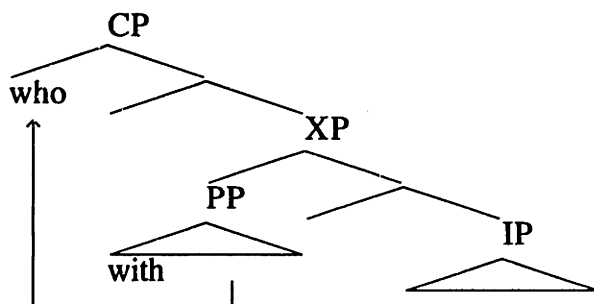
- (101) Mary was talking with John about ergativity,
 a. ...and with Bill about the stock market.
 b. *...and Bill with about the stock market.

This “postpositional” option is only available to a preposition which is the first remnant in a multiple sluicing sentence:

- (102) I know John was talking with somebody about something,
 a. ...but I don't know who with about what.
 b. *...but I don't know with who what about.
 c. *...but I don't know what with who about.
 d. *...but I don't know who with what about.
 e. *...but I don't know who what with about.

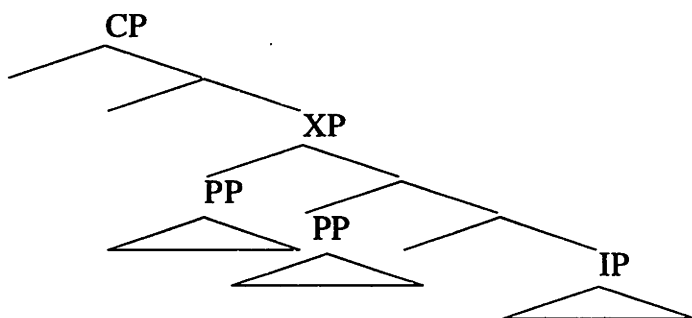
Suppose we assume that the use of a postposition involves movement of the wh-PP to a position outside the elided IP, followed by wh-movement to Spec CP which strands the preposition:

(103)



The data in (102) then follow straightforwardly. The derivations of all of the examples in (102) will involve the representation in (104):

(104)



Here both of the PP remnants have moved out of the elided IP. Movement of a wh-word out of its PP to Spec CP will therefore create a two-membered chain in which both members are visible to PF, since neither is inside the ellipsis site. (102d) and (102e), in which both prepositions follow their objects, are therefore ruled out, since English C^0 has only one strong feature. Given that only one of the wh-words can move to Spec CP, it must apparently be the higher of the two; thus, (102a) is well-formed, and (102b) and (102c) are ill-formed. This is a straightforward Superiority effect; C^0 must attract the closest wh-word.

2.3 Conclusion

The claim that weak features can be able to drive overt movement out of an ellipsis site thus seems to derive various properties of gapping and multiple sluicing in a fairly

unproblematic way. Before moving on to the next section, I will note one potential problem for the account given here, and a potential way of solving it.

Consider first the multiple-wh question in (105):

(105) Who said John bought what?

Now suppose we were to apply VP-ellipsis to (105), giving a sentence like (106):

(106) *(Who said Fred bought what, and) who said what John did buy?

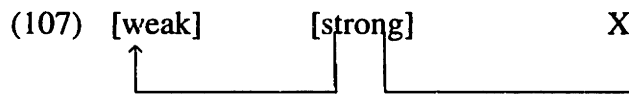
We might expect, on the theory developed so far, that *what* should be able to move out of the elided VP into the specifier of the non-interrogative embedded C⁰; the resulting chain would have only one member which would be a candidate for pronunciation, and would therefore be a well-formed PF object.

One way of accounting for the ill-formedness of (106) would be to say that the declarative/interrogative distinction is a matter of interpretation, rather than of a difference in the featural makeup of the complementizer in question; a complementizer is interrogative just in case it has a wh-word in its specifier. This is similar to Cheng's (1991) notion of clausal typing. The embedded CP in the second conjunct of (106), then, would be interpreted as interrogative rather than declarative.

If all complementizers are alike syntactically, then all English complementizers, for instance, must have a strong wh-feature; the only difference between interrogative and declarative complementizers is that the strong wh-feature on the latter is uninterpretable (and therefore can be deleted under checking). I will discuss this idea further in sections 3.2 and 4.3. We will be led to a particular approach to successive-cyclic movement, according to which the strong wh-feature on a declarative complementizer can attract a moving wh-word to it and subsequently delete when the wh-word moves to a higher, interrogative complementizer.

3 "Bottom-heavy" chains

A third type of overt movement which would be allowed by this theory would create the chain in (107):



In this chain, as in the other well-formed chains discussed here, PF receives instructions to pronounce a single element of the chain, which is associated with a strong feature. This case differs from those previously discussed in that the pronounced position is not the head of the chain. (107) will therefore be difficult to distinguish from a case of overt movement to the position associated with the strong feature, followed by covert movement to the position associated with the weak feature. We will need to look for cases in which overt movement triggers the creation of a relation between the moved element and some higher position which apparently fails to hold in cases in which the moved element is left in situ, and which is more like what we expect in cases of overt movement. The hope would be to show that such cases are in fact cases of movement in the overt syntax (that is, before Spell-Out) to a position associated with a weak feature, creating a chain of the type in (107). As noted in the introduction to this chapter, this theory draws a distinction between a chain of the type in (107) and a similar chain containing only the base position and a weak feature, thus distinguishing itself from Single Output Syntax approaches of the type developed by Bobaljik (1995), Brody (1995b), Groat and O'Neil (1996), and Pesetsky (to appear). On this theory, the relation in the chain in (107) between the highest position of the chain and the intermediate position is created by overt movement, while a chain consisting only of a single movement to check a weak feature is created by covert movement. The movement to the weak feature in (107) thus has something in common with movement to a strong feature, which movement to a weak feature typically lacks. On a Single Output Syntax approach, on the other hand, all movement to check weak features is the same; there is no syntactic distinction drawn between overt and covert movement.

3.1 Japanese

One candidate for a movement of the relevant type is Japanese wh-scrambling, as described by Takahashi (1993). Takahashi claims that scrambling can be a form of wh-

movement in Japanese; we might reinterpret this, in terms of the theory under development here, as meaning that the strong feature which drives scrambling can also drive further overt (but invisible) movement to Spec CP. One piece of evidence Takahashi gives for his conclusion is the observation that scrambling into a clause with a +wh C⁰ fixes the scope of the wh-word:

Japanese (Takahashi 1993, 660)

- (108) a. Kimi-wa [John-ga [Mary -ga **nani -o** tabeta ka] sitteiru to] omotteiru no?
 you TOP John NOM Mary NOM what ACC ate Q know that think Q
 ‘Do you think that John knows what Mary ate?’ OR
 ‘What do you think that John knows whether Mary ate?’
- b. Kimi-wa [**nani -o** John-ga [Mary -ga tabeta ka] sitteiru to] omotteiru no?
 you TOP what ACC John NOM Mary NOM ate Q know that think Q
 ‘Do you think that John knows what Mary ate?’ OR
 ‘What do you think that John knows whether Mary ate?’
- c. **Nani -o** kimi-wa [John-ga [Mary -ga tabeta ka] sitteiru to] omotteiru no?
 what ACC you TOP John NOM Mary NOM ate Q know that think Q
 ‘What do you think that John knows whether Mary ate?’

Thus, a wh-word which is overtly scrambled to a position between two interrogative C⁰’s must take scope at the higher of the two; (108c) has only the matrix question reading, and not the embedded question reading. As Takahashi notes, this is particularly interesting in light of Saito’s (1989) observation that wh-words can in principle be scrambled to a position c-commanding their scope:

Japanese (Takahashi 1993, 656)

- (109) a. John-ga [Mary -ga **nani -o** katta ka] sitteiru

John NOM Mary NOM what ACC bought Q knows

‘John knows what Mary bought’

- b. **Nani-o** John -ga [Mary -ga katta ka] sitteiru

what ACC John NOM Mary NOM bought Q knows

‘John knows what Mary bought’

Thus, scrambled wh-words can in principle undergo reconstruction; such reconstruction is blocked only when there is an interrogative complementizer above the wh-word to which it can move. The theory developed here offers a natural account of these facts. In (108c), the strong feature which triggers overt scrambling into the matrix clause also licenses further overt movement to check the weak interrogative feature of the highest C^0 . The result is a chain of the type in (107); the copy associated with the strong scrambling feature is the one pronounced, and the overt movement to the specifier of the matrix C^0 therefore has no effect on the phonological representation, but does enforce a particular semantic effect, namely matrix scope for the scrambled wh-word. In (109b), by contrast, the matrix C^0 is declarative, and no overt movement to its specifier takes place; reconstruction is therefore possible.

The fact that the scrambled wh-word in (108c) obligatorily has matrix scope forces us, on this theory, to an approach in which overt movement to check weak features is always forced in those cases in which it is possible. This is a conclusion to which we have been led before, notably in the account of null operator movement in Japanese in section 1.4 of this chapter¹⁰.

¹⁰ We are still entitled to wonder what favors the formation of a chain in which the weak feature checked overtly is in the matrix C^0 over one in which the weak feature is that of the embedded C^0 . Both of these chains would be well-formed PF objects and would yield the phonological form in (108c); see section 4.3 of this chapter for evidence that scrambling out of an interrogative Spec CP is syntactically well-formed in Japanese. It seems to me that a derivational account in which the “decision” to overtly check a weak feature is postponed until it is certain that such movement will not create an ill-formed PF object might give us the desired results, but this is yet another area I will have to leave open for the time being. See footnote 16

The line of reasoning outlined here also suggests a possible unification of the two types of languages discussed in chapter 2. In chapter 2 I argued, following an insight of Rudin's (1988), that languages may be divided into those which allow multiple specifiers of CP (which I called CP-absorption languages) and those in which multiple wh-words adjoin to IP-level projections (IP-absorption languages). We saw that one of the distinguishing characteristics of IP-absorption languages was the availability of scrambling. One possibility, then, is that all languages are in fact CP-absorption languages, but that wh-movement past scrambling positions violates Shortest Move. In a language with scrambling, then, wh-words are forced to scramble, and then to move (overtly or covertly) from their scrambled positions to Spec CP. When such movement is overt, it is necessarily invisible, since movement to a strong scrambling feature followed by movement for a strong wh-feature creates an ill-formed chain. Thus, languages with scrambling appear to scramble some or all of their wh-words rather than wh-moving them to Spec CP. A number of interesting issues now arise. For instance, on the (widely held) assumption that there is no covert analogue of scrambling, how would an account of this type extend to covert IP-absorption languages? Also, what is the mechanism whereby certain languages with scrambling seem to allow one wh-word to move directly to Spec CP in the overt syntax (e.g., in German and Serbo-Croatian)? Unfortunately, I will have to leave the full development of a theory of this type to future work.

3.2 Malay

A second candidate for a "bottom-heavy" chain comes from Saddy (1991) and Cole and Hermon's (1997) discussion of wh-movement in Malay. Malay allows either overt or covert wh-movement, and also exhibits partial wh-movement, in which a wh-word moves to a Spec CP below the one in which it takes scope:

for another case in which chains with a strong feature in an intermediate position are preferred over chains with a strong feature in the highest position and weak features in the intermediate positions.

Malay (Saddy 1991, 189)

- (110) a. Bill tahu Tom men-cintai **siapa**?

Bill knows Tom loves who

‘Who does Bill know that Tom loves?’

- b. Bill tahu **siapa** yang Tom cintai t?

Bill knows who FOC Tom loves

‘Who does Bill know that Tom loves?’

- c. **Siapa** yang Bill tahu Tom cintai t?

who FOC Bill know Tom love

‘Who does Bill know that Tom loves?’

Overt wh-movement is restricted by a number of island conditions which do not constrain covert movement. Relative clauses are one such island:

Malay (Cole and Hermon 1997, 8-9)

- (111) a. *Dengan **siapa** kamu sayang [perempuan yang telah berjumpa t]?

with who you love woman that already meet

‘Who do you love the woman who met t?’

- b. Kamu sayang [perempuan yang telah berjumpa **siapa**]?

you love woman that already meet who

‘Who do you love the woman who met t?’

Thus, we can use sensitivity to islands as a diagnostic for overt movement¹¹.

Interestingly, partially moved wh-words behave as though they had undergone overt movement to their scope position, for purposes of island constraints; in particular, islands which do not lie between the pronounced position of the wh-word and its trace can still render questions ill-formed if they intervene between the wh-word and the C⁰ where it takes scope:

¹¹ We saw in Chapter 2 that a “levels approach” of this type is incorrect for wh-islands, but it is conceivable that it is correct for other types of islands. Note, for instance, that although Japanese and Chinese differ with respect to wh-island effects, a fact which was made much of in Chapter 2, they both lack CNPC effects for argument extraction.

Malay (Cole and Hermon 1997, 21)

(112) *Kamu sayang [perempuan (dengan) **siapa** yang telah jumpa t]?

you love woman with who that already meet

‘Who do you love the woman that met?’

This is true even if the Spec CP occupied by the wh-word is an ordinary declarative complementizer, as in (113c):

Malay (Hooi Ling Soh, p.c.)

(113) a. Kamu beli [sebuah buku yang mengatakan

you bought a book that say

bahawa orang Malaysia telah cipta **apa**]?

that person Malaysia already invent what

‘What did you buy a book that says that Malaysians invented t?’

b.* Kamu beli [sebuah buku yang mengatakan

you buy a book that say

bahawa **apa** orang Malaysia telah cipta]?

that what person Malaysia already invent

‘What did you buy a book that says that Malaysians invented t?’

c.* **Apakah** yang kamu beli [sebuah buku yang mengatakan

what FOC you buy a book that say

bahawa orang Malaysia telah cipta t]?

that person Malaysia already invent

‘What did you buy a book that says that Malaysians invented t?’

Thus, the relation between a partially moved wh-word and the Spec CP where it takes scope is syntactically more like overt movement than it is like covert movement. The theory developed here can handle this fact straightforwardly. The wh-feature on C⁰ in Malay is apparently optionally strong, which accounts for the optionality of overt wh-movement. Without any additional assumptions, then, we expect that a sentence could in

principle be base-generated with a weak wh-feature in a matrix C^0 and a strong wh-feature in an embedded C^0 . This representation would yield partial movement; the wh-word would move overtly to check both wh-features, and would be pronounced in the position associated with the strong feature. Again, as in the Japanese case, we have evidence that overt movement to check the weak feature is not only an option, but is actually required; if it were not, partial movement would not exhibit the island effects seen in (108) and (109c), since the partially-moved wh-word would have the option of moving to the matrix C^0 covertly and thus avoiding the effects of the island.

The Malay partial movement facts are another case in which it is useful to assume that declarative and interrogative complementizers are syntactically identical, differing only in the interpretation of their wh-features (this idea was suggested above in section 2.3, and will be useful again below in section 4.3). Consider an example of partial movement like (110b) above, repeated here as (114):

Malay (Saddy 1991, 189)

(114) Bill tahu siapa yang Tom cintai t?

Bill knows who FOC Tom loves

‘Who does Bill know that Tom loves?’

The claim pursued here is that examples like (114) involve a strong wh-feature in the embedded CP and a weak feature in the matrix CP. If this claim is correct, then it must be possible for declarative complementizers to have strong wh-features.

4 An embarrassment of riches: movement to two strong features

Finally, there is one case which Procrastinate would allow which the theory developed here rules out, namely a chain containing two instances of overt movement driven by movement to a strong feature:

(115) [strong] [strong] X


The chain in (115) has two positions which PF is instructed to pronounce. Such a chain is ruled out by the PF filter under consideration in this chapter; PF fails to receive complete instructions about which element of this chain to pronounce, as it has no way of choosing between the two highest positions in the chain.

Extraction of subjects in English is standardly taken to involve a chain of the type in (115):

(116) Who_i did John say t_i t'_i saw Mary?

who in (116) is standardly assumed to originate in an internal subject position, to move (string-vacuously, in this case) overtly to an external subject position, and then to undergo overt wh-movement into the matrix clause. Both of these moves are driven by strong features. In fact, this case would appear to arise quite frequently cross-linguistically; it is not clear that there are any languages in which movement of the subject to the external position is always covert¹². Overt wh-extraction of subjects would therefore seem to be problematic for this theory.

In fact, it seems to be the case that overt extraction is cross-linguistically problematic, in ways which can be accounted for in terms of the theory developed here. In the following sections I will try to show that languages are indeed forced to avoid chains like the one in (115).

4.1 “Weakening” the subject position

Suppose we consider a language in which overt movement of the subject to the external subject position is optional. This theory predicts that in cases of overt subject extraction, the weak feature for attracting the subject to the external position will have to be used.

There is a long literature arguing that in languages in which subjects can be either preverbal or postverbal, wh-extraction of the subject must take place from postverbal position (cf. Rizzi 1982, Jaeggli 1984, Brandi and Cordin 1989, Campos 1997, and

¹² See Bobaljik and Carnie 1996, in particular, for arguments that Irish VSO order results not from covert raising of the subject but from overt raising of the verb to a high position in the clause.

references cited there). One of the clearest pieces of evidence for this comes from certain northern Italian dialects in which preverbal and postverbal subjects are distinguishable by the agreement they trigger on the verb:

Fiorentino (adapted from Brandi and Cordin 1989, 121-122)

- (117) a. La Maria l' è venuta
 the Maria she is come-FEM
 'Maria came'
- b. Gli è venuto la Maria
 it is come the Maria
 'Maria came'

Trentino (adapted from Brandi and Cordin 1989, 121-122)

- (118) a. La Maria l' è vegnuda
 the Maria she is come-FEM
 'Maria came'
- b. È vegnú la Maria
 is come the Maria

Wh-extraction of the subject forces use of the default third person singular verb form, rather than the agreeing verb form associated with overt raising of the subject :

Fiorentino (Brandi and Cordin 1989, 124-125)

- (119) a. Quante ragazze gli è venuto con te?
 how-many girls it is come with you
 'How many girls came with you?'
- b.* Quante ragazze le sono venute con te?
 how-many girls 3.PL.FEM. are come-FEM.PL. with you
 'How many girls came with you?'

Trentino (Brandi and Cordin 1989, 124-125)

(120) a. *Quante putele è vegnú con ti?*

how-many girls is come with you

‘How many girls came with you?’

b. **Quante putele le è vegnude con ti?*

how-many girls 3.PL.FEM. is come-FEM.PL. with you

‘How many girls came with you?’

In the theory developed here these facts are not surprising. Fiorentino and Trentino appear to have both a strong and a weak subject-attracting feature, which are morphologically distinguished. Overt *wh*-extraction of the subject forces use of the weak feature, thus creating a chain with only a single strong feature associated with it.

The use of impoverished or absent subject agreement morphology with overtly extracted subjects is common cross-linguistically, and is sometimes referred to as the “anti-agreement effect” (Ouhalla 1993, Phillips to appear). Other more or less clear cases of languages exhibiting anti-agreement effects include Berber (Ouhalla 1993), Chamorro (Chung 1982, 1983, 1994), Halkomelem (Gerds 1980, 1988), Jacalteco (Craig 1977, 1979), K’ichee’ (Mondloch 1981, Davies and Sam-Colop 1990, Larsen 1987, Trechsel 1993, Hale and Storto 1996), Kinande (Schneider-Zioga 1995, 1996), Palauan (Georgopoulos 1985, 1991), Turkish (Underhill 1972) and Yimas (Phillips 1996), among others^{13, 14}:

¹³ Breton is also often listed among languages with anti-agreement; in terms of the theory developed here, Breton agreement is like Irish agreement, and thus expresses a weak rather than a strong feature. See section 1.6.3 of this chapter for further discussion of this type of agreement and its interaction with *wh*-movement.

¹⁴ In some of the examples below I have constructed an agreeing form of the appropriate verb for the (b) examples which is not present in the texts from which the (a) examples are drawn, and I may well have made morphophonological errors in so doing. The point being made here is unaffected by this; all of the languages cited below are described by the authors of the above papers as exhibiting anti-agreement effects.

Berber (Ouhalla 1993, 479)

- (121) a. man tamghart ay yzrin Mohand?

which woman COMP see-PART Mohand

‘Which woman saw Mohand?’

- b. *man tamghart ay t- zra Mohand?

which woman COMP 3FEM.SG.-saw Mohand

Chamorro (Chung 1982, 50)

- (122) a. Hayi fuma’gasi i kareta?

who washed the car

‘Who washed the car?’

- b. *Hayi ha-fuma’gasi i kareta?

who 3sg washed the car

‘Who washed the car?’

Halkomelem (Gerds 1980, 303)

- (123) a. sténi? θə ni q’wələt tθə səplíl

lady DET ASP bake-TRANS DET bread

‘A lady is the one who baked the bread’

- b. *sténi? θə ni q’wələtəs tθə səplíl

lady DET ASP bake-TRANS-3.ERG DET bread

‘A lady is the one who baked the bread’

Jacaltec (Craig 1979, 150)¹⁵

(124) a. mac xc -ach 7il -ni?

who ASP-2.ABS see-AC

‘Who saw you?’

b. *mac xc -ach y -7il -ni?

who ASP-2.ABS 3.ERG-see-AC

‘Who saw you?’

K’ichee’ (Hale and Storto 1996, 3)

(125) a. Laa aree lee achi x- at- kuna-n (at)?

Q FOC the man ASP-2.SG.ABS-cure-AC (you-SG)

‘Was it the man who cured you?’

b. *Laa aree lee achi x- at- uu- kunaj (at)?

Q FOC the man ASP-2.SG.ABS.-3.SG.ERG-cure (you-SG)

Kikuyu (Clements 1984, 39)

(126) a. nó.o o- tem -íré mote?

who-CL1 WH.AGR cut TNS tree

‘Who cut a tree?’

b. *nó.o á- !tém -íré motě!?

who-CL1 CL1-cut TNS tree

‘Who cut a tree?’

¹⁵ In these examples “AC” stands for “Agentive Construction”--this suffix must appear whenever the subject of a transitive clause is extracted.

Kinande (Schneider-Zioga 1995, 75)

- (127) a. *IyOndI y' U- ka- langİra Marya?*

who-CL1 COMP-CL1 WH.AGR PRES see Marya

'Who sees Mary?'

- b. **IyOndI y' a- ka- langİra Marya?*

who-CL1 COMP-CL1 CL1 PRES see Marya

'Who sees Mary?'

Palauan (Georgopoulos 1991, 88)

- (128) a. *ng- te'a a kileld -ii a sub?*

3SG.SUBJ who R-PERF-heat 3SG.OBJ soup

'Who heated up the soup?'

- b. **ng- te'a a ng- kileld -ii a sub?*

3SG.SUBJ who 3SG.SUBJ R-PERF-heat 3SG.OBJ soup

'Who heated up the soup?'

Turkish (Ouhalla 1993, 484)

- (129) a. *hoca -yi gör -en öğrenciler*

lecturer ACC see PART students

'the students who saw the lecturer'

- b. **hoca -yi gör -en -ler öğrenciler*

lecturer ACC see PART PL students

'the students who saw the lecturer'

Yimas (Phillips 1995, 351)

(130) a. nawm m- Ø- kul- cpul-um?

who-PL COMP WH.ABS 2PL.ACC hit PL

‘Who hit you all?’

b. *nawm pu- kul- cpul?

who-PL 3PL.ABS 2PL.ACC hit

‘Who hit you all?’

Thus, in a number of languages it would appear that extraction of the subject forces the use of impoverished or absent subject agreement morphology. Furthermore, at least in some languages, this morphology is apparently associated with failure to raise the subject to the external subject position. The theory developed here gives us a way of interpreting these facts. We have seen that a chain associated with two strong features is predicted on this theory to be ill-formed, because PF will be unable to decide which of the two positions in the chain to pronounce. One way of improving a structure with respect to this constraint, then, would be to substitute a weak feature for one of the offending strong features. The data above suggest that some languages do make use of this option, substituting a weak subject-agreement feature for a strong subject-agreement feature¹⁶.

¹⁶ Ouhalla (1993) notes that Standard Arabic is an apparent counterexample to the generalization that languages with the option of using weak agreement and a postverbal subject must do so when the subject is wh-extracted. Arabic does have postverbal subjects with weak agreement:

Standard Arabic (Ouhalla 1993, 487, 514)

(i) wasal -a al- tullab -u
arrived 3.M.SG. the students NOM
‘the students arrived’

(ii) al- tullab -u wasal -uu
the students NOM arrived PL

However, wh-extraction of the subject forces the use of strong agreement:

Standard Arabic (Ouhalla 1993, 488)

(iii) ?ayy -u tullabin wasal -uu / *a
which-NOM students arrived PL 3.M.SG
‘Which students have arrived?’

Ouhalla’s suggestion is that the relevant difference between Arabic and the anti-agreement cases discussed in the text has to do with the fact that wh-words can optionally be left in situ in Arabic, an option which is unavailable in the other languages; (iii), on this account, would be a case of wh-in-situ, and the interesting question would be why wh-in-situ must be preverbal rather than postverbal. If Ouhalla is right, wh-extraction of the subject in Arabic would involve a chain in which both of two landing sites could in principle be associated with either a weak or a strong feature (in this case, Spec CP and the specifier responsible for strong verbal agreement). (iii) shows that given a chain of this type, the preference is for

Note that the distribution of overt agreement in these cases differs from that of French participle agreement (see section 1.1). French participle agreement is always associated with a weak feature, but is only morphologically realized when the relevant specifier is occupied at some point in the overt syntax. The presence of agreement in anti-agreement languages, on the other hand, signals the presence of strong agreement features as opposed to weak ones. Since strength and being checked in the overt syntax are often linked, it is perhaps not surprising that these syntactic phenomena receive similar morphological expressions.

4.1.1 Anti-anti-agreement

One interesting property of the anti-agreement effects discussed in the previous section is that they sometimes fail to appear. In the following section I will discuss some of these cases of “anti-anti-agreement”. We will see that the use of weak features for subject agreement is sometimes ruled out by other components of the grammar.

4.1.1.1 Morphological anti-anti-agreement

Building on an observation of Ouhalla (1993), Phillips (to appear) argues convincingly that anti-agreement can sometimes be impossible for morphological reasons. Some but not all languages fail to show anti-agreement effects in negative questions:

Berber (Ouhalla 1993, 499)

(131) man tamghart ay ur t- ssn Mohand?

which woman COMP NEG 3FEM.SG. know Mohand

‘Which woman does not know Mohand?’

the lower of the two features to be the one realized as a strong feature. See footnote 10 above for another case in which this preference seems to be in effect.

Turkish (Ouhalla 1993, 500)

(132) a. hoca -yi gör -me -yen öğrenciler

lecturer ACC see NEG PART students

‘the students who did not see the lecturer’

b. *hoca -yi gör -me -yen -ler öğrenciler

lecturer ACC see NEG PART PL students

‘the students who did not see the lecturer’

Ouhalla (1993) notes that the languages in which anti-agreement effects vanish in negative questions are those, like Berber, in which negation appears further from the verb stem than the agreement morpheme in question; in languages like Turkish, on the other hand, where agreement is further from the verb stem than negation, negation has no effect on anti-agreement. The generalization, then, appears to be that subject agreement can only be eliminated if it is on the “edge” of the word¹⁷.

Phillips (1996, to appear) notes a particularly interesting incarnation of this phenomenon in Yimas. Yimas subject agreement prefixes attach to the left of third person object agreement prefixes, but to the right of first and second person object agreement prefixes¹⁸:

Yimas (Phillips (to appear, 17))

(133) a. pu- nan- tay

3PL.ABS 2SG.ACC see (subject prefix precedes object prefix)

‘They saw you’

b. pu- n- tay

3PL.ABS. 3SG.ERG. see (subject prefix follows object prefix)

‘He saw them’

¹⁷ There are subtleties involving optionality here which I will not attempt to discuss; for instance, anti-agreement with extraction of subjects from negative sentences is optional in Berber, but impossible in Breton (and obligatory in Turkish).

¹⁸ This by no means does justice to the very complex agreement system of Yimas; see Foley 1991, Phillips 1993, 1996, to appear for further discussion.

Anti-agreement appears in Yimas when the object is first or second person (and hence the subject agreement morpheme is on the “edge” of the word), but not when the object is third person:

Yimas (Phillips (to appear, 17))

(134) a. *nawm m- kul- cpul-um?*

who-PL COMP 2PL.ACC hit PL

‘Who hit you all?’

b. *nawrm na- mpi- tpul?*

who-DU 3SG.ABS 3DU.ERG hit

‘Which two people hit him?’

Phillips’ proposal for this morphologically conditioned anti-anti-agreement, which I will adopt, is that anti-agreement represents failure of the verb to raise to the functional head responsible for subject agreement. In cases in which the verb must raise past this position for independent reasons, then, as when a morpheme above the subject agreement morpheme needs morphological support, anti-agreement cannot appear.

One potential problem for Phillips’ proposal is anti-agreement in K’ichee’, where the subject agreement morpheme which is eliminated by subject extraction appears to be preceded by a number of other inflectional prefixes:

K’ichee’ (Hale and Storto 1996, 3)

(135) a. *Laa aree lee achi x- at- Ø- kuna-n (at)?*

Q FOC the man ASP 2.SG.ABS ERG cure AC (you-SG)

‘Was it the man who cured you?’

Here the null subject agreement morpheme is preceded by overt object agreement and aspect morphology. A possible solution to this problem might be to question the assumed distribution of word boundaries in the K’ichee’ verb. Craig (1977, 1979) posits a word boundary between the absolutive agreement morpheme and the ergative agreement morpheme in the related language Jacalteco:

Jacaltec (Craig 1977, 111)

(136) xc -ach w- abe

ASP 2.ABS 1.ERG hear

‘I heard you’

If a word boundary can be posited here for K’ichee’ as well, Phillips’ proposal faces no problems with that language.

We have seen, then, that anti-agreement is sometimes impossible for what would appear to be purely morphological reasons. We are left with two potential ways of thinking about anti-anti-agreement. One would be to say that substitution of a weak subject-attracting feature for a strong one is impossible in these cases for morphological reasons, and the feature attracting the subject is strong. Another possibility would be to say that a weak subject-attracting feature is always used in subject extraction (in these languages, or possibly in all languages), but that for morphological reasons this weak feature is sometimes spelled out morphologically with the same morpheme as that used for the strong subject-attracting feature. In the next sections I will discuss some facts which seem to argue in favor of the first proposal; anti-agreement does reflect a syntactically distinct subject agreement feature, and when it does not occur that agreement feature is not present in the syntax.

4.1.1.2 Anti-anti-agreement and anaphora

In some of the languages exhibiting anti-agreement, anti-agreement fails to appear when the subject binds an anaphor or a pronominal variable:

Jacaltec (Craig 1977, 217-218)¹⁹

- (137) a. mac x- Ø- s -potx' s- ba?

who ASP 3.ABS 3.ERG kill 3.ERG self

'Who killed himself?'

- b. mac x- Ø- s- mak s- mam

who ASP 3.ABS 3.ERG kill 3.ERG self

'Who_j hit his_j father?'

Chamorro (Nakamura to appear)²⁰

- (138) a. ?Hayi **ha-** li'i' gui'?

who 3.SG.ERG see him

'Who saw himself?'

- b. Hayi **ha-** taitai i lepblo -ña?

who 3.SG.ERG. read the book 3.SG.POSS

'Who_j read his_j book?'

These facts suggest that anti-agreement is not simply a morphological phenomenon; the morphological impoverishment of the subject's agreement features has consequences for the syntactic properties of the subject. See Richards 1995b for an attempt at developing a binding theory in which the subject's agreement features play an important role.

Constructing a full theory of the interaction between anti-agreement and anaphora is beyond the scope of this thesis, but if the theory developed here is on the right track, these facts should tell us something about the nature of weak subject-attracting inflectional features.

¹⁹There is a (possibly important) contrast between (126a) and (126b), which is that although anti-agreement is impossible in cases of anaphor binding, it is merely optional for pronominal variable binding.

²⁰Note that in this regard Chamorro subject extraction differs from its very similar Tagalog counterpart (see section 1.5); Tagalog verbs always bear actor-topic morphology when the subject is extracted, regardless of whether the subject binds an anaphor:

Tagalog

- (i) Sino ang kumagat sa kanyang sarili?

who T AT-bit Loc his self

'Who bit himself?'

Presumably this has to do with the fact that Tagalog verbs never show agreement with their subjects; Tagalog thus fails to show an "anti-agreement effect". Another difference between Tagalog and Chamorro, which on this account is irrelevant to this phenomenon, is that Tagalog distinguishes morphologically between anaphors and pronouns, while Chamorro uses the same forms for both.

One possibility is that the difference between weakness and strength is not just a diacritic, but has something to do with the content of the features involved; weak features might actually involve fewer morphosyntactic features than strong features.

4.1.1.3 Person-animacy hierarchies and anti-agreement

Some of the languages which exhibit anti-agreement also show effects of a person-animacy hierarchy:

Chamorro (Chung 1994a, 20)

(139) *Kao ha- toktuk hao i chi'lu -hu lahi?

Q 3.SG.ERG hug you the sibling 1.SG.POSS male

'Did my brother hug you?'

Halkomelem (Gerdtz 1988, 31)

(140) *ni ləm -əθamə -s ɬə sténi?

AUX look.at TRANS+2.ABS 3.ERG DET woman

'The woman looked at you'

(139) and (140) are both ruled out by person-animacy hierarchies which forbid sentences with third person subjects and second person objects. Interestingly, anti-agreement rescues sentences of this type in both languages:

Chamorro (Chung 1994a, 177)

(141) Hayi um-ayuda hao?

who AT-help you

'Who helped you?'

Halkomelem (Gerdtz 1988, 32)

(142) niɬ ɬə sténi? ni ləm -əθamə

3.EMPH DET woman AUX look.at TRANS+2.ABS

'It's the woman who looked at you'

In Chamorro, anti-agreement with subject extraction only occurs in realis transitive clauses; in the irrealis, which exhibits a nominative-accusative agreement system, no anti-agreement is found, and the person-animacy hierarchy effects reappear:

Chamorro (Chung 1994a, 177)

- (143) *Hayi pära u- ayuda hao?
 who FUT IRR.3.SG help you
 ‘Who is going to help you?’

Evaluating the significance of these person-animacy hierarchy facts is made difficult in part by the uncertain status of such hierarchies in the theory. If the effects of such hierarchies were purely morphological--if, for instance, they followed purely from cooccurrence restrictions on (possibly null) agreement morphemes--then these interactions between anti-agreement effects and person-animacy hierarchies would not be entirely unexpected. However, Chung (1989, 1994a) discusses facts from Chamorro which suggest that the person-animacy hierarchy in that language is sensitive to the syntactic properties of the subject.

Chamorro clauses with nonpronoun subjects cannot have a third person animate pronoun direct object, but this requirement is suspended when the subject binds an anaphor:

Chamorro (Chung 1994a, 21, 23)

- (144) *Ti pära u- dispensa gui’ i ma’estra
 not FUT IRR.3.SG forgive him/her the teacher
 ‘The teacher is not going to forgive him/her’

- (145) Ha- bira gui’ si Santa Maria...
 3SG.ERG turn her Unm Santa Maria
 ‘The Virgin Mary would turn (herself)...’

Similarly, Chamorro transitive clauses cannot have third person plural nonpronouns as their subject, but this effect vanishes if the subject binds a pronominal variable:

Chamorro (Chung 1994a, 17, 23)

(146) *Pära uma- fa'gasi i lalahi i nā'yan

FUT IRR.3.PL. wash the men the dishes

'The men are going to wash the dishes'

(147) Ma- fa'gasi i lalahi i kannai -ñiha

3PL.ERG wash the men the hand 3PL.POSS

'The men washed their hands'

Thus, the Chamorro person-animacy hierarchy interacts with binding in a way reminiscent of the interaction of anti-agreement and binding. As in the case of anti-agreement, the fact that the person-animacy hierarchy is sensitive to facts about binding suggests that the hierarchy has a syntactic dimension; it is not simply a matter of morphology.

The Chamorro anti-agreement facts, taken together, argue for a particular answer to a question posed above in section 4.1.1.1. In this chapter I have been developing a theory according to which anti-agreement effects signal the substitution of a weak subject-agreement feature for a strong subject-agreement feature; this is a strategy for licensing overt wh-movement, creating a chain with only a single strong feature, which PF is then able to pronounce. Consider irrealis verbs in Chamorro, which fail to exhibit anti-agreement effects, apparently for morphological reasons. There are at least two possible approaches to morphologically conditioned anti-agreement of this kind. One would be to say that the morphology is a reliable guide to the strength of the subject agreement feature; when no anti-agreement is exhibited, the subject agreement feature is in fact strong, and some other method for licensing overt extraction of the subject is involved. Another approach would be to say that these verbs do in fact exhibit anti-agreement in the syntactic sense; the subject feature is always weak in cases of subject extraction, and the only difference between irrealis and realis verbs in Chamorro is that the latter overtly signal this difference in the strength of the subject agreement feature by means of a distinct subject

agreement morpheme, while the former realize strong and weak subject agreement features with identical morphology.

The interaction of anti-agreement with the person-animacy hierarchy in Chamorro suggests that the first approach is the right one. We have seen in this section and the previous one that Chamorro subjects in clauses involving anti-agreement have certain special syntactic properties, including inability to bind anaphors and bound variables and immunity to the person-animacy hierarchy. The ill-formedness of (143) shows that when the verb is morphologically unable to exhibit anti-agreement, immunity to the person-animacy hierarchy no longer holds, even in cases of subject extraction. This argues against an approach in which these verbs are only morphologically distinct from verbs which can exhibit anti-agreement. In other words, it looks as though the morphology is in fact a reliable guide to the strength of the subject agreement features; when anti-agreement is not morphologically expressed, it is in fact syntactically absent.

4.1.2 Anti-agreement and ergativity

Several of the languages listed above as exhibiting anti-agreement are ergative languages. The phenomenon of anti-agreement seems to be related in many ways to the antipassive (and in fact, the anti-agreement constructions in the Mayan languages discussed above are sometimes referred to as “agentive antipassive” in the Mayanist literature; cf. Smith-Stark 1978. For Minimalist accounts of the Mayan facts, see Tada 1993 and Watanabe 1996; these accounts are quite similar in a number of ways to the one developed here). To the extent that these structures share syntactic properties, our conclusions about the syntax of the one will have repercussions for our theories about the other.

In the case of the antipassive, a transitive subject exchanges its ergative agreement morphology for absolutive agreement, which is commonly null; in anti-agreement, ergative agreement morphology is lost. In all of the ergative languages that I know of that exhibit anti-agreement, anti-agreement is limited to transitive subjects, or at least cannot be demonstrated to apply to intransitive subjects. For instance, the Jacaltec suffix *-n(i)* that

obligatorily cooccurs with loss of the ergative agreement morpheme in subject extraction constructions cannot appear in cases of extraction of an intransitive subject:

Jacaltec (Craig 1977, 196-7)

- (148) a. x- Ø- s- watx'e naj hun -ti'

ASP 3.ABS 3.ERG make he one-this

'He made this'

- b. x- Ø- w- il naj x- Ø- watx'e-**n** hun-ti'

ASP 3.ABS 1.ERG see him ASP 3.ABS make AC one this

'I saw the man who made this'

- (149) a. x- Ø- to naj ewi

ASP 3.ABS go he yesterday

'He went yesterday'

- b. x- Ø- w- il naj x- to(*-**n**) ewi

ASP 3.ABS 1.ERG see him ASP go AC yesterday

'I saw the man who went yesterday'

Because the third person absolutive agreement morpheme is null, it is impossible to tell whether agreement with intransitive subjects is deleted, but the failure of the suffix *-n(i)* to appear on the lower verb in (149b) suggests that subject agreement "weakening" cannot take place in this case.

Similarly, in Chamorro, anti-agreement is found only with extraction of transitive subjects; an intransitive verb must still agree in number with an extracted subject:

Chamorro (Cooreman 1987, 42)

- (150) a. *I famagu'on mu-nango gi tasi

the children AT-swim LOC sea

- b. I famagu'on man-nango gi tasi

the children PL swim LOC sea

Thus, anti-agreement seems limited in ergative languages to transitive subjects; in terms of the theory developed here, this means that only transitive subjects are associated with a strong agreement feature. Antipassives, of course, are also limited to transitive clauses.

A third similarity between anti-agreement and antipassives is that there are languages which use the same morphology to signal both. In the Mayan language Ixil, for instance, the suffix *-on* marks both antipassives and verbs undergoing anti-agreement:

Ixil (Lengyel 1978, 87)

(151) a. Kat un- b'an -Ø -ah

ASP 1.SG.ERG do 3.SG.ABS.

'I did it'

b. Kat b'an-**on** -ʔin

ASP do AP 1.SG.ABS

'I did (something)'

(152) a. Kat i- q'os -Ø -ah

ASP 3.SG.ERG hit 3.SG.ABS

'He hit it/him/her'

b. ʔab'il kat q'os-**on**-Ø

who ASP hit AC

'Who hit it/him/her?'

Finally, antipassives, like anti-agreement, have been observed to be obligatory for extraction of transitive subjects in ergative languages that possess antipassives (see Nakamura 1996, in particular, for discussion of this)

Dyirbal (Dixon 1972)

- (153) Payi yara [pakal-nga -nyu pakul yuriku]
 there-ABS man-ABS spear AP NONFUT-ABS there-DAT kangaroo-DAT
 ‘the man who speared the kangaroo’

Chukchi (Comrie 1979, 229)

- (154) [en-agtat -kə -lʔ -a qaa -k] ʔaaček -a
 AP chase NEG PART ERG reindeer LOC youth ERG
 ‘the youth who does not chase the reindeer’

Inuit (Bittner 1994, 58)

- (155) angut [aalaam-mik tigu -si-sima -su -q]
 man gun INST take AP PERF REL.TRANS ABS.SG
 ‘the man who took the gun’

On the other hand, there is reason to believe that the anti-agreement cases in ergative languages cited in the previous sections represent a phenomenon distinct from that classically referred to as antipassive, if that term is understood to involve constructions in which the direct object loses its syntactic status as a direct object. The Halkomelem anti-agreement construction, for instance, has a transitive suffix on the verb, and the object appears to retain its status as an argument:

Halkomelem (Gerds 1980, 303)

- (156) sténiʔ ʈə ni q'wólət tʈə səplíl
 lady DET ASP bake-TRANS DET bread
 ‘A lady is the one who baked the bread’

Halkomelem does have an antipassive, which substitutes an antipassive suffix for the transitive suffix and marks the direct object as an oblique:

Halkomelem (Gerds 1980, 300)

- (157) ni q'wələm θə sténi? ʔə tθə səplíl
 ASP bake-AP DET lady OBL DET bread
 'The lady baked the bread'

Similarly, in Jacaltec and Ixil, the direct object appears to be the absolutive argument in anti-agreement constructions, since it controls absolutive agreement on the verb:

Jacaltec (Craig 1979, 150)

- (158) mac xc -ach 7il -ni?
 who ASP 2.ABS see AC
 'Who saw you?'

Ixil (Lengyel 1978, 87)

- (159) ʔeš kat q'os-on -ʔin
 you-PL.ABS ASP hit AC 1.SG.ABS
 'You are the ones that hit me'

Like Halkomelem, Jacaltec has a morphologically distinct antipassive, in which the verb takes a special antipassive suffix and the object is expressed as a postpositional phrase:

Jacaltec (Craig 1979, 140)

- (160) x- Ø- col -wa naj t- aw -iñ
 ASP 3.ABS help AP he AUG 2.ERG to
 'He helped you'

Antipassives and anti-agreement, then, must apparently be regarded as related but distinct constructions. If the approach taken here is right, both constructions involve substitution of a weak subject-attracting feature for a strong one. The two constructions appear to differ in their treatment of the object; anti-agreement leaves the object unaffected, while antipassives alter the syntactic representation of the object. Moreover, most transitive sentences can be antipassivized, while anti-agreement is often (though not always; cf. the Fiorentino and Trentino facts above) restricted to cases of subject extraction.

It is tempting to regard these facts in terms of the constraint on “overcrowding” alluded to in section 1.7 above. We have seen a number of phenomena suggesting that leaving multiple nominal arguments in situ in the VP is ruled out. If the theory developed here is on the right track, antipassives and anti-agreement have in common that the subject is left in situ in the VP, along with the object. We predict that this configuration can only be licensed if one of the arguments is extracted or made into an adjunct. Anti-agreement would represent the first of these strategies, and antipassive the second.

This account leaves much to be explained, of course, which I cannot at the moment explain. One question is why anti-agreement can only be licensed by subject extraction, and antipassive by making the object into an adjunct; why is the reverse not possible? There are also empirical problems with the approach. As just noted, there are languages in which anti-agreement is possible even if the subject is not *wh*-extracted (namely, Fiorentino and Trentino); this is apparently a case in which neither of the strategies for licensing multiple in-situ arguments is used. The opposite case is also attested, in which both strategies are used; K’ekchi has a construction which is only usable for subject extraction (thus patterning with anti-agreement) which involves expressing the direct object as an adjunct (thus patterning with antipassive):

K’ekchi (Berinstein 1985, 164)

(161) Ani x- Ø- sac’-o -c re?

who TNS 3.ABS hit AC ASP ERG.3-DAT

‘Who hit him?’

Interestingly, the K’ekchi subject-extraction construction is not required (or allowed) for extraction of subjects which bind reflexives, a fact which is reminiscent of the anti-anti-agreement facts discussed in 3.1.1.2:

K'ekchi (Berinstein 1985, 108)

- (162) Ani x- Ø- x- yoc' rib?
 who TNS ABS.3 ERG.3 cut ERG.3-self
 'Who cut himself?'

Thus, many questions remain. Still, it seems reasonable to think that the phenomenon of anti-agreement and its similarity to antipassive may shed some light on the syntax of ergativity; if the account developed here is correct, ergative languages always move the subject out of the VP in the overt syntax only in transitive clauses, and antipassives involve failure to overtly move the subject.

4.1.3 Interlude: Chamorro *-um-*

The above discussion of anti-agreement offers a way of simplifying somewhat our assumptions about Chamorro verbal morphology. Chamorro is traditionally (cf. Topping 1973, Chung 1982, 1983, 1989, 1994a, 1994b, Cooreman 1987) described as having several distinct infixes *-um-*. One *-um-* is the one which appears when the subject of a transitive realis clause is extracted; this was glossed in section 1.5 above as an "Actor-Topic" morpheme, signalling movement of the subject through the "topic" position:

Chamorro (adapted from Chung 1982, 49-50)

- (163) Hayi fuma'gasi i kareta?
 who AT-wash the car
 'Who washed the car?'

A second *-um-* is the one which appears in control clauses:

Chamorro (Chung 1982, 49)

- (164) a. Malägu' gui' bumisita si Rita
 want he AT-visit Unm Rita
 'He wants to visit Rita'

In section 1.5 I suggested that these two instances of *-um-* could be unified. Movement to the topic position is driven by a weak feature. As a result, such movement can only occur

in the overt syntax if the moving element moves further to check a strong feature (thus satisfying (3) by giving PF a position associated with a strong feature to pronounce) or if the moving element is null (and therefore immune to PF requirements like (3)).

A third *-um-* in Chamorro is an agreement morpheme. Chamorro has a split-ergative agreement system; realis verbs have an ergative agreement pattern, and irrealis verbs have a nominative-accusative pattern. The subjects of transitive realis verbs control a rich agreement system which distinguishes all the persons and numbers (Chung 1994a, 12):

(165)	<u>singular</u>		<u>dual/plural</u>
1st	<i>hu-</i>	incl.	<i>ta-</i>
		excl.	<i>in-</i>
2nd	<i>un-</i>		<i>in-</i>
3rd	<i>ha-</i>		<i>ma-</i>

Intransitive verbs, by contrast, exhibit only a number distinction (Chung 1994a, 12):

(166)	<u>singular/dual</u>	<u>plural</u>
	<i>-um-</i>	<i>man-</i>

Thus, the third *-um-* is an agreement marker. It is easy to see why this *-um-* is distinguished from the other cases of *-um-* in the literature on Chamorro. The *-um-* which signals subject extraction appears only when transitive subjects are extracted; that is, it only replaces the paradigm in (165), of which it is not a member, and not the one in (166).

Plural intransitive subjects still force use of the plural prefix *man-* when extracted²¹:

Chamorro (Cooreman 1987, 42)

- (167) a. *I famagu'on mu-nango gi tasi
 the children AT-swim LOC sea
 b. I famagu'on man-nango gi tasi
 the children PL swim LOC sea

²¹Here the infix *-um-* has undergone phonologically conditioned metathesis, and thus surfaces as *mu-*.

Moreover, subject extraction does not simply detransitivize the verb for purposes of agreement; the intransitive plural prefix *man-* never occurs when transitive plural subjects are extracted (Sandra Chung, p.c.).

Both of these differences between the *-um-* which marks subject extraction and the agreement marker *-um-* are expected on this theory, however. Suppose that the use of *-um-* with subject extraction really is a case of anti-agreement; extraction of a subject forces the use of a weak agreement feature, which is spelled out as a default agreement morpheme, the least specific one in the realis paradigm. The situation in Chamorro would then be very similar to that in languages like Fiorentino and Trentino, in which anti-agreement appears as the default third person singular masculine form:

Fiorentino (Brandi and Cordin 1989, 124-125)

(168) a. Quante ragazze gli è venuto con te?

how-many girls it is come with you

'How many girls came with you?'

b.* Quante ragazze le sono venute con te?

how-many girls 3.PL.FEM. are come-FEM.PL. with you

'How many girls came with you?'

Trentino (Brandi and Cordin 1989, 124-125)

(169) a. Quante putele è vegnú con ti?

how-many girls is come with you

'How many girls came with you?'

b. *Quante putele le è vegnude con ti?

how-many girls 3.PL.FEM. is come-FEM.PL. with you

'How many girls came with you?'

Thus, we would not expect the plural agreement marker *man-* to be used in anti-agreement; anti-agreement is not a matter of switching from the transitive agreement paradigm to the intransitive agreement paradigm, but involves use of the most unmarked subject agreement

morpheme, namely *-um-*. The fact that anti-agreement only appears with transitive verbs is also explained under this theory. We have seen that in ergative languages, anti-agreement is often restricted to transitives:

Jacaltec (Craig 1977, 196-7)

(170) a. x- Ø- w- il naj x- Ø- watx'e-n hun-ti'

ASP 3.ABS 1.ERG see him ASP 3.ABS make AC one this

'I saw the man who made this'

b. x- Ø- w- il naj x- to(*-n) ewi

ASP 3.ABS 1.ERG see him ASP go AC yesterday

'I saw the man who went yesterday'

In *Jacaltec*, the suffix *-n(i)* which occurs in anti-agreement constructions cannot occur on an intransitive verb. I suggested in section 4.1.2 of this chapter that only the transitive subject in ergative languages is associated with strong agreement features, and therefore it is only in cases of transitive subject extraction that the agreement features of the subject must be made weak to permit extraction. Thus, the failure of Chamorro anti-agreement to affect extraction from intransitive verbs is not unexpected. It would appear, then, that we can reduce the instances of *-um-* in Chamorro to one, a default non-plural agreement marker which is syntactically weak. The other "topic markers" discussed in section 1.5 can still be understood as they were in that section, as indicators of movement to a topic position, but the topic marker for subject extraction is apparently null in Chamorro. Note that the subject does still control agreement when the object is extracted, although it is of a different morphological form than ordinary subject agreement (a fact for which I have no account):

Chamorro (adapted from Chung 1982, 49-50)

(171) Hafa finahan-ña si Maria gi tenda

what GT-buy-3sg Unm Maria Loc store

'What did Maria buy at the store?'

4.1.4 Anti-object-agreement

The discussion thus far has centered on anti-agreement for subjects. I have claimed that subject anti-agreement represents the substitution of a weak subject agreement feature for a strong feature, thus licensing overt extraction of the subject by limiting the number of strong features with which the chain is associated to one. We expect to also find cases of anti-agreement for objects; in languages in which strong and weak object agreement features are morphologically distinct, the weak features should be obligatorily used, modulo the anti-anti-agreement effects discussed above, in cases of object extraction. I know of two clear cases of this phenomenon:

Karitiana (Hale and Storto 1996, 20)²²

- (172) a. Mora-mon y- 'it ti- oky -t
 what ABS my father OT kill NONFUT
 ‘What did my father kill?’
 b.*Mora-mon y- 'it i- oky -t
 what ABS my father 3.OBJ kill NONFUT
 ‘What did my father kill?’

Yimas (Phillips to appear, 30)

- (173) a. Wara ipa- na- am -n?
 what 1PL.ABS. DEF eat PRES
 ‘What are we going to eat?’
 b.*Wara na- kay- am -n?
 what 3SG.ABS. 1PL.NOM. eat PRES
 ‘What are we going to eat?’

4.2 Subject-complementizer interactions

In the previous section we saw fairly extensive evidence that in some languages, overt extraction of an argument necessitates the use of special “weak” agreement with that

²²Here “OT” stands for “Object Topicalization”, a morpheme which obligatorily replaces the object agreement morpheme when the object is extracted.

argument, a phenomenon sometimes referred to as “anti-agreement”. Anti-agreement receives a natural account under the theory developed here, in which overt extraction from a position with strong agreement is ill-formed.

We have also seen evidence that anti-agreement is sometimes unavailable, for morphological or syntactic reasons, and that the absence of anti-agreement morphology is a reliable indicator of the unavailability of the syntactic strategy signalled by anti-agreement; that is, when anti-agreement is not employed, the feature responsible for attracting the subject to the external subject position is in fact strong, and wh-extraction takes place from that position. In this section I will try to provide an account of the method of extraction involved in these cases; the account will owe a great deal to the insights of Rizzi (1990), but is quite tentative.

In a number of languages, subject extraction imposes certain restrictions on the form of the complementizer of the clause immediately containing the subject:

(174) a. Who do you think (*that) left?

b. Who do you think (that) Bill saw?

Norwegian (Taraldsen 1986, 150)

(175) a. Vi vet hvem *(som) snakker med Marit.

we know who that talks with Mary

‘We know who is talking with Mary’

b. Vi vet hvem (*som) Marit snakker med.

we know who that Mary talks with

‘We know who Mary is talking with’

French (Rizzi 1990, 56)

- (176) a. l'homme que je crois qui/*que viendra
 the-man that I think who/that will-come
 'the man that I think will come'
- b. l'homme que je crois que/*qui Jean connaît
 the-man that I think that/who Jean knows
 'the man that I think Jean knows'

Thus, subject extraction can force a complementizer to be absent, as in English, or present, as in Norwegian, or to have a particular form, as in French. Building on an observation of Perlmutter's (1971), Rizzi (1982) points out that languages exhibiting such effects on complementizers in subject extraction are always ones lacking the option of postverbal subjects; in our terms, only languages without anti-agreement put conditions on the form of the complementizer of a clause out of which a subject is extracted. The method for licensing subject extraction in languages in which the feature attracting the subject to the external position is always strong, then, apparently has something to do with the complementizer.

Moreover, there is at least one important difference between anti-agreement and the effects of subject extraction on the complementizer. We have seen evidence that anti-agreement can be rendered unavailable for morphological or syntactic reasons, and that when anti-agreement is impossible it is not used. This does not appear to be true of complementizer-extraction interactions, at least in English. Consider the *that*-trace paradigms in (177)-(178), for example:

- (177) a. John said (that) he would prune the rosebush
 b. Who did John say (*that) t would prune the rosebush?
 c. What did John say (that) he would prune t?
 d. Who is John saying (?that) just yesterday t pruned the rosebush?
- (178) a. John muttered darkly *(that) he would chop down the tree
 b. Who did John utter darkly *(that) t would chop down the tree?
 c. What did John utter darkly *(?that) he would chop down t?
 d. Who is John uttering darkly *(??that) just yesterday t chopped down the tree?

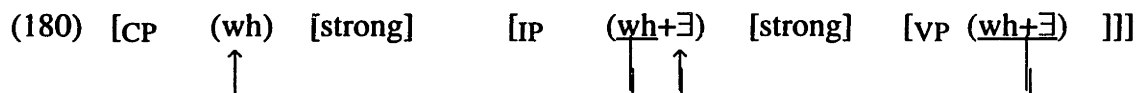
For whatever reason, deletion of the complementizer introducing the clausal complement of *mutter* in this case is impossible, and extraction from this clausal complement is rather degraded. Still, the contrasts which are typical of *that*-trace phenomena seem to hold for (178) as for (177); extraction from the subject is generally ill-formed if a complementizer is present, and can be improved by the presence of an adverb between the complementizer and the extraction site (Culicover 1993). Thus, conditions on the form of the complementizer in cases of subject extraction are enforced even when independent requirements force the use of the “wrong” complementizer. This does not appear to be the case for anti-agreement, as we have seen.

It would appear, then, that in some languages subject extraction involves a chain of the kind in (179), which can only be saved by the use of a particular kind of complementizer; if this complementizer is unavailable for independent reasons, extraction is impossible:



In section 1.9 I suggested, following a large literature, that *wh*-words consist of two parts: a *wh*-component and an existential component. I claimed that *wh*-extraction can either move the entire *wh*-word or simply target the *wh*-component, leaving the existential component behind. Suppose that instead of the chain in (179), subject extraction in the

languages under investigation always involves extraction of the wh-component alone from the external subject position:



The representation in (180), unlike that in (179), involves two distinct chains, one headed by [wh], and the other by [wh+ \exists]. Each chain is associated with a single strong feature, and is therefore a well-formed PF object.

In English, of course, only one of the above chains is pronounced. We might expect to find languages in which both chains are pronounced. *Vata* (Koopman 1983, 1984) and *Yoruba* (Carstens 1985, 1987, Lawal 1987, Sonaiya 1989) are arguably cases of this; these are languages in which wh-extraction from the subject position obligatorily leaves a resumptive pronoun behind:

Vata (Koopman 1984, 37)

(181) a. àlÓ *(Ò) lē saká la'

who he eat rice Q

'Who is eating rice?'

b. yI kòfí lē (*mÍ) la'

what Kofi eat it Q

'What is Kofi eating?'

Yoruba (Carstens 1987, 62)

(182) a. Tani *(ó) n kọrin

who he ASP sing

'Who is singing?'

b. Kíni Ànà kà

what Aina read

'What did Aina read?'

In the discussion of movement of a bare *wh*-component in section 1.9, we noted that such movement has a number of distinctive properties. One of them is strong sensitivity to islands; extraction from a weak island forces movement of an entire *wh*-word, rather than a bare *wh*-component. We thus expect subject extraction to be especially susceptible to weak islands in languages without anti-agreement. This does appear to be the case:

(183) a. *Who are you trying to find out whether *t* wrote *War and Peace*?

b. ?What are you trying to find out whether Tolstoy wrote *t*?

On the other hand, we also noted that extraction of a bare *wh*-element has certain semantic consequences; such *wh*-extraction cannot yield a reading in which the existential component of the *wh*-word has scope over scopal elements between the base position of the *wh*-word and its landing site. Because bare *wh*-elements cannot be extracted from weak islands, this is the reading which vanishes under extraction from weak islands:

(184) How many people do you think I should talk to?

a. For what *n*: there are *n*-many people *x_i* such that you think I should talk to *x_i*.

b. For what *n*: you think that it should be the case that there be *n*-many people that I talk to.

(185) How many people do you wonder whether I should talk to?

a. For what *n*: there are *n*-many people *x_i* such that you wonder whether I should talk to *x_i*.

b. *For what *n*: you wonder whether it should be the case that there be *n*-many people that I talk to.

All other things being equal, then, we would expect subject extraction to lack the ambiguity in (184), yielding only the (b) reading. But this is clearly false:

(186) How many people should I say will talk to me?

a. For what n : there are n -many people x_i such that I should say that x_i will talk to me.

b. For what n : I should say that there are n -many people x_i that will talk to me.

We are thus led to the question of what permits the (a) reading in examples like (186). I would like to claim that the availability of this reading is a result of properties of a certain class of complementizers.

Complementizers in a number of languages agree with the *wh*-words in their specifiers. Kinande is one such language (Schneider-Zioga 1995, 1996):

Kinande (Schneider-Zioga 1995, 71, 75, 82)

(187) a. *EkIhI kyO Mary' akalangIra*

what-CL7 that-CL7 Mary sees

'What does Mary see?'

b. *IyOndI yO Yosefu alangIra*

who-CL1 that-CL1 Joseph sees

'Who did Joseph see?'

c. *IyOndI y' UkalangIra Marya*

who-CL1 that-CL1 sees Mary

'Who sees Mary?'

d. *EkIhI kyO Mary' akaBula nga-kyO Yosefu akalangIra*

what-CL7 that-CL7 Mary wonders if-CL7 Joseph sees

'Who does Mary wonder if Joseph sees?'

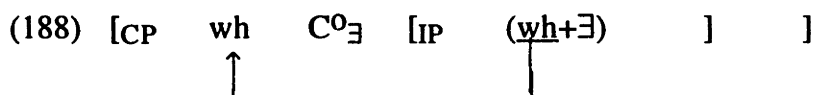
e. *IyOndI yO Mary' akaBula nga-y' UbiIrgEnda*

who-CL1 that-CL1 Mary wonders if-CL1 leave

'Who does Mary wonder if has left?'

In the discussion of French participial agreement, we saw that agreement is associated with the existential component of the *wh*-word. I would like to suggest that the agreement on

complementizers is capable of supplying the semantic force of this existential component. The nominal features on the complementizer, together with the *wh*-component of the *wh*-word, are interpreted as if a *wh*-word with its existential component had moved into Spec CP:



Wh-movement of the subject can therefore be semantically ambiguous as *wh*-movement typically is, albeit via a different mechanism.

This different mechanism gives subject extraction a special status with respect to the constraint on *wh*-movement across weak islands discussed above. Recall that weak islands force pied-piping of the existential component; the reading associated with leaving this component in situ is ruled out:

(189) How many people do you think I should talk to?

- a. For what *n*: there are *n*-many people *x_i* such that you think I should talk to *x_i*.
- b. For what *n*: you think that it should be the case that there be *n*-many people that I talk to.

(190) How many people do you wonder whether I should talk to?

- a. For what *n*: there are *n*-many people *x_i* such that you wonder whether I should talk to *x_i*.
- b. *For what *n*: you wonder whether it should be the case that there be *n*-many people that I talk to.

The nature of this constraint is still fairly mysterious, and I will not try to determine its exact nature here. Consider, however, the status of subject extraction with respect to the constraint in question. Subject extraction always involves movement of a bare *wh*-element, without an accompanying existential component, and in this way might be said to violate the constraint in question, if the constraint applies throughout the derivation. On the other hand, subject extraction results in a representation in which a *wh*-element is paired with an

existential element supplied by an agreeing complementizer. Representationally, then, the semantic constraint illustrated in (189-190) might be said to be satisfied by subject extraction.

This “mixed” status with respect to the constraint on pied-piping of the existential element might give us an account of the “surprising asymmetry” noticed by Pesetsky (1984):

- (191) a. ?Who do you know why I think you like t?
 b. ??Who do you know why I think t likes you?

The asymmetry in (191) is surprising because the link which crosses the weak island created by *why* is the same in both cases; it connects the scope position of the moving wh-word *who* with an intermediate Spec CP position in the most deeply embedded clause. The contrast in grammaticality suggests that extraction of a subject is syntactically distinguished from extraction of an object throughout the derivation, not just in the link from the base position to the next position in the chain. The theory developed here gives us this result; subject extraction involves successive-cyclic movement of a bare wh-element, which violates the constraint on weak island extraction if that constraint is understood derivationally.

On the other hand, subject extraction fails to violate the constraint on extraction from weak islands representationally, which is presumably the reason for the three-way contrast in (192) (Rizzi 1990, 81):

- (192) a. ?*Who do you wonder whether we believe t can help us?
 b. ??Who do you wonder whether we believe we can help t?
 c. *How do you wonder whether we believe we can help Bill t?

The adjunct *how* violates the constraint on extraction from weak islands both derivationally and representationally, as it cannot be understood as involving existential quantification over an individual-level variable.

Finally, we are entitled to wonder why the subject must make use of the semantic assistance of the complementizer at all, especially in cases which do not involve extraction across weak islands. Why can the question in (193a), involving the “wrong” complementizer downstairs, not simply have the representation in (193b)?

(193) a. *Who do you think that *t* left?

b. [CP wh ... [CP wh *that* [IP (wh+ \exists) [VP (wh+ \exists)]]]

Unfortunately, I have no very good answer to this question. The representation in (193) would yield an reading in which the existential component of the *wh*-word was interpreted in Spec IP. One possibility is that such a reading is impossible. It is not clear that the external subject position makes a semantic contribution to the meaning of the sentence. A representation like (193b) might be ruled out, then, by a principle of representational economy, since it forces the semantics to assign an interpretation to Spec IP, a position which is ordinarily left uninterpreted. *Wh*-movement of the object, by contrast, can leave the existential component of the *wh*-word in the object's theta-position, which is clearly interpreted, and thus need not make use of the agreeing complementizer's semantic contribution.

A related question has to do with the contrast in acceptability between (194a) and (194b) (Rizzi 1990, 81)

(194) a. *Who do you wonder whether can help us?

b. ?*Who do you wonder whether we think can help us?

This contrast will be discussed in chapter 5, section 2.4.

4.3 Improper movement, and proper improper movement

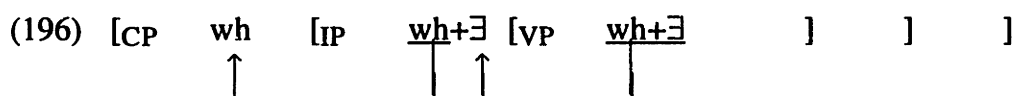
Consider cases of improper movement, such as the one in (195):

(195) *[How many people are known [*t* [it was told *t'*]]]

In (195), *how many people* begins as the object of *told*. It undergoes A'-movement to the specifier of the embedded CP, where it checks the strong [+wh] feature there. Finally, it A-moves into the external subject position of the highest clause, where it checks Case. The sentence should thus be able to mean something like *It is known how many people were told*. Locality seems to be respected throughout the derivation, as is Procrastinate, and feature-checking takes place as it should.

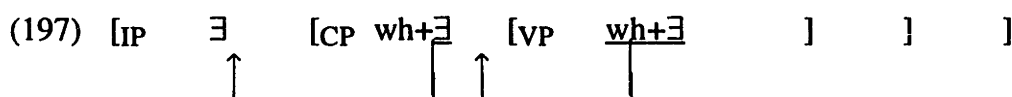
The theory developed here can rule out the derivation in (195), without need of an additional stipulation ruling out Improper Movement. Both of the movements in (195) check strong features, and the resulting chain is therefore an ill-formed PF object. Improper Movement can thus be ruled out on independent grounds.

In the previous section we saw apparent well-formed cases of a chain with two positions associated with strong features; these were the cases of overt wh-extraction from the subject position in languages like English. I suggested that such cases were well-formed because they in fact involve two chains; the entire wh-phrase undergoes A-movement to the external subject position, and wh-movement then targets only the wh-component of the wh-phrase:



We should make sure that this means of licensing chains of this type is not available to license improper movement.

In fact, there is reason to believe that this tactic would lead to a semantically uninterpretable chain in the case of improper movement. In this case the movement to the higher position does not involve wh-feature checking; rather, it involves checking of a feature of the type we associate with the existential component of the wh-phrase. The derivation in (195) would therefore have to involve movement of the type in (197):



The structure in (197) will involve an instance of vacuous quantification, assigning the sentence in (195) an interpretation like that in (198):

(198) There are n -many people x_i such that it is known for what n : x_i was told.

The variable n in the existential component which must be bound by the *wh*-component is not bound in (198), and the structure is therefore semantically ill-formed.

This approach to improper movement allows two classes of well-formed improper movements. One would be a case in which A' -movement with no quantificational force is followed by A -movement. Improper movement of this type would be immune to the semantic factors which rule out (198); if the feature driving A' -movement is unlike the *wh*-component of a *wh*-word in that it does not bind a variable in the remainder of the moving element, then no problem of vacuous quantification should arise.

Tough-movement is one candidate for a well-formed improper movement of the relevant type:

(199) John is tough [t PRO to please]'

The diagram shows the phrase 'John is tough' followed by a bracketed structure '[t PRO to please]''. An upward arrow points from 'PRO to please' to 't', indicating movement.

In this case the feature which drives A -bar movement to the intervening Spec CP is not quantificational, and a structure like that in (198) can therefore be well-formed:

(200) [IP \exists [CP X+ \exists [VP X+ \exists]]]

The diagram shows the hierarchical structure [IP \exists [CP X+ \exists [VP X+ \exists]]]. An upward arrow points from the 'X+ \exists ' in the VP to the 'X+ \exists ' in the CP specifier, indicating movement.

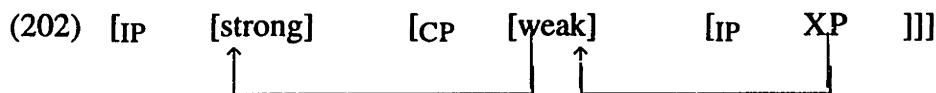
Thus, this approach might alleviate somewhat the problematic status of *tough*-movement in the theory.

Note that this analysis of *tough*-movement combines with the approach to subject extraction in English outlined above to predict that *tough*-movement out of a subject position should never be possible. Overt extraction out of subject position involves leaving the existential component of the extracting nominal behind, and *tough*-movement crucially involves moving this component into a specifier above the A' -specifier through which movement passes. These two requirements should be incompatible. As has been widely

discussed in the literature (cf. Chomsky 1973, Stowell 1986, Browning 1987, Cinque 1990, Takahashi 1997), this is in fact the case²³:

- (201) a. ?That kind of car is difficult to believe that John would buy t
 b. *That kind of guy is difficult to believe t would buy a Ferrari

There is a second class of improper movements which are allowed by the approach developed here. Recall that on this theory, improper movement is a special case of a chain associated with two strong features; in English, for instance, overt wh-movement cannot be followed by overt movement to check Case. On the other hand, in a language in which the feature driving wh-movement is weak, such improper movement should be entirely possible:



The chain in (202) is a well-formed PF object; it is associated with a single strong feature, which instructs PF to pronounce the head of the chain. If it is the case, as I have suggested, that improper movement is ruled out solely because of well-formedness conditions imposed by PF, then improper movement should be possible when the feature driving the intermediate A'-movement is weak.

In fact, there is some evidence from Japanese that this is the case. Saito (1992) notes that long-distance scrambling and local scrambling typically differ in that only the latter can remedy weak crossover:

²³ Similar facts have been noted for parasitic gaps:

- i. a. Which book did you buy t before finding out that John already had t?
 b. *Which guy did you buy a book for t before finding out that t had already bought it?

It seems to me that the approach developed above could be expanded to deal with parasitic gaps as well, but I will not try to do so here.

Japanese (Saito 1992, 115)

- (203) a. *Dono hon-ni -mo sono tyosya -ga t keti-o tuketa*
 which book on also its author NOM threw-cold-water
 ‘Every book_i, its_j author threw cold water on’
- b. ?**Dono hon-ni -mo sono tyosya -ga [Hanako -ga t keti-o tuketa to]*
 which book on also its author NOM Hanako NOM threw-cold-water that
 itteiru
 is-saying
 ‘Every book_i, its_j author says that Hanako threw cold water on’

On the other hand, Saito notes that long-distance scrambling of a wh-word can in fact remedy weak crossover:

Japanese (Saito 1992, 108-109)

- (204) a. ?*Dare -o soitu -no hahaoya-ga t aisiteiru no*
 who ACC guy GEN mother NOM love Q
 ‘Who_i, his_j mother loves’
- b. ?*Dare -o soitu -no hahaoya-ga Hanako -ga t aisiteiru to omotteiru no*
 who ACC guy GEN mother NOM Hanako NOM love that think Q
 ‘Who_i, his_j mother thinks that Hanako loves’

On the account of improper movement developed here, the contrast between long-distance scrambling of a quantifier and long-distance scrambling of a wh-word is unsurprising. The wh-word has the option of stopping in the intermediate Spec CP²⁴, a position associated with a weak feature in Japanese; from this position, scrambling into the higher clause can be local, and thus can have the properties of A-movement. A long-distance scrambled quantifier, by hypothesis, lacks such an intermediate landing site²⁵.

²⁴ Recall from Chapter 2 that even in an IP-absorption language like Japanese, Spec CP is available as an intermediate landing site for wh-movement; see section 5 of that chapter, in particular, for evidence to that effect.

²⁵ Saito notes that negative polarity items pattern with wh-words in this regard:

Japanese (Saito 1992, 109)

(i) ?*Dono hito -mo soitu-no hahaoya-wa [Hanako -ga t aisiteiru to] omottenai*

As we expect on this approach, if the intermediate landing site for the wh-word is not available, the ability of long scrambling to remedy weak crossover vanishes:

Japanese (Shigeru Miyagawa, p.c.)

(205) a. ?Dare-o soitu -no okaasan -ga [John -ga t sikatta to] itta no

who ACC guy GEN mother NOM John NOM scolded that said Q

‘Who_i, his_j mother thinks that John scolded’

b. *Dare-o soitu -no okaasan -ga [John -ga t sikatta ka] siritagatteiru no


who ACC guy GEN mother NOM John NOM scolded Q wonders Q

‘Who_i, his_j mother wonders whether John scolded’

(205b) is much worse than (205a), and is also worse than an ordinary wh-island violation in Japanese; in fact, it has the status of a weak crossover violation. This is what we expect; in (205b), the intermediate landing site for long-distance wh-movement is unavailable, and improper movement therefore cannot take place.

Let us return now to English improper movement:

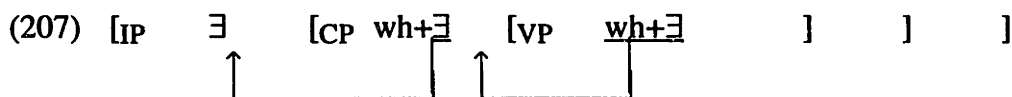
(206) *[How many people are known [t [it was told t']]]



(206) is a derivation involving A-bar movement to an interrogative complementizer in the embedded clause, followed by A-movement into the matrix clause, yielding a sentence which should be able to mean something like *It is known how many people were told*. The result is a chain associated with two strong features, which is an ill-formed PF object. This chain cannot be salvaged by separating the moving NP into two related chains, as we did for wh-extraction out of the subject position in English, because the result is semantically ill-formed:

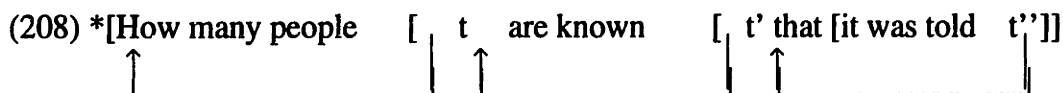
which person also guy GEN mother TOP Hanako NOM love that think-NEG
 ‘Anyone_i, his mother does not think that Hanako loves t_i’

If the account developed here is on the right track, the well-formedness of (i) must indicate that negative polarity items are licensed by a weak feature in Neg⁰, and can use Spec NegP as an intermediate landing site, just as wh-words use Spec CP as an intermediate landing site. We are also driven to the conclusion that Spec NegP is situated below the highest possible landing site of A-scrambling.



The representation in (207) involves a violation of the ban on vacuous quantification; the wh-component of the wh-word fails to bind a variable in the existential component.

Ideally, we want the above account to generalize to rule out the ill-formed derivation in (208):



In (208) the wh-phrase *how many people* undergoes A-bar movement into the specifier of the embedded, non-interrogative C⁰, followed by A-movement into the matrix clause, followed by A-bar movement into Spec CP of the matrix clause; the resulting representation should be assigned a meaning like that of *How many people are known to have been told?* If the only strong features in this derivation are those in the matrix clause, it is not clear how the derivation can be ruled out on the account developed here.

In order to rule this derivation out, then, I will need to assume two things. First, as was suggested above in section 2.3, let us assume that the difference between interrogative and declarative complementizers is purely interpretive; interrogative and non-interrogative complementizers are formally identical, both having a single strong feature in English, and differ only in the semantic component. The first two steps in the derivation in (208) are thus identical to the derivation in (206); both steps are driven by a strong feature. The second crucial assumption will be that the derivation terminates when the semantically ill-formed representation in (208) is created. Thus, the fact that the wh-component will be raised by a strong feature into the matrix clause at a later point in the derivation, where it can bind the variable in the existential component, cannot save the derivation.

This account raises difficult questions about the nature of successive-cyclic movement, an operation which is already problematic in a Minimalist approach. If both interrogative and declarative complementizers have a single strong feature in English, then

wh-movement from one complementizer to the next should always involve creation of a chain with multiple strong features. One way of circumventing this problem would be to claim that the strong feature on a declarative complementizer, being uninterpretable, vanishes once the wh-word moves on to check another strong feature on another complementizer. Crucially, the feature would have to wait to delete until the wh-word moved to a complementizer; otherwise, this escape hatch could rule back in the improper movement derivation just discussed. This theoretical move would therefore involve a comparatively complex notion of “interpretable”; the computational system would be unable to determine that the strong feature on the intermediate complementizer was uninterpretable until the wh-word moved to a different complementizer, thus demonstrating that the intermediate complementizer was not its final scopal position.

We have seen that English and Japanese treat Spec CP in crucially different ways; in Japanese, but not in English, Spec CP can be used as an intervening landing site on the way to an A-position²⁶. In the account developed here, this contrast is linked to the fact that English, but not Japanese, has overt wh-movement. The availability of overt wh-movement in a language indicates that all complementizers have a strong feature, while its absence indicates that only weak features are present on the complementizer. Thus, even complementizers which are not interrogative differ in the two languages, as desired.

4.4 Japanese subject scrambling

Saito (1985) notes that scrambling of subjects is impossible in Japanese:

²⁶ See also Baek (in progress) for evidence from Korean ECM constructions that Spec CP can be used as an intermediate landing site for A-movement in Korean.

Japanese (Saito 1985, 185, 193)

(209) a.* Sono okasi -ga John -ga [t oisii to] omotteiru

this candy NOM John NOM tasty that thinks

‘This candy, John thinks is tasty’

b.* Kono giron -ga Mary -ga John -ni [t okasii to] itta

this argument NOM Mary NOM John DAT strange that said

‘This argument, Mary told John is strange’

As Saito notes, one might claim that the ill-formedness of (209a) arises from a processing strategy which assumes the least amount of scrambling possible. On an account of this kind, the two nominative-marked NPs at the beginning of (209a) would be interpreted as being in their base positions, thus giving the sentence the semantically anomalous meaning ‘this candy thinks that John is tasty’.

Saito argues that this is not the correct approach, however. For one thing, it is not clear how an account of this kind would deal with the ill-formedness of (209b), where failure to posit scrambling of the first nominative-marked NP leads to a subcategorization violation, there being no syntactically appropriate place to put the dative NP *John-ni*. Another problem is that the facts in (209) contrast crucially with a similar restriction on sequences of dative-marked NPs, which are preferentially parsed as being unscrambled (an observation Saito attributes to Kuno 1980):

Japanese (Saito 1985, 190)

(210) John -ga Bill -ni Mary -ni hana -o todokesasetta

John NOM Bill DAT Mary DAT flower ACC deliver-CAUS

‘John made Bill deliver flowers to Mary’

* ‘John made Mary deliver flowers to Bill’

Sentences like (210) are given a reading in which the dative-marked NPs are in their base order, and cannot be interpreted as if they had undergone scrambling. At first glance, as Saito notes, this phenomenon might be taken to account for the ill-formedness of (209a) as

well; a sequence of nominative-marked NPs, like a sequence of dative-marked NPs, must be interpreted as though no scrambling had taken place.

Crucially, however, the parsing preference for sequences of dative NPs can be overridden by pragmatic considerations:

Japanese (Saito 1985, 192)

(211) John -ga [Bill -no ie -ni] Mary -ni t_j hana -o todokesaseta

John NOM Bill GEN house DAT Mary DAT flower ACC deliver-CAUS

‘John made Mary deliver flower’s to Bill’s house’

The sentence in (211) can be interpreted as though the dative-marked NP *Bill-no ie-ni* ‘Bill’s house-DAT’ had scrambled over the other dative-marked NP *Mary-ni* ‘Mary-DAT’; thus, it need not receive the nonsensical interpretation ‘John made Bill’s house deliver flowers to Mary’.

In the case of nominative arguments, however, pragmatic considerations have no such effect, as the example in (212) shows:

Japanese (Saito 1985, 185)

(212) *Sono okasi -ga John -ga [t oisii to] omotteiru

this candy NOM John NOM tasty that thinks

‘This candy, John thinks is tasty’

(212) cannot be interpreted as involving scrambling of one nominative argument over another, even though an interpretation in which both nominative arguments are in situ (‘this candy thinks that John is tasty’) is pragmatically implausible. Saito thus concludes that the parsing preference for sequences of dative NPs is not what rules out scrambling of nominative NPs; such scrambling is apparently impossible for syntactic reasons.

In the theory developed here, we expect scrambling of subjects to be ill-formed, on the assumption that movement to the external subject position is obligatorily driven by a strong feature in Japanese. As Watanabe (1996) points out, there is some independent motivation for believing that this feature is strong. The relevant data have to do with *ga-no*

conversion in Japanese, a process whereby the subject of a relative clause is marked with Genitive case rather than Nominative:

Japanese (Watanabe 1996, 210)

(213) [John -ga / no nihon-e kaetta] hi

John NOM GEN Japan to returned day

‘the day on which John came back to Japan’

Interestingly, *ga-no* conversion is subject to a transitivity restriction:

Japanese (Watanabe 1996, 211)

(213) [John -ga / *no LGB -o kashita] hito

John NOM GEN LGB ACC lent person

‘The person to whom John lent LGB’

The transitivity restriction does not hold, however, if it is the direct object which is relativized:

Japanese (Watanabe 1996, 211)

(214) [John -ga / no katta] hon

John NOM GEN bought book

‘the book that John bought’

As Watanabe notes, the distribution of *ga-no* conversion is very reminiscent of that of French stylistic inversion; it is possible with intransitive verbs, and impossible when the verb is transitive, unless it is the direct object which is extracted:

French (Kayne and Pollock 1978, 595; Valois and Dupuis 1992, 327; Pollock 1981, 230)

(215) a. *Quand partira ton ami?*

when will-leave your friend

‘When will your friend leave?’

b. **Je me demande quand mangera sa pomme Marie*

I me ask when will-eat her apple Mary

‘I wonder when Mary will eat her apple’

c. *Que crois -tu que manquent un grand nombre d’étudiants?*

what believe-you that be-absent-from a great number of-students

‘What do you think that many students are absent from?’

Thus, it seems reasonable to equate genitive marking of the subject of a Japanese relative with French stylistic inversion; if this is correct, then genitive marking indicates that the feature driving raising to the subject position is weak, and nominative marking indicates that it is strong. Thus, we have evidence that movement to the external subject position is overt in Japanese, at least when the subject is marked with nominative case. Raising to the external subject position and scrambling should therefore yield a chain associated with two strong features, an illegitimate PF object, according to the theory developed here. We thus expect subject scrambling to be impossible, which it is.

4.6 Conclusion and Expansions

In this section we have seen a number of strategies in various languages for dealing with chains containing multiple strong features. The theory sketched here appears to give a natural account of a number of phenomena by means of a fairly simple constraint on well-formed PF objects. I have claimed that a chain involving multiple positions associated with strong features is intolerable. Such chains can in principle be redeemed in a number of ways; the methods discussed above include substituting a weak feature for one of the offending strong features (the method used in the anti-agreement cases, for instance),

deleting an uninterpretable strong feature (used in successive-cyclic wh-movement in languages like English), and extraction of a subpart of a moving element, thus creating two chains instead of one (the strategy for extraction of subjects in some languages, which is responsible for complementizer-trace phenomena). These strategies are sometimes restricted by semantic factors. For instance, a strong feature cannot be deleted unless it is semantically uninterpretable; this strategy is therefore available for successive-cyclic wh-movement, for instance, but not for wh-extraction of subjects.

In the previous sections we saw a number of cases in which null operators were capable of overt movement to check weak features; I suggested that this was due to the freedom of such operators from the PF constraint banning overt movement to check weak features. Given the theory as it has been presented thus far, we might also expect to find that null operators are also free to undergo overt movement to check multiple strong features.

In fact, this does not appear to be the case. Anti-agreement phenomena, which were analyzed here as a means of avoiding chains associated with multiple strong features by substituting a weak feature for a feature which is ordinarily strong, appear with null relativization operators, as well as with overt wh-words. In Fiorentino and Trentino, for instance, weak agreement must be used when a subject is relativized:

Fiorentino (Brandi and Cordin 1989, 126)

(216) a. *Le ragazze che gli è venuto ieri*

the girls that it is come yesterday

‘the girls that came yesterday’

b. **Le ragazze che le sono venute ieri*

the girls that 3.PL.FEM. are come-FEM.PL. yesterday

Trentino (Brandi and Cordin 1989, 126)

(217) a. Le putele che è vegnú algeri

the girls that is come yesterday

‘the girls that came yesterday’


b. *Le putele che le è vegnude algeri

the girls that 3.PL.FEM. is come-FEM.PL. yesterday

Thus, null operators are apparently not immune to the ban on being associated with multiple strong features. As the theory has been developed thus far, this is surprising. I have treated chains associated with multiple weak features and no strong features and chains associated with multiple strong features as ill-formed for essentially the same reason; both of these types of chains contain multiple candidates for pronunciation which are all equally viable alternatives, and PF has no way of choosing between them. The fact that null operators trigger anti-agreement phenomena suggests that this is too simple.

Consider again the two relevant types of offending chains:

(218) a. [weak] [weak]



b. [strong] [strong]



Recall that feature strength, on this theory, is an instruction to PF to pronounce the copy in an immediate checking relation with the feature in question. The chain in (218a), then, contains no instructions to PF of any kind. The chain in (218b), on the other hand, contains contradictory instructions, requiring pronunciation of both features. One possible way of dealing with the different status of null operators with respect to these two types of chains would be to capitalize on this difference between them. PF needs no instructions in how to pronounce a null element, so the chain in (218a) is well-formed if the element undergoing movement is null. Contradictory instructions as to which position to pronounce, on the other hand, are “confusing” for PF regardless of how the moving element is to be pronounced.

5 Overall conclusions

In this chapter I have developed a theory of well-formed PF objects which is effectively a representational version of Chomsky's (1993) *Procrastinate*. The basic idea is that in order for a chain to be a well-formed PF object, PF must receive unambiguous instructions from the syntax as to which part of the chain to pronounce. Feature strength, on this account, is essentially an instruction to pronounce the copy in a chain in a checking relation with the strong feature. I have tried to show that this approach has empirically desirable consequences.

The approach has consequences for the architecture of the theory, as well.

Consider the requirement of Featural Cyclicity which played a crucial role in chapter 3:

(219) A strong feature must be checked as soon as possible after being introduced into the derivation.

Given the theory developed here, (219) can be modified to (220), a constraint very reminiscent of Pesetsky's (1989) *Earliness*:

(220) A feature must be checked as soon as possible after being introduced into the derivation.

(220) will be constrained by the requirement that it create well-formed PF and LF objects, and will thus typically be prevented from triggering overt movement to check a weak feature. (220) and (219) do differ empirically in at least one case, that being the case of overt movement to check weak features. The prediction of (220) is that this type of movement should always occur whenever it is possible (for instance, when the moving element is phonologically null, or has previously undergone movement to check a strong feature); weak features should behave just like strong features in this case. In fact, there is some evidence that this is true. As we saw in section 1.4 of this chapter, Japanese null relative operators must apparently undergo overt movement. In sections 3.1 and 3.2, we saw that overt movement to check a weak feature is apparently obligatory when licensed by

an earlier movement to check a strong feature; these were the cases of wh-scrambling in Japanese and partial wh-movement in Malay.

Another interesting result of the analysis proposed here was a particular approach to successive-cyclic wh-movement. We saw evidence from the properties of improper movement in English and Japanese and of partial wh-movement in Malay, in particular, that suggested that interrogative and declarative complementizers have identical sets of features, differing only in interpretation. Thus, English C^0 always has a strong wh-feature; when the complementizer is declarative, this feature is uninterpretable, and therefore vanishes after checking. More specifically, I claimed that the uninterpretable feature vanishes after the wh-word moves to another complementizer, this being the point in the derivation at which the computational system can “see” that the wh-feature on the intermediate complementizer must be uninterpretable. This assumption was useful in sections 2.3, 3.2, and 4.3.

One aspect of the theory developed here which is clearly undesirable is its crucial reference to the property of feature strength. Ideally, we should try to develop a theory in which the decision between overt and covert movement is made on the basis of general principles, rather than by stipulative labels like “strong” and “weak”. The approach developed here will hopefully be helpful in the eventual development of such a theory. This approach is predicated on the assumption that all strong features share some property in common, namely that of being an instruction to PF to pronounce a particular copy. A more explanatory theory which replaces strength with some set of more general syntactic principles might not have this property; the effects of what is currently called “strength” might be derived in one way in one case and in another in another case. Such a theory might make different predictions, then, from the one developed here. For instance, a chain with multiple strong features, in this theory, is always intolerable, and can be redeemed in certain ways, subject to semantically based restrictions. A theory which derives strength from basic principles and thus allows for different types of strong features might allow a

chain with multiple strong features if these features were of different types; the putative cases of chains of the offending type (e.g., successive-cyclic wh-movement, or extraction from subject position in a language without anti-agreement) might be licensed in a different way. I will have to leave the development of a theory of this type for later work.

Chapter Five: The Principle of Minimal Compliance

1. Introduction

In chapter 2 I introduced and defended an approach to the data in (1) (Baker 1970):

- (1) a. *What do you know [who bought t]?
- b. Who knows [who bought what]?

Baker (1970) observed that (1b) can have a reading in which *what* has matrix scope; covert movement of *what* is apparently immune to the effects of the *wh*-island which rules out (1a). I argued that the best type of approach to these facts was what I referred to as a “Subjacency Tax” approach, in which the well-formed overt movement of *who* made the matrix C⁰ immune to the effects of Subjacency for the rest of the derivation. In this chapter I will argue that this kind of interaction between dependencies is a general property of the grammar, not just of Subjacency phenomena.

The phenomena exemplified in (2-6) seem to share a common quality with the Subjacency cases we have been discussing; in all of these cases, a dependency which

would be ill-formed in isolation is somehow “saved” by the presence of a well-formed dependency. In other words, it looks as though in these cases, at least, a given constraint only has to be satisfied once in a certain domain:

Reflexivity (Reinhart and Reuland 1993)

- (2) a. *Henk_i wees zich_i aan mij toe
Henk assigned self to me
- b. Henk_i wees mij aan zichzelf_i toe
Henk assigned me to self-self
- c. Henk_i wees zich_i aan zichzelf_i toe
Henk assigned self to self-self

Weak Crossover (Hornstein 1995)

- (3) a. *Who_i did his_i mother introduce t_i to Mary?
- b. Who_i did John introduce t_i to his_i teacher?
- c. ?Who_i did his_i mother introduce t_i to his_i teacher?

VP-ellipsis¹ (Danny Fox, p.c.)

- (4) a. *I introduced John and Mary_i to Bill before they_i could (introduce themselves_i to Bill)
- b. John introduced himself to Mary before Bill_i could (introduce himself_i to Mary)
- c. I introduced John and Mary_i to each other_i before they_i could (introduce themselves_i to each other_i)

Connectedness (Kayne 1983a, 1983b)

- (5) a. *a person who_i people that talk to e_i usually like him
- b. a person who_i people that talk to John usually like e_i
- c. a person who_i people that talk to e_i usually like e_i

¹ Here the material in parentheses corresponds to the elided VP.

That-trace effects (Lasnik and Saito 1984)

- (6) a. *Who_i do you think t'_i that t_j left?
 b. Who_i do you think t'_i t_j left?
 c. Who_i do you think t''_i that John said t'_i t_j left?

The (a) cases all violate some constraint which the (b) cases obey. Our standard assumptions about the way in which constraints operate lead us to expect that sentences containing both the configuration in (a) and that in (b) should be ill-formed; if the configuration in (a) is what violates the constraint in question, then any sentence containing that configuration should be doomed. In fact, this is not the case; the (c) sentences are better than the (a) sentences. There have been a variety of attempts to explain these facts, of course, but no unified account of all of them has ever been given, to my knowledge. In this chapter I will try to isolate the factor that (1-6) have in common, thus simplifying the constraints involved. I will then go on to apply this principle to the notion of Shortest Move, deriving a number of additional-wh effects in various languages.

The idea developed here is essentially that our standard assumptions about the way in which constraints operate are incorrect. Traditionally, constraints are thought of as checking the entire structure for violations, and rejecting structures that contain any violations. The data in (1-6) suggest that in some cases, at least, the presence of a dependency that satisfies a constraint can allow the computational system to ignore another dependency which would be ill-formed in isolation. The idea I want to develop here is that the computational system tries to avoid checking the same constraint in the same portion of the structure more than once². As a result, if a constraint is obeyed, the portion of the structure in which it is obeyed can be ignored thenceforth with respect to that constraint.

Let us formalize this idea as the Principle of Minimal Compliance (PMC), given in (7):

² We might think of this as an application of Economy to the process of constraint-checking.

(7) Principle of Minimal Compliance

For any dependency D that obeys constraint C, any elements that are relevant for determining whether D obeys C can be ignored for the rest of the derivation for purposes of determining whether any other dependency D' obeys C.

The notion of “relevance” involves at least the following qualifications:

(8) An element X is relevant to determining whether a dependency D with head A and tail B obeys constraint C iff

- a. X is along the path of D (that is, X=A, X=B, or A c-commands X and X c-commands or dominates B).

AND b. X is a member of the class of elements to which C makes reference.

(8a) allows the PMC to apply only to elements along the “path” of a dependency. We could try to build (8a) into the statement of the PMC, but it is interesting to note that there seem to be no constraints making reference to elements that are not in the set defined by (8a) (for example, there are no constraints rendering a dependency ill-formed if some element is present in a position c-commanded by the head which does not c-command or dominate the tail). However this fact is to be reflected in our theory, it seems clear that it should not be specific to the PMC; rather, let us assume that the PMC is “aware” of the fact that elements not lying on the path of the dependency are always irrelevant for determining whether the dependency is well-formed. (8b) is more straightforward; it effectively tells us not to bother with things not mentioned by the constraint, and is roughly akin to the basic insight of Rizzi’s (1990) Relativized Minimality.

2. Applications

Let us go on to see how the data in (1-6), among others, follow from simple constraints in conjunction with the PMC.

2.1 Reflexivity

The facts in (9) now follow straightforwardly from application of the PMC to something like the constraint in (10)³:

- (9) a. *Henk_i wees zich_i aan mij toe
 Henk assigned self to me
- b. Henk_i wees mij aan zichzelf_i toe
 Henk assigned me to self-self
- c. Henk_i wees zich_i aan zichzelf_i toe
 Henk assigned self to self-self

(10) In Dutch, all anaphors bound within their clause must be *zichzelf*.

Here the dependency between *Henk* and *zichzelf* obeys the constraint on reflexives in (10).

The PMC then allows us to ignore all elements which are relevant for determining that the dependency between *Henk* and *zichzelf* obeys (10) from then on for purposes of determining whether other dependencies obey (10). *Henk* and *zichzelf* are themselves among the relevant elements, since the constraint refers to an anaphor and its binder and they are along the path of the dependency. Thus, in particular, we will be allowed to ignore *Henk*⁴ for purposes of determining whether other anaphors in the sentence obey (10). Next we look at *zich*, to determine whether it obeys (10). If we ignore *Henk*, the anaphor is not bound within its clause. (10) is therefore satisfied⁵.

Of course, *zich*, being an anaphor, does in fact have to be bound by *Henk* in order to be interpreted. For the account sketched above to be maintained, we must apparently understand *zich* as being subject to two separate constraints, the one in (10) and another

³ We will obviously want a less language-specific principle to deal with the phenomena in (9), but (10) will suffice for our purposes. For an attempt to derive (10) from general principles of Economy, see Richards 1995b.

⁴ We can also ignore *zichzelf*, of course, although this is not relevant here.

⁵ Veraar. (1996) argues for a different way of understanding these data; constraints of time prevent me from addressing her proposal here, but if she is correct the Reflexivity facts will not fall under the PMC at all (although the rest of the cases discussed here are unaffected, of course).

which requires it to have an antecedent. The PMC allows us to ignore *Henk* with respect to the constraint in (10), but has no effect on the constraint that requires *zich* to be bound, which is still in force. This could be for a number of reasons; for instance, the requirement that *zich* be bound might be made to follow from properties of the meaning of *zich*, which would presumably be unaffected by the PMC. It seems clear that in principle we do want the PMC to apply separately to different constraints; we do not want an anaphor which is in a well-formed binding relation to be immune to the Case Filter (or its successors), for instance.

At the same time, there is something conceptually awkward about separating the constraints on *zich* in this way. We might prefer, for instance, a theory of anaphoric binding which derives all the properties of anaphoric binding from a single requirement, rather than from several distinct ones. Such a theory would derive both the requirement that *zich* be bound and that it not be locally bound from a single constraint. If we were to construct such a theory, the version of the PMC sketched above would be inadequate; ignoring a binder for purposes of the requirement that *zich* not be locally bound would entail ignoring it for purposes of the requirement that *zich* be bound, and would not improve the structure.

One possible move to make at this point would be to reformulate the PMC as marking parts of the structure as having already obeyed a constraint, rather than allowing us to “ignore” portions of the structure. On such an account, the well-formed relation between *Henk* and *zichzelf* would mark *Henk* as obeying the generalized condition on anaphora, and would allow *Henk* to locally bind *zich*.

Despite the conceptual attractiveness of this move, I will not make it here. The approach as sketched immediately above is too strong. Suppose we were to succeed in developing a theory in which all the properties of anaphoric binding, including the

requirement that anaphors be bound, the choice between *zich* and *zichzelf*, and the locality constraints on certain anaphors (such as *zichzelf*), were built into a single constraint. Certain parts of this constraint would appear to be subject to the PMC, including the choice between *zich* and *zichzelf*, as we have seen. Others clearly are not; for instance, a binder which binds an anaphor in an appropriately local fashion is not then licensed to illegitimately bind another anaphor long-distance. The version of the PMC which was just sketched cannot distinguish between these cases; if a binder enters into a well-formed binding relation, it should then be invulnerable to all the constraints on binding, which is not the case. We must apparently understand anaphors as being subject to a number of separate constraints, which interact with the PMC in different ways.⁶

2.2 Weak Crossover

The facts in (11) follow in a similar way from the PMC applied to a straightforward constraint on weak crossover:

Weak Crossover (Hornstein 1995)

- (11) a. *Who_i did his_i mother introduce t_j to Mary?
 b. Who_i did John introduce t_j to his_i teacher?
 c. ?Who_i did his_i mother introduce t_j to his_i teacher?
- (12) All pronouns bound by a wh-word must also be bound by a trace of that wh-word in an A-position.

In (11c), the relation between *who* and the second instance of *his* obeys (12). This will enable us to ignore *who* for purposes of determining whether other pronominal variables obey (12), just as we could ignore *Henk* in the Reflexivity case. Thus, the first instance of *his* is effectively not bound by a wh-word, since we are allowed to ignore the wh-word

⁶ This need not be as conceptually awkward as it might seem. For instance, a theory of anaphor binding in which the various constraints on anaphors follow from general constraints on dependencies which are not specific to anaphors (for instance, in which the locality conditions follow from Shortest Move, while the matter of selection among different anaphoric forms is left to Structural Economy) would have this character.

that does bind it, and the structure is well-formed. The facts are quite similar to those in the Reflexivity case, and fall out in a similar way; a constraint on possible relations between binders and bound elements need only be satisfied once by a given binder. After a binder obeys the constraint once, the PMC allows us to ignore it thenceforth for purposes of evaluating the constraint, and the constraint thus fails to apply to anything else bound by that binder.

2.3 VP-ellipsis

Similar reasoning yields the paradigm of VP-ellipsis described below:

VP-ellipsis⁷ (Danny Fox, p.c.)

- (13) a. *I introduced [John and Mary]_i to Bill before they_i could (introduce themselves_i to Bill)
- b. John introduced himself to Mary before Bill_i could (introduce himself_i to Mary)
- c. I introduced [John and Mary]_i to each other_i before they_i could (introduce themselves_i to each other_i)
- (14) Bound variables in an elided VP must correspond to bound variables in the non-elided VP model.

The facts in (13) seem to show that an R-expression cannot have a correspondent in an elided VP which is an bound variable, unless the elided VP also contains a bound variable which corresponds to a bound variable in the unelided VP. Thus, (13a), where the R-expression *John and Mary* corresponds to a bound variable in the elided VP, is ill-formed. (13c), which contains the same configuration but also contains a bound variable corresponding to the bound variable *each other*, is well-formed. The Principle of Minimal Compliance allows us to understand this pattern as involving the parallelism requirement (14) (probably a specific case of a more general requirement of parallelism between elided VPs and their non-elided counterparts). Thus, the relation between *they* and *each other* in

⁷ Here the material in parentheses corresponds to the elided VP.

the elided VP in (13c) is one of variable-binding, but this relation obeys the constraint in (14), since *each other* is a bound variable corresponding to a bound variable in the unelided original. This allows us to ignore *they* and *each other* for purposes of computing the well-formedness of the rest of the structure with respect to (14). As a result, we can ignore the binder for *themselves* in the elided VP, and *themselves* is thus effectively not a bound variable, since we can ignore its binder; (14) is therefore satisfied. The pattern is thus entirely parallel to the weak crossover and Reflexivity facts discussed above; an element of a certain kind cannot be bound unless another element of a different kind is also bound. The well-formed binding relation allows us to ignore the potentially offending binder, and all other elements bound by that binder are thus free from the constraints in question.

2.4 *that*-trace effects

In section 4.2 of chapter 4 I sketched an account of contrasts like those in (15-16):

- (15) a. *Who_i do you think t'_i that t_i left?
 b. Who_i do you think t'_i t_i left?
 c. Who_i do you think t''_i that John said t'_i t_i left?
- (16) a. *Who_i do you wonder t'_i whether t_i can help us?
 b. ?*Who_i do you wonder t''_i whether we think t'_i t_i can help us?

Following much work on this topic (cf. Lasnik and Saito 1984, Rizzi 1990), I suggested that traces of *wh*-moved subjects must enter into a certain type of relation with the complementizers whose specifiers serve as landing sites for extraction. I further assumed that *that* and *whether* are unable to enter into this relation. The account is thus very similar to that of Lasnik and Saito (1984), which assumes a version of the ECP including a principle something like (17):

- (17) Certain traces must be antecedent-governed, and *that* and *whether* block antecedent-government.

Let us assume, following Lasnik and Saito, that all of the traces in (15-16) are of the type requiring antecedent-government (in their terms, none of these traces are lexically governed). (15a), on this story, is ill-formed because *that* intervenes between t'_i and t_i , blocking antecedent-government of t_i . This then leads to the question of why (15c) is better than (15b), given that antecedent-government of t'_i by t''_i is presumably also blocked by *that*; similarly, the improvement of (16b) over (16a) is unexpected, assuming that *whether* prevents antecedent-government. Lasnik and Saito posit a process of deleting intermediate traces, which accounts for the well-formedness of (15c) by claiming that t'_i is deleted after it marks t_i as properly antecedent-governed but before obedience to (17) is checked. According to the theory developed here, the relation between t'_i and t_i in (15c) obeys (17), thus enabling us to ignore both t'_i and t_i with respect to (17). As a result, the relation between t''_i and t'_i is rendered well-formed by the PMC; since we can ignore t'_i , the fact that t'_i is not antecedent-governed is no longer a problem. This lets us avoid some⁸ of the complications of Lasnik and Saito's (1984) analysis, and derives a problematic section of the *that*-trace paradigm from a principle which seems to be independently necessary.

Note that the constraint in (17) demands that for every trace there be some antecedent that stands in an appropriate relation to the trace. As a result, ignoring a binder which fails to properly antecedent-govern will never help a dependency which obeys (17), since the offending traces will still lack an antecedent-governor if their binder is ignored. In other words, sentences like (18) are correctly predicted to be ill-formed:

(18) *Who_i did you think t''_i John said t'_i that t_i left?

⁸ One complication which does not fall out of the account developed here is the distinction between argument and adjunct traces. I will follow Szabolcsi and Zwarts (1993) and Sauerland (1996b) in assuming that the behavior of adjuncts with respect to islands is at least partly due to semantic factors; see section 3.2 of chapter 4 for some discussion.

Here the relation between t''_i and t'_i obeys (17), which allows us to ignore t'_i for purposes of evaluating the structure for (17). However, this does not save the structure; t_i is still not antecedent-governed, even if we do ignore t'_i , and (18) is still ill-formed.

2.5 Subjacency, CED

Subjacency has classically (Huang 1982 and much subsequent work) been claimed to apply only to overt movement, on the basis of contrasts like those in (19):

- (19) a. *What_j do you wonder who bought t_j?
 b. Who_i t_j wonders who bought what_j?

(19a) is a standard wh-island violation; *what*, by crossing the +wh comp whose specifier is occupied by *who*, violates Subjacency. (19b) has an interpretation in which *what* and *who_i* both have matrix scope; this reading is classically taken to be derived by LF movement of *what* to the specifier of the matrix CP. This move presumably also violates Subjacency, just like its overt counterpart in (19a), yet the structure is well-formed. The conclusion which has typically been drawn from this is that Subjacency only applies to overt movement.

In chapter 2 I defended a different approach to this kind of phenomenon, claiming that movement of *what* in (19b) need not obey Subjacency because movement of *who* into the matrix Spec CP has already obeyed Subjacency. One piece of evidence I offered for this conclusion was the fact that the effect of adding an additional wh-word shown in (19b) can be found regardless of the level at which the movement takes place; thus, facts similar to those in (19) can be found in languages like Japanese, which do all wh-movement covertly, and in languages like Bulgarian, which do all wh-movement overtly:

Japanese (Watanabe 1992)

- (20) a. ?? John-wa [Mary -ga nani -o katta ka dooka] siritagatte-iru no?
 John TOP Mary NOM what ACC bought whether know-want Q
 ‘What does John want to know whether Mary bought?’
- b. John-wa [Mary-ga nani -o katta ka dooka] dare -ni tazuneta no?
 John TOP Mary NOM what ACC bought whether who DAT asked Q
 ‘Who did John ask whether Mary bought what?’

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c.)

- (21) a. * Koja kniga_i otreče senatorăt [mălvata će iska da zabrani t_i]?
 which book denied the-senator the-rumor that wanted to ban
 ‘Which book did the senator deny the rumor that he wanted to ban?’
- b. ? Koj senator koja kniga_i otreče [mălvata će iska da zabrani t_i]?
 which senator which book denied the-rumor that wanted to ban
 ‘Which senator denied the rumor that he wanted to ban which book?’

In (20) and (21) a *wh*-extraction which is ill-formed in isolation (as shown in the (a) cases) is remedied by the addition of another *wh*-word outside the island (as shown in the (b) cases). The relevant difference between (19a) and (19b) thus seems to have to do not with a distinction between overt and covert movement but with a phenomenon of the general type under discussion here; in (19b), a well-formed dependency is able to “help out” an ill-formed dependency.

For purposes of the following discussion, it will be useful to separate the constraint responsible for Subjacency and for Huang’s (1982) Condition on Extraction Domains (CED) into two parts:

- (22) a. An attractor must trigger a dependency consisting of a (possibly singleton) set of well-formed links which share an index.
- b. A well-formed link must consist of a head α and a tail β such that there is no γ , γ a member of some set of barriers (including wh-islands, complex noun phrases, adjuncts, and subjects), such that α c-commands γ and γ dominates β .

The distinction above between links and dependencies will become important in the case of parasitic gap constructions such as (23):

- (23) What did you file t [without reading t']?

Here the matrix [+wh] complementizer will be said to enter into a single wh-dependency which consists of two links, one linking Spec CP to t and the other linking Spec CP to t'.

The notion of “index” is employed here to distinguish parasitic gap constructions from cases of multiple interrogation; the links in (23) share an index, while multiple wh-words do not.

The PMC may apply to either half of this constraint; that is, we expect to see cases in which well-formed wh-dependencies rescue ill-formed wh-dependencies, as well as cases in which well-formed links of a dependency rescue ill-formed links. The behavior of the two halves of the constraint will be rather different, however, as we will see.

2.5.1 Dependencies

Consider again the facts in (24-26) and the constraint (27):

- (24) a. *What_j do you wonder who bought t_j?
- b. Who_j t_i wonders who bought what_j?

Japanese (Watanabe 1992)

- (25) a. ?? John-wa [Mary -ga nani -o katta ka dooka] siritagatte-iru no?
 John TOP Mary NOM what ACC bought whether know-want Q
 ‘What does John want to know whether Mary bought?’
- b. John-wa [Mary-ga nani -o katta ka dooka] dare -ni tazuneta no?
 John TOP Mary NOM what ACC bought whether who DAT asked Q
 ‘Who did John ask whether Mary bought what?’

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c.)

- (26) a. * Koj kniga; otreče senatorăt [mălvata će iska da zabrani t_j]?
 which book denied the-senator the-rumor that wanted to ban
 ‘Which book did the senator deny the rumor that he wanted to ban?’
- b. ? Koj senator koja kniga; otreče [mălvata će iska da zabrani t_j]?
 which senator which book denied the-rumor that wanted to ban
 ‘Which senator denied the rumor that he wanted to ban which book?’
- (27) a. An attractor must trigger a dependency consisting of a (possibly singleton) set of well-formed links which share an index.
- b. A well-formed link must consist of a head α and a tail β such that there is no γ , γ a member of some set of barriers (including wh-islands, complex noun phrases, adjuncts, and subjects), such that α c-commands γ and γ dominates β .

Here we are considering cases in which the wh-dependencies do all consist of single links. The reasoning is essentially identical to that used in the three cases we have just reviewed. In the (a) cases above, the +wh comp violates Subjacency by being associated with *what_j* across a wh-island. In the (b) cases, the matrix +wh comp obeys Subjacency when it associates with *who_i*. This allows us to ignore the matrix +wh comp when evaluating the

rest of the structure for Subjacency. Next we can evaluate the association between the matrix +wh comp and *what_j* to determine whether it obeys Subjacency. If we ignore the matrix +wh comp, it does; Subjacency, in this approach, is a constraint on +wh complementizers, so this dependency is no longer subject to Subjacency at all. The well-formedness of (24b) follows.

Making this move will apparently require us to adopt a derivational perspective on the grammar (contra, e.g., Brody (1995b), whose proposal for dealing with Subjacency is similar in many ways to the one developed here). In particular, we will apparently need to assume that in cases in which one move triggers the PMC in a way which allows another move to occur, the first move must precede the second move. That is, Subjacency is not simply a constraint on representations; moves are evaluated for obedience to Subjacency as they occur, and if a move violates Subjacency it cannot be saved by a later move which triggers the PMC.

(24b) involves a case in which, according to our standard ways of looking at derivations, the licit move precedes the illicit move, thus triggering the PMC and allowing the illicit move to take place. If Subjacency were purely representational, we would expect to find cases of the opposite kind, in which an overt Subjacency-violating move is licensed by a covert Subjacency-obeying move. This does not appear to be the case:

- (28) a. *Which_i car did John persuade the man who bought t_j to sell the hubcaps?
 b. Who_j t_j persuaded the man who bought which car_j to sell the hubcaps?
 c. *Which_i car did John persuade the man who bought t_j to sell which hubcaps_j?

(28a-b) are parallel to (24a-b). (28a) is a Subjacency violation, in this case a violation of the Complex NP Constraint. In (28b) we can see the now-familiar “saving” effect of a second wh-word. (28c) is the case which shows that Subjacency must be interpreted derivationally; here the Subjacency-violating move will be followed at LF by the Subjacency-obeying move of *which hubcaps*. This is not enough to save the structure,

however; Subjacency is still violated. It looks as though the order in which moves take place is still relevant for Subjacency; licit moves must precede illicit moves in the derivation⁹.

The cases we have discussed thus far in this section have all involved Subjacency, but a similar case can be constructed for CED islands. As with Subjacency islands, an additional *wh*-word outside the island which moves to the same [+*wh*] complementizer remedies the violation:

(29) a. *Who_i does John want to say a prayer [before we interview t_j]?
 b. Who_i t_j wants to say a prayer [before we interview who]?
Bulgarian (Roumyana Izvorski, p.c.)

(30) a. *Kogoj iska Ivan da kaže molitva [predi da intervjuirame t_j]?
 whom wants Ivan to say prayer before we-interview

b. Koi_j kogoj t_j iska da kaže molitva [predi da intervjuirame t_j]?
 who whom wants to say prayer before we-interview

Again, the Bulgarian facts show that the phenomenon has nothing to do with the overt/covert distinction; the presence of a well-formed dependency is what remedies the CED violation of the other dependency.

2.5.2 Links; Connectedness

Now let us turn to the behavior of the PMC as it applies to links of a single *wh*-dependency. Here we expect the outcome to be somewhat different. Consider again the constraints in (27), repeated as (31):

⁹ Note that there may be a difference between Subjacency and Reflexivity in this regard. (i) and (ii) are both well-formed (Reinhart and Reuland 1993, 668):

(i) Henk wees zich aan zichzelf toe

Henk assigned self to self-self

(ii) Henk wees zichzelf aan zich toe

Henk assigned self-self to self

We must apparently assume either that the constraint driving the Reflexivity effects is purely representational, or that the formation of the dependencies to which the constraint applies may occur in any order (these accounts are identical, as far as I can see, for the facts dealt with here).

- (31) a. An attractor must trigger a dependency consisting of a (possibly singleton) set of well-formed links which share an index.
- b. A well-formed link must consist of a head α and a tail β such that there is no γ , γ a member of some set of barriers (including wh-islands, complex noun phrases, adjuncts, and subjects), such that α c-commands γ and γ dominates β .

These constraints, like any constraints, may be thought of as consisting of two parts; they apply to a certain set of structures, and state a well-formedness requirement for those structures. (31a), for instance, applies to [+wh] complementizers and their wh-dependencies, while (31b) applies to links in a wh-dependency. (31a) requires the dependencies to which it applies to consist entirely of well-formed links. (31b) requires links to consist of a head and a tail in a certain relation in order to be well-formed.

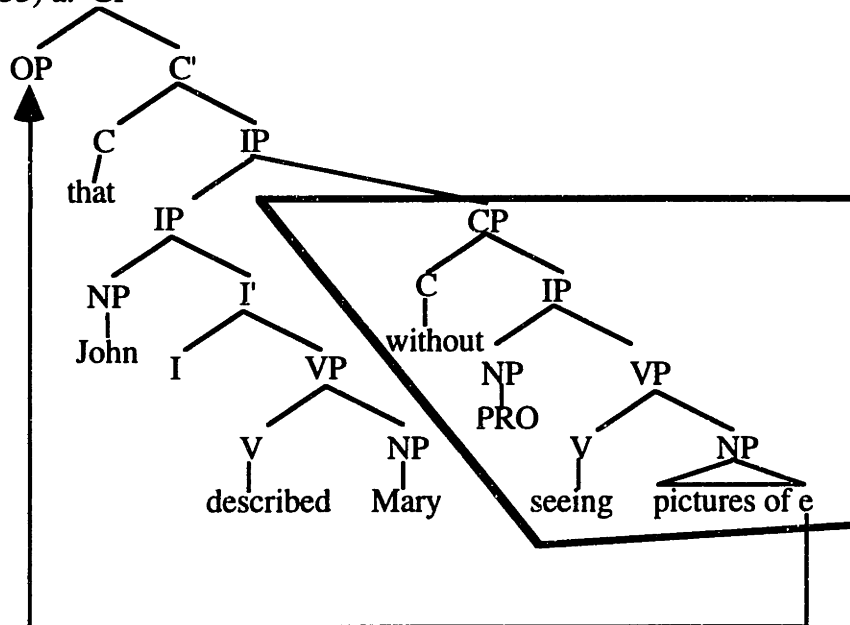
The effect of the PMC in the cases we have surveyed thus far has been to render a structure unrecognizable as belonging to the class of structures to which a constraint belongs. In the previous section, for instance, the PMC was triggered by one wh-dependency to render the [+wh] complementizer invisible for purposes of the constraint. This prevented the constraint from applying to that complementizer for the rest of the derivation, since the constraint in question only applies to structures containing [+wh] complementizers.

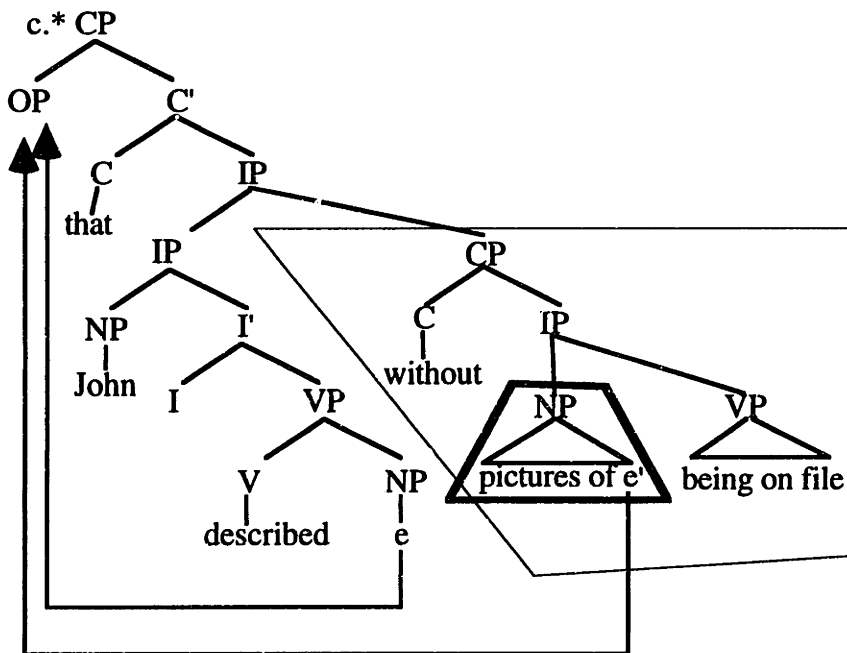
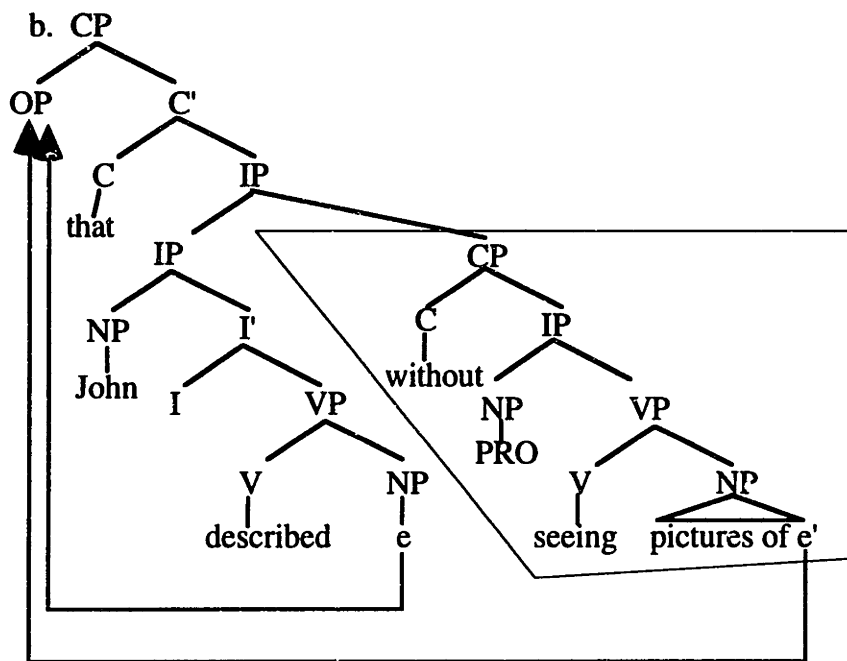
We should also expect to see cases in which the PMC grants invisibility to elements which cause a structure to violate the well-formedness requirements imposed by a constraint. This would leave the structure recognizable as one of the type to which the constraint applies, but would make structures which are in fact ill-formed appear well-formed. Kayne's (1983a, 1983b) Connectedness facts will be a case of this, in the theory developed here. Kayne's insight was that a CED-obeying dependency can eliminate the effect of a single CED island lying along its path. This is shown in (32):

- (32) a. *the person Op_i that John described Mary[without examining any pictures of e_j]
 b. ?the person Op_i that John described e_i [without examining any pictures of e'_i]
 c. *the person Op_i that John described e_i [without [any pictures of e'_i] being on file]

In (32b), the link between OP and e' crosses one CED island (namely, the adjunct CP *without examining...*), while in (32c) the link crosses two CED islands (the adjunct CP and its subject *any pictures of*). The contrasts in (32) follow from the theory developed here, as can be seen from the simplified trees in (33), showing the structure of the relative clauses in (33):

(33) a. *CP





(33a) shows an ordinary CED violation; the link between OP_i and e_j crosses an ungoverned XP boundary (namely, the adjunct CP) and is rendered ill-formed. In (33b), the link between OP_i and e_j obeys the CED. The CED makes reference to the head and tail

of a link, and also to XPs of a certain type between them; thus, we can ignore all of these for purposes of determining whether other dependencies obey the CED. The link between OP_i and e'_i has an XP which qualifies as a barrier (that of the adjunct *without seeing...*) intervening along its path. However, this XP is one of the elements that we are now allowed to ignore for purposes of computing violations of the CED. Thus, the link between OP and e'_i is well-formed, since the only ungoverned XP boundary intervening between the head and tail of the chain can be ignored via the PMC.

In (33c), on the other hand, OP_i and e'_i are separated not only by an adjunct but by a subject island; e'_i is contained in the subject of the adjunct clause. We can ignore the adjunct, as before, but the PMC does not allow us to ignore the subject; not being on the path of the link between OP_i and e_i , it is irrelevant for checking the well-formedness of the link between OP_i and e'_i . The ill-formedness of (33c) follows.

Examples very reminiscent of Kayne's (1983a, 1983b) Connectedness cases can be constructed with islands of the traditional Subjacency type, as well (see Chomsky (1986) for some discussion):

- (34) a. *Who_i did John ask Mary [whether he should invite t_i]?
 b. ?Who_i did John ask t_i [whether he should invite t'_i]?
 c. *Who_i did John ask t_i [whether he should find out [who would invite t'_i]]?

Here the reasoning is the same as for the classic Connectedness cases. The link between *who* and its trace in (34a) violates a *wh*-island, and the dependency consisting of this link is therefore not well-formed. In (34b), the link between *who* and t_i is well-formed, and thus triggers the PMC with respect to the *wh*-island along its path; this makes the link between *who* and t'_i well-formed, since the only island intervening between the head and tail of the link has been made invisible. In (34c), the embedded *wh*-island [*who would*

invite t'j] is not along the path of the link between *who* and t_i , and the link between *who* and t'_j is therefore ill-formed¹⁰.

An interesting fact about Connectedness effects is that the set of elements which a well-formed link allows us to ignore is the same as the set of elements which are potentially barriers for Shortest Move; namely, the set of elements along the “path” of the dependency in question. This theory allows us to state this correspondence directly; the elements along the path of the link are the ones which are “relevant” in the technical sense developed here.

2.5.3 Interlude: Subjacency/CED vs. Connectedness

The above account divided the constraint responsible for Subjacency and CED effects into two parts, repeated below as (35):

- (35) a. An attractor must trigger a dependency consisting of a (possibly singleton) set of well-formed links which share an index.
- b. A well-formed link must consist of a head α and a tail β such that there is no γ , γ a member of some set of barriers (including *wh*-islands, complex noun phrases, adjuncts, and subjects), such that α c-commands γ and γ dominates β .

The PMC might in principle be expected to be able to apply to either part of the constraint. I claimed above that this was in fact the case; the PMC applies to (35a) to give the “saving” effect of an additional *wh*-word in (36), and to (35b) to give the “saving” effect of an additional gap in (37):


- (36) a. *What do you wonder [who bought t]?
 b. Who wonders [who bought what]?



¹⁰ Another possibility, following Chomsky 1986, would be to postulate a distinct operator inside the lower clause, which must obey Subjacency separately; obviously, this would involve relinquishing the account developed here of the Connectedness effects. See section 2.6.2.2 for a similar problem raised by Shortest Move, along with an argument against an approach which postulates distinct operators.

- (37) a. *What did John read the catalog [before buying t]?
 b. What did John read t [before buying t]?

The phenomena in (36) and (37) are distinguished in a number of ways which follow from the different properties of the constraints in (35). The additional-wh-effect in (36) can save a dependency which crosses arbitrarily many barriers:

- (38) a. *What_{tj} did you wonder [who asked Bill [who bought t_j]]?
 b. Who_i t_i wondered [who asked Bill [who bought what_{tj}]]?
- 

The Connectedness facts in (37), by contrast, can only involve a single barrier, as we saw before¹¹:

- (39) a. *the person Op_i that John described Mary[without examining any pictures of e_i]
 b. ?the person Op_i that John described e_i [without examining any pictures of e'_i]
 c. *the person Op_i that John described e_i [without [any pictures of e'_i] being on file]

Similarly, Connectedness phenomena are sensitive to the placement of the barrier, which must be along the extraction path of the well-formed link:

- (40) a. What did you file t [before I could read t]?
 b. Who t filed the paper [before you could talk to t]?

This is not the case, however, with the additional-wh effect for Subjacency or the CED:

¹¹ Brody (1995b) claims that Subjacency and Connectedness do not differ in this regard, on the basis of sentences like (i) (from Brody 1995b, 56):

(i) *Who_i t_i was against [proposals to leave [without waiting for who]]?

(i) involves an LF-moved wh-word which is embedded in both a complex NP and an adjunct island. Brody concludes from this that in both the Connectedness and the Subjacency cases, only a single island can be rendered transparent by a well-formed dependency. His account thus has nothing to say about the well-formedness of (39b). It is not clear to me (or to a number of English speakers I have consulted) that (i) is actually ill-formed, either. If there is a grammaticality contrast between (i) and (39b), this theory has nothing straightforward to say about it.

(41) a. Who should I talk to *t* about [what to read about what]?



b. Who *t* asked about [what to read about what]?



In the terms developed here, the additional-wh effects involve ignoring a [+wh] complementizer, while the Connectedness effects involve ignoring an island. (35a) is a constraint on [+wh] complementizers and the wh-dependencies they trigger. Once a [+wh] complementizer has triggered a well-formed dependency, the PMC makes it immune to (35a) for the rest of the derivation. The placement of the island is therefore irrelevant.

(35b), on the other hand, is a condition on well-formed links. Here we must prevent the line of reasoning used for the additional-wh effect (and for the Reflexivity, WCO, and VP-ellipsis facts) from applying; we do not want it to be the case that a well-formed link can render any other links with the same head well-formed. The constraint in (37b) achieves this result; it requires links to consist of a head and a tail in a certain configuration. Ignoring the head of a link, then, will simply result in a headless link, which is still an ill-formed object by this constraint. The only way to remedy an ill-formed link is to ignore the island that triggers the ill-formedness. Thus, there must be a well-formed link for which the island counts as “relevant” in the technical sense introduced in (8); that is, the island must be “along the path” of a well-formed link. The placement of the island with respect to the well-formed link is therefore relevant, and Connectedness cases will never involve crossing multiple islands, since only one island can be along the path of a well-formed dependency¹².

¹² The PMC thus derives the anti-c-command requirement on parasitic gaps posited by Chomsky (1982); this is a reflex of the requirement that the island be along the path of a well-formed link. Examples like (i) are therefore problematic for this proposal:

(i) Who did John tell *t* [that we were planning to hire *t*]?

(i) is apparently a case of a parasitic gap with no islands involved at all. If the anti-c-command requirement truly follows from the properties of the PMC as a way of circumventing islands, parasitic gaps without islands should be free of the anti-c-command requirement. This seems false; (i) is clearly better than (ii):

Thus, the PMC will have very different effects on constraints like the one on well-formed links, which force dependent elements to have binders of a certain kind, than it will on constraints like Subjacency, which constrain binders, or on constraints like the one responsible for the Dutch Reflexivity effects discussed in section 2.1, which force dependent elements not to have binders of a certain kind. It is not uncommon for binders to enter into multiple dependencies, and the PMC will allow constraints on binders to be violated by all but a single dependency headed by a given binder. On the other hand, dependent elements typically do not enter into more than one dependency of a given type, and the effects of the PMC on locality constraints on dependent elements will therefore be less frequently seen; the PMC will only have an effect if it can render invisible everything which causes a particular dependent element to violate the constraint. In some cases, of course, this will mean the PMC will have no effect at all. For instance, it is reasonable to assume that local anaphors are subject to a requirement that they be locally bound. If this requirement were like the constraint on Dutch anaphora discussed in section 2.1, we would expect the PMC to operate on it in a similar way, rendering (42c) well-formed:

- (42) a. John_i told himself_i [that Mary liked Susan].
 b. * John_i told Susan [that Mary liked himself_i].
 c. * John_i told himself_i [that Mary liked himself_i].

In (42c), the second instance of *himself* is bound by an antecedent which has already entered into a well-formed binding relation with the first instance of *himself*. If the requirement that *himself* be locally bound were a constraint on the binder, like Subjacency, we would expect (42c) to be well-formed. On the other hand, if the constraint in question is, like the constraint on well-formed links, a constraint on dependent elements of a certain

(ii) *Who_t said [that we should plan to hire t]?

I will have to leave this problem for future research. One solution which suggests itself would be to say that the tensed embedded clauses in these examples are in fact islands for the type of across-the-board movement involved in parasitic gaps; this solution demands a detailed account of what makes different structures islands for different types of extractions, and I am unable to supply such an account at this point.

kind which requires them to have a sufficiently local binder, then the ill-formedness of (42c) is expected. Ignoring *John* in (42c) will not improve the status of the dependency between *John* and the second instance of *himself*; the latter will still lack a sufficiently local binder.

Of course, the constraint on well-formed links and the requirement that certain anaphors be locally bound presumably make reference to different sets of opacity-inducing elements. One result of this difference is that the PMC apparently never interacts with the locality condition on local anaphors at all; there is no equivalent of Connectedness in this domain. The PMC interacts with the constraint on well-formed links to bring about Connectedness effects in cases in which one of the elements which can render a dependency ill-formed (an ungoverned XP, in this case) lies along the path of a well-formed dependency. In the theory being developed here, the absence of Connectedness effects for anaphora would follow from the fact that the elements which render an anaphoric dependency insufficiently local can never lie along the path of a well-formed anaphoric dependency. For instance, the relevant opacity-inducing element in (42b-c) might be *Mary* (following Chomsky 1973 and much subsequent work). *Mary* does not lie along the path between *John* and the first instance of *himself*; if it did, that dependency would be ill-formed and would still fail to trigger the PMC¹³. Thus, the locality constraint on local anaphors fails to show PMC effects at all. However, we can derive this result from properties of the relevant constraint, rather than simply stipulating it.

The PMC apparently needs to be able to distinguish between two classes of constraints. In one class, which includes the constraints discussed in sections 2.1-2.3, the constraint is on dependencies of a certain kind, and the PMC can circumvent the effects of the constraint by allowing us to ignore parts of the structure which identify the dependency

¹³If this approach is on the right track, it will place certain constraints on our theories about what the binding domain for a local anaphor is. If, for instance, the embedded CP in (42c) c-commands *himself*, then this CP cannot be the binding domain, or we would expect the PMC to save the structure.

as a member of the relevant kind. In the second class, the constraint is on dependent elements of a certain kind, and requires them to be associated with appropriate licensors; constraints of this type include the requirement that anaphors be bound, as well as the one responsible for Connectedness effects (section 2.5). Here we have to limit the power of the PMC to remedy violations; triggering the PMC with respect to the licensor cannot make structures legitimate. The version of the PMC in (7) is intended to be capable of making this distinction; in the second class of cases, ignoring the licensor will not improve the structure, because the constraint in question demands a licensor¹⁴. At the same time, as we have seen, it is in principle possible to remedy violations of such constraints using the PMC, as in the Connectedness case. In other words, it is not simply that the PMC arbitrarily fails to apply to certain constraints.

2.6 Attractors and Islands

In the previous section we saw two ways of obviating the effects of an island. One is to render the attractor invulnerable to the effects of the island, by having it participate in one well-formed dependency before the dependency which is rendered ill-formed by the presence of the island is formed. This is the strategy involved in the “Subjacency tax” phenomena which were one focus of chapters 2 and 3. A second way is to render the island harmless by having it be along the path of a well-formed dependency; this is the strategy responsible for Connectedness effects. We have seen that these strategies differ in a number of ways, one of them being sensitivity to the placement of the island. In this

¹⁴ In principle, it might be possible to distinguish between these two classes of cases, not by assuming two different classes of constraints, but by assuming general constraints which apply to different structures in different ways. Distinguishing between an approach of this kind and the one developed in this paper would not be straightforward; we would need to find a case in which the PMC interacts with a constraint differently depending on what kind of structure is being constrained, and try to show that the differences in structure can account for the differences in behavior of the PMC. One such case might be the distinction between head-movement and XP-movement with respect to Pesetsky's (1982) Path Containment Condition (discussed in section 2.6.2.2); PCC effects are apparently found with XP-movement, but not with head-movement (a fact for which I have no account). I will not try to develop this proposal further here, however.

section we will review a number of other phenomena which may be understood as involving the PMC rendering either attractors or islands invisible.

One question we will be attempting to answer, in so doing, has to do with the requirements placed on the timing of the derivation by the two strategies. We have seen that the Subjacency tax strategy imposes a particular order of operations on the derivation; the well-formed move must precede the ill-formed move. It is difficult to determine whether any such requirement exists in the Connectedness cases, since only a single operator is visible (whether more than one operator is actually involved is unclear; see section 2.6.2.2.2 for an argument that parasitic gap constructions are a form of across-the-board movement). One type of case we will be looking for, then, is a clear case of multiple operator-movement in which one operator renders an island transparent for movement by another operator.

2.6.1 Islands

In this section we will investigate cases in which the PMC appears to render islands transparent for extraction. These cases will therefore involve mechanisms which are formally identical to that involved in licensing English parasitic gaps, but will differ in that multiple operators will demonstrably be involved. We will therefore be able to investigate more fully the requirements placed on the timing of the derivation by this strategy.

2.6.1.1 Additional-wh effects in Bulgarian and Japanese

In section 1 of chapter 2 we noted that in languages like Bulgarian and Japanese, as in English, the effects of a Subjacency island may be obviated by the presence of a well-formed extraction. In some cases, this effect demonstrably imposes a requirement that the well-formed move precede the ill-formed move, as it does in English. On the theory developed here, these are cases in which the well-formed move triggers the PMC with respect to the attractor, rendering it impervious to the effects of the relevant constraint and allowing it to attract the second wh-word from a position from which extraction would

ordinarily be impossible. In (43-44), the (a) examples involve extraction from some island, which is unsurprisingly ill-formed. The (c) examples contain an additional well-formed extraction from a position c-commanding the relevant island; Superiority forces the well-formed extraction to precede the ill-formed extraction in these cases, and the sentences are better than the (a) examples, as the PMC predicts. The interesting cases are the (b) examples; here extraction takes place from a position which does not c-command the island. The Bulgarian (44b) demonstrably involves well-formed movement which follows ill-formed movement, and the result is ill-formed, as we expect. In the Japanese (43b), since movement is covert, it is impossible to determine whether well-formed movement precedes ill formed movement in the derivation. Speaker's intuitions about the grammaticality of (43b) seem to differ; see chapter 3, section 1 for some discussion of this fact.

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (43) a.*John -ga [Bill -ga [Mary -ga nani -o katta ka dooka] sitteita to] itta no?

John NOM Bill NOM Mary NOM what ACC bought whether knows that said Q

'What did John say that Bill knows whether Mary bought?'

- b.*/?John -ga [Bill -ga [Mary -ga nani -o katta ka dooka]

John NOM Bill NOM Mary NOM what ACC bought whether

sitteita to] dare -ni itta no?

knows that who DAT said Q

'Who did John tell that Bill knows whether Mary bought what?'

- c.?John -ga dare-ni [Bill -ga [Mary -ga nani -o katta ka dooka]

John NOM who DAT Bill NOM Mary NOM what ACC bought whether

sitteita to] itta no?

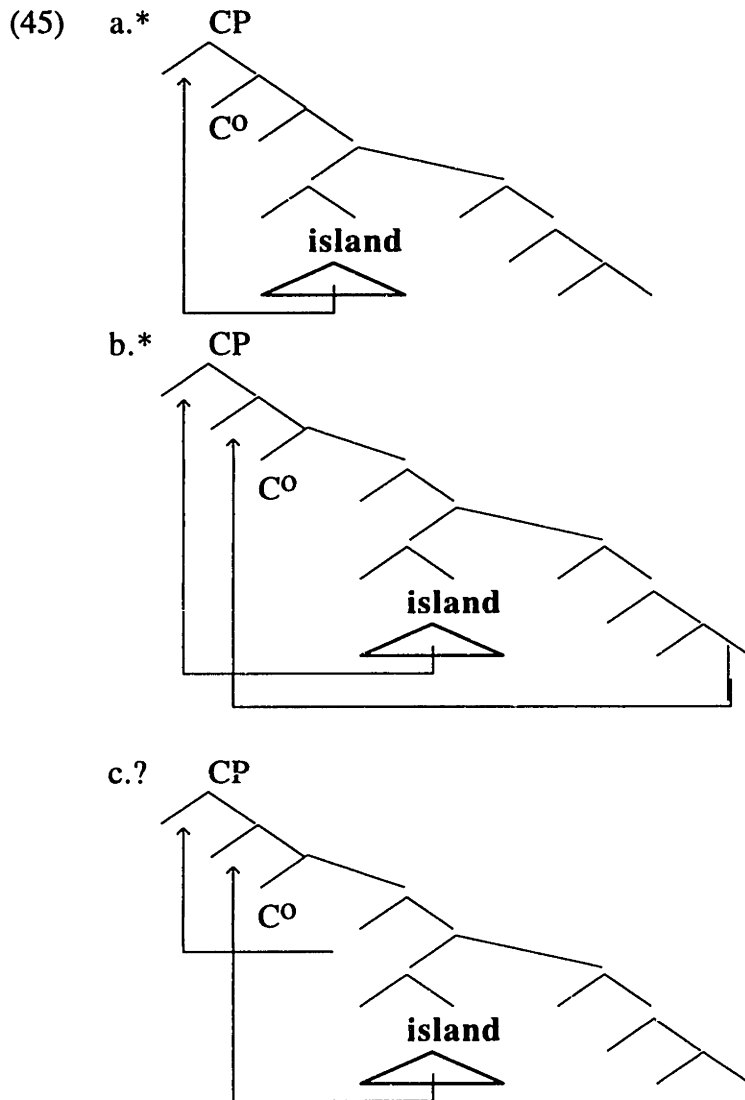
knows that said Q

'Who did John tell that Bill knows whether Mary bought what?'

Bulgarian (Roumyana Izvorski, p.c.)

- (44) a. *Kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat [mălvata,
 what tells this official to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘What does this official tell journalists who are investigating the rumor that the
 government wants to ban that the communists have deceived their editors?’
- b. *Kakvoj kogo_k kazva tozi služitel na [žurnalistite, kojto razsledvat [mălvata,
 what who tells this official to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili t_k?
 that the-communists AUX deceived
 ‘What does this official tell journalists who are investigating the rumor that the
 government wants to ban that the communists have deceived who?’
- c. ??Koji kakvoj kazva t_j na [žurnalistite, kojto razsledvat [mălvata,
 who what tells to the-journalists who investigate the-rumor
 če pravitelstvoto iska da zabrani tj]],
 that the-government wants to ban
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘Who tells journalists who are investigating the rumor that the government wants to
 ban what that the communists have deceived their editors?’

In (43-44), then, the properties of the additional-wh effect seem to be roughly like their English counterpart; well-formed movement must precede ill-formed movement in the derivation, at least in cases where the PMC is affecting an attractor. The facts in (43-44) may be given the simplified diagrams in (45):



(43-44) have in common the property that the attractor is the only thing the PMC is in a position to affect. The islands in these examples are all embedded in such a way that they are not along the path of any of the well-formed wh-movements involved. If we consider cases in which the potentially offending island is along the path of a well-formed

wh-movement, the facts change. The equivalents of the (b) sentences above become well-formed, even better (for some speakers, at any rate) than the (c) sentences:

Japanese (adapted from Watanabe 1992, 270-271)

- (46) a. *John-wa [Mary -ga nani -o katta ka dooka] Tom -ni tazuneta no?

John TOP Mary NOM what ACC bought whether Tom DAT asked Q

‘What did John ask Tom whether Mary bought?’

- b. John-wa [Mary -ga nani -o katta ka dooka] dare -ni tazuneta no?

John TOP Mary NOM what ACC bought whether who DAT asked Q

‘Who did John ask whether Mary bought what?’

- c. ?John-wa dare -ni [Mary -ga nani -o katta ka dooka] tazuneta no?

John TOP who DAT Mary NOM what ACC bought whether asked Q

‘Who did John ask whether Mary bought what?’

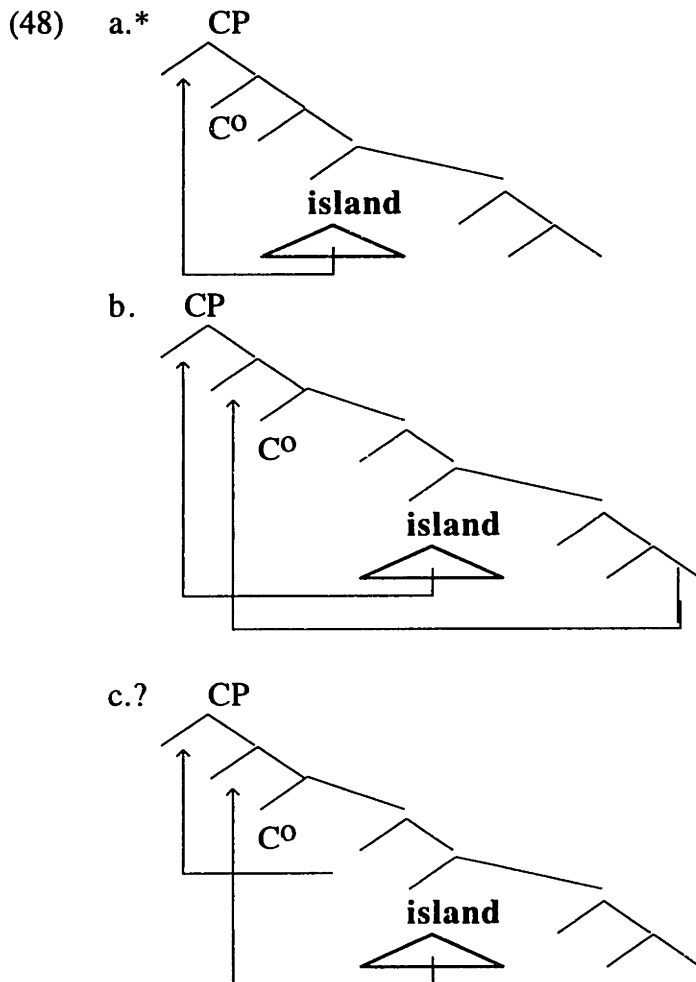
Bulgarian (Roumyana Izvorski, p.c.)

- (47) a. *Kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat tj],
 what tells this official to the-journalists who investigate
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘What does this official tell journalists who are investigating that the communists
 have deceived their editors?’
- b. ?Kakvoj kogo_k kazva tozi služitel na [žurnalistite, kojto razsledvat tj],
 what who tells this official to the-journalists who investigate
 če komunistite sa zabludili t_k?
 that the-communists AUX deceived
 ‘What does this official tell journalists who are investigating that the communists
 have deceived who?’
- c. ??Koji kakvoj kazva t_j na [žurnalistite, kojto razsledvat tj],
 who what tells to the-journalists who investigate
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘Who tells journalists who are investigating what that the communists have
 deceived their editors?’

Here the resemblance between additional-wh effects in Japanese and Bulgarian and their English counterparts breaks down. In the (b) examples above, the well-formed extraction takes place from a position c-commanded by the offending island, and in the Bulgarian case this extraction appears to follow the ill-formed extraction in the derivation; at least, the landing site of the well-formed movement is lower than that of the ill-formed movement, and it has apparently been the case so far that movement to multiple specifiers involves

landing in successively lower specifiers (see chapter 3). Thus, in the (b) examples well-formed movement appears to follow ill-formed movement in the derivation. Still, there is an additional-wh effect in these cases, and in fact the (b) examples are preferable to the (c) examples, where well-formed wh-movement precedes ill-formed movement, as desired.

These cases may be diagrammed as in (48):



The puzzling fact here is the comparative well-formedness of the structure in (48b). Just in case an island is along the path of a well-formed wh-movement, it appears, the restriction on the relative ordering of well-formed and ill-formed movement goes away. This is precisely the case in which the PMC ought to be able to obviate the effects of the offending island, rather than rendering the attractor immune to the constraint. In other words, the

structure in (48b) is well-formed for the same reason that the Connectedness cases are well-formed; the PMC renders the island transparent (and for some reason, rendering the island transparent, as in (48b), improves the sentence more than triggering the PMC with respect to the attractor, as in (45c) and (48c); I have no account of why this should be so). The contrast between (45b) and (48b), then, follows from the same structural conditions on the effects of the PMC that are responsible for the conditions on parasitic gap licensing in English; only islands which are along the path of a well-formed dependency may be rendered transparent. For ease of reference, then, I will refer to this strategy as “parasitic wh-movement”. We appear to have evidence that the island-obviating strategy involved in parasitic wh-movement, unlike the strategy which affects the relevant attractor, imposes no conditions on the order of operations.

In fact, this is not quite right. Parasitic wh-movement does in fact impose some kind of requirement on the timing of the wh-movements involved, as can be seen by the absence of this effect in English:

- (49) a. *Which car did you persuade [the man who bought t] to sell the hubcaps?
 b. *Which car did you persuade [the man who bought t] to sell which hubcaps?

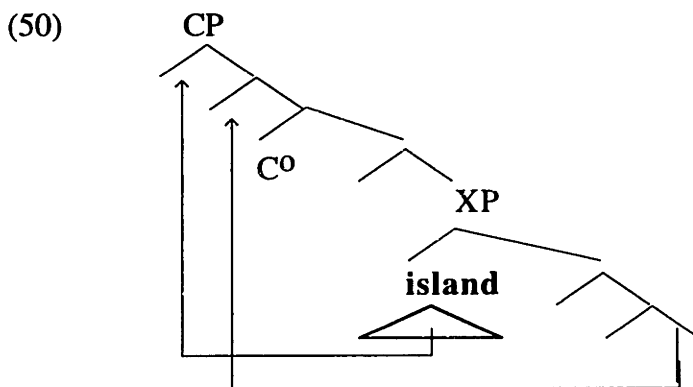
(49b) is no better than (49a), despite the fact that (49b) is an instance of the structure in (48b); there is a wh-phrase in situ which will undergo well-formed covert movement in a path which crosses the offending island. Note that covert movement is in fact capable of making islands of this type transparent, as the Japanese facts show. Parasitic wh-movement, then, is apparently only available in languages such as Bulgarian and Japanese, in which all wh-movement takes place on a single level, and not in languages like English. Why should this be so?

Suppose we assume an approach to wh-movement like that defended in Chomsky (1986), Saito and Fukui (1996), Agbayani (1997), and Fox (1997), among others, in which wh-movement (and possibly all movement) involves successive-cyclic adjunction to

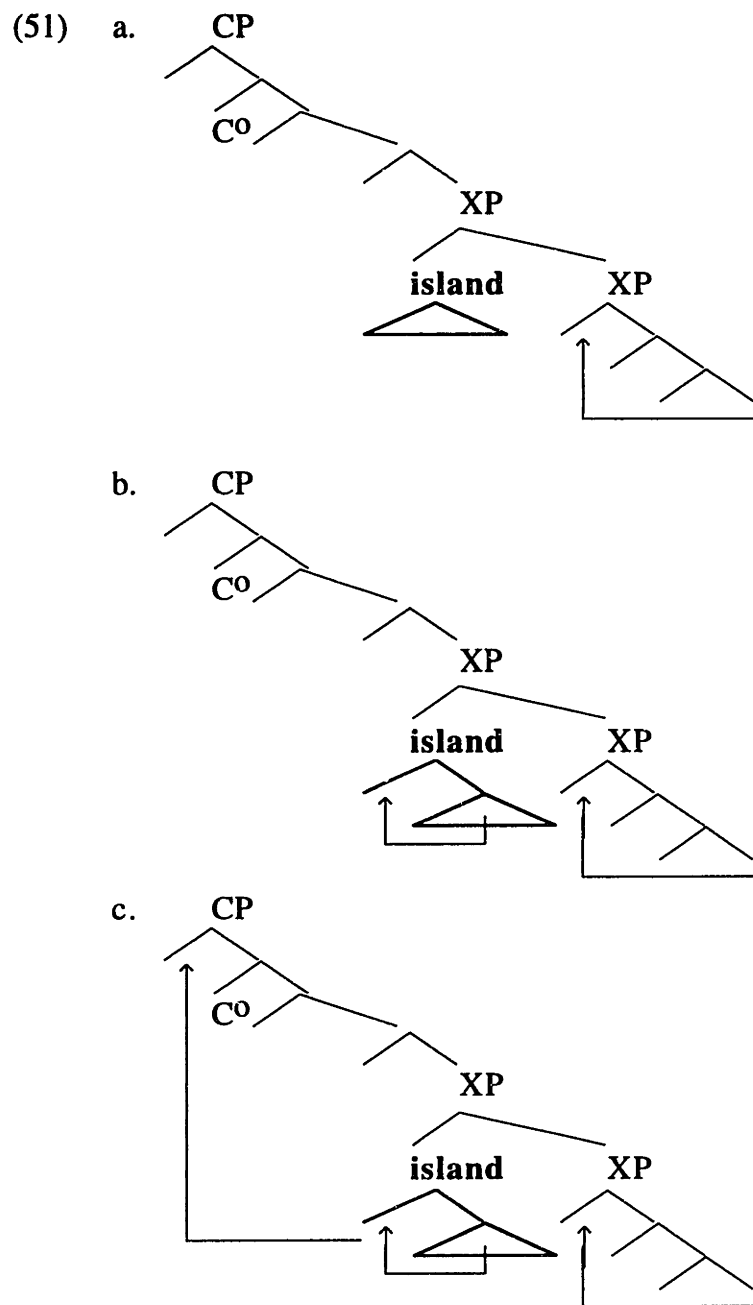
some or all of the maximal projections which dominate the extraction site and are c-commanded by the eventual landing site. The reasons for wh-movement to proceed in this manner will be unimportant for our purposes; they may, as the references above suggest, have to do with locality constraints. Crucially, this approach to wh-movement would allow movement of one wh-word can “stop” midway along the path to its eventual destination, while another wh-movement proceeds. In cases in which one wh-word c-commands another, a derivation of this type will presumably be impossible; Shortest will force attraction of the closest wh-word.

Suppose we consider, however, the cases of parasitic wh-movement in Bulgarian and Japanese. Here, as we saw before, the extraction sites are not in a c-command relation to each other. There is therefore no reason to attract the wh-words in a particular order. In fact, the wh-words could move in an alternating fashion, each moving for several steps, followed by movement of the other, until they entered into a c-command relation with each other, at which point Shortest would force a particular order of movement. Such a derivation seems counterintuitive, but nothing seems to rule it out.

Consider, then, a structure for parasitic wh-movement, like the one in (48b), repeated as (50):



The derivation of (50) might proceed as follows (ignoring irrelevant steps):



In (52), the movement corresponding to (51a) does not take place until the covert syntax, after the ill-formed overt *wh*-movement from inside the island; the island is therefore not transparent at the point in the derivation at which extraction takes place from inside it.

Thus, the behavior of the PMC is uniform with respect to the restrictions it places on the properties of the derivation; the PMC must be triggered before a potentially offending operation is performed. In the cases in which the PMC renders an attractor impervious to a constraint, this has the effect that one *wh*-word must land in a position associated with the attractor before the other does, since it is the act of landing which triggers the PMC. In the second class of cases investigated here, in which the PMC renders an island transparent for movement, the requirement is simply that a well-formed movement must cross the island before the ill-formed movement does. On a theory in which *wh*-movements consist of a number of successive-cyclic adjunctions, this is consistent with the potentially ill-formed movement landing in a position associated with the attractor before the well-formed movement does so; the properties of the attractor are not relevant in this case, since it is the island which is being rendered transparent.

In the next section I will investigate another class of cases which appear to involve the PMC obviating the effect of an island. We will see that the requirements placed by the PMC on the derivation may in fact be slightly different than has been suggested thus far.

2.6.1.2 Inside jobs; additional additional-*wh* effects

Certain island effects in Japanese appear only for extraction of *naze* 'why' (this is similar to the situation described by Huang (1982) in Chinese, in which only adjuncts are subject to the CED):

Japanese (Saito 1996, 22)

- (53) a. John -wa [nani -o katta hito -o] sagasiteru no?

John TOP what ACC bought person ACC looking-for Q

‘What is John looking for [the person that bought t]?’

- b. *John-wa [naze sono hon -o katta] hito -o sagasiteru no?

John TOP why that book ACC bought person ACC looking-for Q

‘Why is John looking for [the person that bought that book t]?’

As discussed in Saito (1994a, 1994b, 1996), Sohn (1994), and Grewendorf and Sabel (1996), the presence of an argument wh-word inside the offending island improves the status of examples like (53b):

Japanese (Saito 1996, 26)

- (54) ? John-wa [nani -o naze katta] hito -o sagasiteru no?

John TOP what ACC why bought person ACC looking-for Q

‘What is John looking for [the person that bought t why]?’

This looks like a PMC effect; well-formed extraction improves the status of extraction which would be ill-formed in isolation. To make the contrast between (53b) and (54) follow from the PMC, the relevant constraint will have to be one which constrains both arguments and adjuncts but which arguments always pass; it might be roughly stated as in (55):

- (55) In Japanese, if wh-movement takes place out of an island, the extracted wh-word must be an argument.

Obviously we will ultimately want to derive (55) from more general considerations, but (55) will suffice for our purposes.

Interestingly, (54) compares favorably with (56), in which the additional wh-word is not inside the island:

Japanese (Shigeru Miyagawa, p.c.)

(56) ?? dare -ga [sono hon -o naze katta] hito -o sagasiteru no?

who NOM that book ACC why bought person ACC looking-for Q

‘Who is looking for [the person that bought that book why]?’

(56) is a case in which the PMC can only operate by rendering the relevant attractor immune to the constraint involved, since the island is not along the path of the well-formed extraction. As expected, the PMC does improve the sentence somewhat; (56) is better than (53b). We saw in the last section, however, that (in languages like Japanese and Bulgarian, at least) cases in which the PMC renders islands transparent are more acceptable than ones in which the PMC affects an attractor, for reasons which are still mysterious. We might take the contrast between (54) and (56), then, as evidence that the additional wh-word in (54) is rendering the island transparent by undergoing well-formed wh-movement out of it prior to the potentially ill-formed wh-extraction of *naze*.

(54) and (56) differ in two potentially relevant regards. The two wh-words are clausemates in (54), but not in (56), and the additional wh-word is inside the offending island in (54), but not in (56). On the account being developed here, it is the second of these differences which is important. More complex examples suggest that this is correct:

Japanese (Shigeru Miyagawa, p.c.)

- (57) a. *Taroo-wa [Hanako -ga [sono hon -o **naze** katta to] itta] hito -o
 Taroo TOP Hanako NOM that book ACC why bought that said person ACC
 sagasiteru no?
 looking-for Q
 ‘Why is Taroo looking for the person [that Hanako said [bought that book t]]?’
- b. ?***dare-ga** [Hanako -ga [sono hon -o **naze** katta to] itta] hito -o
 who NOM Hanako NOM that book ACC why bought that said person ACC
 sagasiteru no?
 looking-for Q
 ‘Who is looking for the person [that Hanako said [bought that book why]]?’
- c. ?Taroo-wa [**dare-ga** [sono hon -o **naze** katta to] itta] hito -o
 Taroo TOP who NOM that book ACC why bought that said person ACC
 sagasiteru no?
 looking-for Q
 ‘Who is Taroo looking for the person [that t said [bought that book why]]?’
- d. ?Taroo-wa [Hanako -ga [**nani** -o **naze** katta to] itta] hito -o
 Taroo TOP Hanako NOM what ACC why bought that said person ACC
 sagasiteru no?
 looking-for Q
 ‘What is Taroo looking for the person [that Hanako said [bought t why]]?’

In (57b), as in (54), the two wh-words are not clausemates and the additional wh-word is not inside the island. In (57d), as in (56), the two wh-words are clausemates, and the additional wh-word is inside the island. The example which tells us which of these differences is relevant is (57c), in which the additional wh-word is inside the island but the

wh-words are not clausemates. We can see that the relevant factor is apparently the placement of the additional wh-word with respect to the island, and not the presence of clause boundaries between the wh-words; (57c) is better than (57b). This is what the theory developed here leads us to expect, if the phenomenon under investigation is in fact a case of the PMC rendering islands transparent.

In the last section we saw evidence that when the PMC renders an island transparent (as when it affects an attractor) the well-formed move must trigger the PMC with respect to the potentially offending structure before the potentially ill-formed move can take place. There is additional evidence for this conclusion from properties of the phenomenon under discussion in this section, to which we now turn.

Saito (1994a, 1994b, 1996) points out that the additional wh-word which improves the status of adjunct extraction from an island must c-command the adjunct; adding an argumental wh-word below *naze* 'why' does not improve the sentence:

Japanese (Saito 1994a, 204-5)

- (58) a. ??John-wa [nani -o naze katta] hito -o sagasiteru no?

John TOP what ACC why bought person ACC looking-for Q

'What is John looking for [the person that bought t why]?'

- b. *John-wa [naze nani -o katta] hito -o sagasiteru no?

John TOP why what ACC bought person ACC looking-for Q

'Why is John looking for [the person that bought what t]?'

This is as we expect. Shortest will require that the higher of the two wh-words be attracted first; in (58a), this is the wh-word which can be extracted out of the island, while in (58b) it is the wh-word which cannot be extracted. The contrast thus follows, if we assume that well-formed movement must precede ill-formed movement in the derivation for the PMC to have an effect.

There is a potential confound in (58b), unfortunately, which weakens the force of this argument. As was noted in chapter 2, section 3.3, many speakers of Japanese find sequences of *wh*-words like that in (58b) ill-formed even if no islands are present, a phenomenon known as “Anti-Superiority” in the literature:

Japanese (Saito 1994a, 195)

(59) a. John -ga nani -o naze katta no?

John NOM what ACC why bought Q

‘What did John buy why?’

b. *John-ga naze nani -o katta no?

John NOM why what ACC bought Q

There are speakers of Japanese for whom this contrast is weak or nonexistent, however. Thus, the argument from the data in (58) for a requirement that well-formed movement precede ill-formed movement in the derivation is convincing only to the extent that the contrast in (58) is sharper or more widespread among speakers than that in (59)¹⁵.

Another set of facts which bears on the question of how the PMC interacts with the timing of *wh*-movement comes from Sohn (1994), who points out that overt extraction of arguments out of islands can improve the status of overt extraction of *naze* ‘why’ out of the same island:

¹⁵ Kazuko Yatsushiro (p.c.) informs me that she gets neither the contrast in (58) nor that in (59), and Saito is clearly a speaker who gets both; if these contrasts have roughly an identical distribution across speakers, this argument is indeed in trouble. The judgment in (59) is intended to be contrastive; as noted, speakers differ on the strength of this effect.

Japanese (Sohn 1994, 317)

- (60) a. *Naze_j John-wa [[Mary -ga t_j sono hito -o uttaeta to iu] uwasa]-o kiita no?
 why John TOP Mary NOM the man ACC sued that rumor ACC heard Q
 ‘Why did John hear [the rumor that Mary sued the man t_j]?’
- b. ?Dare -o_j John-wa [[Mary -ga t_j uttaeta to iu] uwasa] -o kiita no?
 who ACC John TOP Mary NOM sued that rumor ACC heard Q
 ‘Who did John hear [the rumor that Mary sued t_j]?’
- c. ?Dare -o_j naze_j John-wa [[Mary -ga t_j t_j uttaeta to iu] uwasa] -o kiita no?
 who ACC why John TOP Mary NOM sued that rumor ACC heard Q
 ‘Who did John hear [the rumor that Mary sued t why]?’
- d. ?Sono hito -o_j naze_j John-wa [[Mary -ga t_j t_j uttaeta to iu] uwasa] -o
 the man ACC why John TOP Mary NOM sued that rumor ACC
 kiita no?
 heard Q
 ‘The man, why did John hear [the rumor that Mary sued t t_j]?’

Thus, long-distance scrambling of an argument out of an island (regardless of whether the argument is a wh-word) renders the island transparent for scrambling of *naze* out of it¹⁶.

This follows straightforwardly from the account given above of the facts for covert movement.

Sohn points out, however, that overt extraction of an argument does not improve the status of covert extraction of *naze*:

¹⁶ As Sohn (1994) points out, the contrast between (60d) and examples like (57a) is an argument that long-distance scrambling of arguments is not freely available in the covert component; if it were, we would expect to be able to save (57a) by covert long-distance scrambling of one of the arguments of the clause from which *naze* is to be extracted.

Japanese (Sohn 1994, 332)

- (61) *Dare -o John -wa [[Mary -ga naze t uttaeta to iu] uwasa] -o kiita no?
 who ACC John TOP Mary NOM why sued that rumor ACC heard Q
 ‘Who did John hear [the rumor that Mary sued t why]?’

The ill-formedness of (61) is surprising, on our view. Well-formed extraction of *dare* ‘who’ clearly precedes ill-formed extraction of *naze* ‘why’, at least on the assumption (which we have made up until now) that overt movement precedes covert movement in the derivation. The well-formed extraction should therefore be able to make the island transparent, allowing later movement of *naze* out of the island.

I have described the PMC as effecting a change in the status of a representation which lasts throughout the derivation; the creation of well-formed dependencies makes parts of the structure immune to certain constraints for the rest of the derivation. Another approach to the PMC would make its effect temporary. We might say, for instance, that an element which is in the process of participating in a well-formed dependency makes parts of the structure which are in a certain local relation to it immune to the constraints it obeys, just while it is in a local relation to those parts of the structure; when it moves on, those parts of the structure become vulnerable again to the constraints.

For most of the cases we have looked at so far, this approach to the PMC will have no effect. In cases in which a binder or an attractor is rendered invisible to a constraint by participating in a well-formed dependency, for instance, the head of the dependency is always in a local relation with the binder or attractor in question at the point in the derivation at which the potentially ill-formed dependency is being evaluated. In cases in which it is an island which is being rendered transparent, on the other hand, a “temporary” PMC will predict that the island in question will only be transparent while the well-formed wh-word is in a local relation to it. Recall that we are crucially assuming an approach to

wh-movement in which wh-words successive-cyclically adjoin to some or all of the maximal projections intervening between the extraction site and the eventual landing site. Thus, it should be possible for a well-formed wh-movement to “stop” in a position adjoined to an island, making the island transparent and allowing another, potentially ill-formed wh-movement to pass through¹⁷.

This will have the desired result for Saito and Sohn’s cases; well-formed extractions will be able to render islands transparent for potentially ill-formed extractions only if the relevant movements are either both overt or both covert, and not if one movement is overt and the other covert. In the latter case, the well-formed overt wh-movement fails to make the island transparent for the ill-formed wh-movement, since it moves on to its eventual destination before the ill-formed movement begins.

Thus, Sohn’s case might be taken to be evidence for a “temporary” PMC, the effects of which would wear off once the moving element had moved beyond the portions of the structure to be rendered invisible for purposes of the constraint. Since this version of the PMC differs from the “permanent” PMC assumed up until now only with respect to Sohn’s case, I will have to leave the ultimate choice between them to future research.

2.6.2 Attractors

In this section we will investigate several cases in which the PMC renders an attractor invisible for purposes of a constraint. We have already seen one such case above, in section 2.5.1 of this chapter; this was the case in which a well-formed wh-movement renders an attractor impervious to the effects of Subjacency, thus allowing a second wh-movement which would violate Subjacency in isolation to be ill-formed. The cases in this

¹⁷ This will require an elaboration of our approach to Superiority effects, which I will not attempt to sketch fully here. We might envision, for instance, an approach in which the c-command relation between the wh-words cannot be disturbed once established by Merge, but which would allow the lower wh-movement to raise to a position just below that of the higher wh-word. The wh-words would then both be able to be in positions adjoined to the island at the same point in the derivation, as long as the higher wh-word adjoined there first.

section will concentrate on the interaction of the PMC with the constraint responsible for forcing movement to be maximally local, discussed in section 7 of chapter 3. This constraint is repeated as (62), along with the definition of the Attract relation which it constrains:

(62) *Shortest*

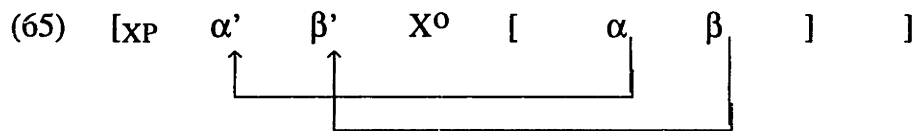
A pair P of elements $\{\alpha, \beta\}$ obeys Shortest iff there is no well-formed pair P' which can be created by substituting γ for either α or β , and the set of nodes c-commanded by one element of P' and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominating the other.

(63) *Attract*

An attractor K attracts a feature F , creating a copy α' of an element α containing F , and Merging α' with K . The relations between α' , K , and F must all obey Shortest.

- (64) a. $\begin{array}{ccccc} & & K & & [\alpha \quad F \quad] \\ & & & & \\ & & & & \end{array}$
 b. $\begin{array}{ccccc} [\alpha' \quad F' \quad] & & K & & [\alpha \quad F \quad] \\ & & & & \\ & & & & \end{array}$

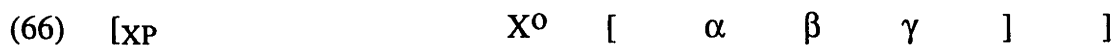
Following Chomsky 1995, I have defined Attract as involving the creation of a copy of some minimal domain α containing the attracted feature F . In some cases, α will be only the set of formal features containing F ; in others, properties of the PF interface will force α to be some larger structure (for instance, an entire XP). Shortest will constrain the relation between the attractor K and the attracted feature F , forcing the attractor to attract the nearest possible feature. Shortest also constrains the relation between F and the copy α' of α , thereby requiring that movement be as short as possible (that is, forbidding movement past an attractor which could attract the moving element and forcing movement to multiple specifiers to be to the lowest available specifier). In the case of movement to multiple specifiers, then, Shortest will have the desired effect of forcing maximally crossing paths:



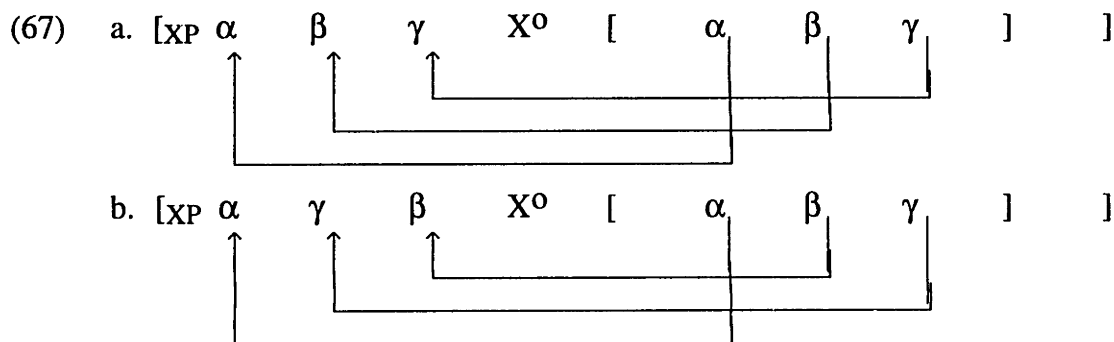
We expect that α should move first, and β next, and that their paths should cross.

2.6.2.1 Superiority

Suppose we consider how this theory might interact with the PMC. We have seen that the PMC allows an attractor which has obeyed a constraint to be free of that constraint for the rest of the derivation. In the case given in (65), the constraint to which the attractor is subject is *Shortest*; once X^0 has attracted α , it is no longer subject to *Shortest*. In (65), there is only one other possible target for attraction, namely β , so the fact that X^0 is now unconstrained by *Shortest* is irrelevant. Suppose we consider a case with more than two moveable elements:



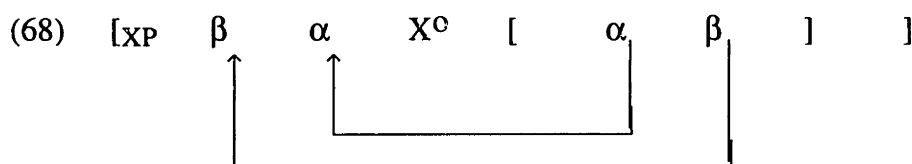
X^0 will have to attract α first, as before, but now it will be free to attract either β or γ next, since *Shortest* no longer applies to X^0 . We thus expect that either of the derivations in (67) should be possible:



After attracting α , X^0 may attract either β or γ ; we thus predict that the word order of the second and third wh-words should be free. In fact, we will see that this is the case; this was first observed by Bošković (1995a, to appear), who proposes an account very similar

in spirit to the one developed here. Thus, the PMC is apparently capable of “deactivating” Shortest for an attractor which has already obeyed that constraint.

On the other hand, we do not want the PMC to have a similar effect on the effects of Shortest on the relation between the attracted feature F and the copy α' of the domain containing it; that is, it should not be the case that a well-formed move allows the creation of a move to the same attractor whose head and tail are not maximally close to each other. If this were possible we would expect movement to multiple specifiers of a single head to be able to yield nested paths, as in (68):



We might expect Shortest-obeying movement of α to be able to license a subsequent Shortest-violating movement of β triggered by the same attractor, just as a movement which obeys Subjacency can improve the status of a subsequent Subjacency-violating movement to the same attractor. Why does this not happen?

We encountered a similar phenomenon above in our discussion of the Connectedness cases. Recall that these were cases in which a well-formed link rendered islands along its path transparent for purposes of the creation of a second link which would be ill-formed in isolation. The PMC was only able to render such islands transparent if they were along the path of a well-formed link:

- (69) a. *the person Op_i that John described Mary [without examining any pictures of e_j]
 b. ?the person Op_i that John described e_j [without examining any pictures of e'_j]
 c. *the person Op_i that John described e_j [without [any pictures of e'_j] being on file]

The interesting difference here is that between (69b) and (69c). Here the presence of a second island renders (69c) ill-formed, since this island is not along the path of a well-formed link. The ill-formedness of (69c) shows that the attractor in the Connectedness cases cannot be rendered impervious to the relevant constraint by the presence of a well-formed link; if this were possible, the placement of the islands would be irrelevant¹⁸.

In the Connectedness case, this problem was circumvented by defining the relevant constraint as applying not to the attractor, but to each link in the dependency, and as constraining the links in such a way that the PMC could not improve a structure by rendering the head of a link invisible. The constraint in question was (70b):

- (70) a. An attractor must trigger a dependency consisting of a (possibly singleton) set of well-formed links which share an index.
- b. A well-formed link must consist of a head α and a tail β such that there is no γ , γ a member of some set of barriers (including wh-islands, complex noun phrases, adjuncts, and subjects), such that α c-commands γ and γ dominates β .

(70b) requires links to have heads, so ignoring the head of a link would simply result in a headless link, which would violate (70b). Given that (70b) does not constrain attractors, invoking the PMC to make the attractor invisible would also have no effect in this case.

The definitions of *Attract* and *Shortest* given above will have a similar effect. *Shortest* does constrain the relation between the attractor and the attracted feature, and as we saw, the attractor can be rendered impervious to the effects of *Shortest*, allowing it to

¹⁸ The discussion in section 2.5.1.1 complicates this conclusion somewhat; we have seen that, for some reason, dependencies which the PMC saves by rendering the island transparent are better than dependencies in which the attractor is affected by the PMC. The contrast between (69b) and (69c), then, would still not be surprising even if it was the attractor which the PMC affected in (69c). The crucial datum now is the absence of any contrast between (69c) and (i) (imagine that John is a photographer whose job it is to provide photographs of important people for a newspaper's files):

(i) *the person Op_i that John retired [without [any pictures of e_i] being on file]

If (69c) is no better than (i), then the PMC is unable to affect the attractor in these cases.

attract a feature which is not the closest one. However, Shortest also constrains the relation between the attracted feature F and the copy α' of the structure α containing the attracted feature:

(71) *Shortest*

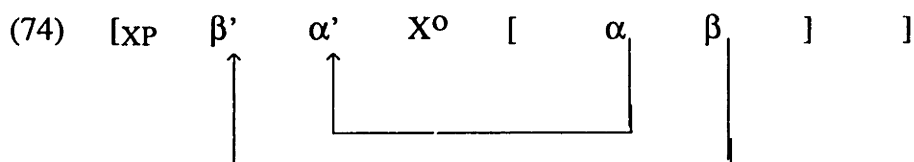
A pair P of elements $\{\alpha, \beta\}$ obeys Shortest iff there is no well-formed pair P' which can be created by substituting γ for either α or β , and the set of nodes c-commanded by one element of P' and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominating the other.

(72) *Attract*

An attractor K attracts a feature F , creating a copy α' of an element α containing F , and Merging α' with K . The relations between α' , K , and F must all obey Shortest.

- (73) a. $\quad \quad \quad K \quad [\alpha \quad F \quad]$
 b. $[\alpha' \quad F' \quad] \quad K \quad [\alpha \quad F \quad]$

Thus, consider again the case of nested movements to multiple specifiers, given in (74):



Here the well-formed movement of α occurs first, thus rendering the attractor X^0 impervious to Shortest. Of the three elements involved in the operation of attracting β , then (namely, X^0 , β' and β), Shortest will only constrain the relation between two, namely β' and β . This relation, however, is unaffected by the PMC; neither of these elements has participated in a well-formed dependency at this point. (74) is thus correctly predicted to be

ill-formed, since the relation between β' and β could be shorter than it is; β' could be in a specifier lower than α' ¹⁹.

Thus, the effects of Shortest on the relation between the head and tail of a chain are like the constraint on links in the Connectedness case, in that the relevant constraint is not a constraint on an attractor and thus cannot be affected by application of the PMC to make the attractor invisible. In section 2.6.2.2 we will discuss a case in which the relation between the head and tail of a chain is improved by the PMC. As in the Connectedness case, then, we will see that this constraint is in fact subject to the PMC, although the ways in which the PMC can affect it are limited by the way in which the constraint is stated.

To summarize, then; we expect the PMC to apply to Shortest, rendering attractors immune to Shortest once they have obeyed it once. The effects of Shortest on the relation between the head and tail of a chain, on the other hand, should be unaffected by the PMC for purposes of movement to multiple specifiers. Thus, every movement to multiple specifiers will have to land in a lower specifier than any existing specifier, but we expect to see some freedom in the order in which possible movers are attracted; in particular, once the highest mover is attracted first, the other movers may be attracted in any order.

2.6.2.1.1 Bulgarian

Bošković (1995a, to appear) points out that multiple wh-movement in Bulgarian behaves as the theory sketched above predicts it should. Consider first the rigid ordering between a pair of non-subject wh-words:

¹⁹ Note that since α' has participated in a Shortest-obeying relation, we might expect to be able to ignore its presence in (74) and thus remedy the ill-formed dependency between β' and β . The particular definition of Shortest in terms of the length in nodes between the two members of the constrained pair avoids this result; Shortest is not measured by comparison of the c-command domains of the head and the tail, for instance, so the presence or absence of α' is of no importance. If we wanted to maintain a version of Shortest in terms of c-command (requiring, essentially, that the difference in the sets of elements c-commanded by the head and tail of a pair of constrained elements be minimized), we might draw a distinction between α' and the structure dominated by α' , with the PMC affecting only the former. On this approach, β' in (74) would still “needlessly” c-command the structure dominated by the node α' , though effectively not α' itself, and would still be rejected in favor of a position c-commanded by α' .

Bulgarian (Bošković 1995a, 13-14)

- (75) a. Kogo kakvo e pital Ivan
 whom what AUX asked Ivan
 ‘Who did Ivan ask what?’
 b. ?*Kakvo kogo e pital Ivan
 what whom AUX asked Ivan

Shortest forces the attractor to attract the wh-words in a particular order, and also applies to each movement to chain to require it to be maximally short (and thus to land in a specifier below any existing specifiers). The prediction is that if the subject is also a wh-word, the second and third wh-words will be freely ordered. This prediction is borne out:

Bulgarian (Bošković 1995a, 13-14)

- (76) a. Koj kogo kakvo e pital
 who whom what AUX asked
 ‘Who asked whom what?’
 b. Koj kakvo kogo e pital

Movement of *koj* to Spec CP renders C⁰ impervious to Shortest; it is now free to attract either *kogo* or *kakvo* next. Whichever is attracted, however, the relation between the head and the tail of the resulting chain will also have to obey Shortest, and thus each wh-word will land in a specifier below any existing specifiers. The freedom of word order in (76) follows. In (75), on the other hand, movement of the higher wh-word *kogo* still renders C⁰ impervious to Shortest, but now there is only a single wh-word left to attract, namely *kakvo*; the PMC therefore has no visible effects in this case.

2.6.2.1.2 Japanese

An arguably similar phenomenon is found in Japanese wh-movement. The Japanese wh-word *naze* ‘why’ cannot c-command another wh-phrase which takes the

same scope, a fact discussed by Saito (1982, 1994a, 1994b, 1996), Maki (1994), and Watanabe (1992), among others; this phenomenon is known as “Anti-Superiority” in the literature:

Japanese (Saito 1994a, 195)

(77) a. John-ga nani -o naze katta no?

John NOM what ACC why bought Q

‘Why did John buy what?’

b. *John -ga naze nani -o katta no?

John NOM why what ACC bought Q

I will discuss one possible reason for this restriction later in this section; for the time being, let us simply note as a descriptive fact that *naze* must be lower in the clause than other wh-words.

Interestingly, the addition of a third, c-commanding wh-word can improve the status of (78b):

Japanese (Saito 1994a, 196)

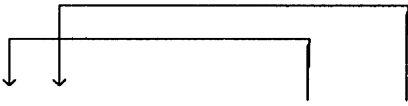
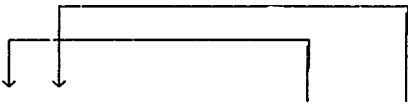
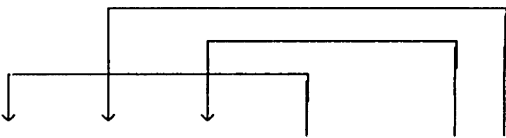
(78) Dare -ga naze nani -o katta no?

who NOM why what ACC bought Q

‘Who bought what why?’

It seems reasonable to analyze the contrast between (77b) and (78) in a way which relates these facts to the Bulgarian facts discussed in the previous section. We saw that when three wh-words are attracted by a single C^0 in Bulgarian, the second and third are freely ordered; in particular, the second wh-word can be attracted last, as though it were the lowest wh-word in the clause. We have seen that *naze* must usually be the lowest wh-word in its clause, except when it is the second of three wh-words. If the constraint on *naze* is a representational one requiring it to be the lowest wh-word after all wh-words

have been attracted, then the Japanese and Bulgarian data can be unified. At LF, (77)-(78) would have the following representations:

- (79) a.  John-ga nani -o naze katta no?
John NOM what ACC why bought Q
- b. *  John -ga naze nani -o katta no?
John NOM why what ACC bought Q
- c.  Dare -ga naze nani -o katta no?
who NOM why what ACC bought Q

In (79a-b), the two wh-words are forced by Shortest to follow crossing paths, and *naze* is the lower wh-word at the end of the derivation only in (79a). In (79c), as in Bulgarian, the additional wh-word renders the relevant attractor invulnerable to Shortest, allowing it to attract *nani* 'what' second and *naze* last. As a result, the well-formedness condition requiring *naze* to be the lowest wh-word in Spec CP is satisfied.

Tanaka (in preparation) discusses a number of phenomena in Japanese with the same general character, to which we now turn. One relevant phenomenon is the distribution of nominals marked with *sika* 'only'. *sika* NPs are negative polarity items, and must be associated with sentential negation either in their own clause or in a higher clause:

Japanese (Tanaka (in preparation))

- (80) a. John -ga [Mary-sika ko -nai to] itta
 John NOM Mary only come NEG that said
 ‘John said nobody but Mary had come’
 b. John-ga [Mary-sika kita to] iwa -nakatta
 John NOM Mary only come that say NEG-PAST
 ‘John didn’t say that anybody but Mary had come’

Like *naze* ‘why’, NPs marked with *sika* cannot c-command a wh-word:

Japanese (Tanaka (in preparation))

- (81) a. *John-sika nani -o yoma-nai no?
 John only what ACC read NEG Q
 ‘What did only John read?’
 b. Nani-o_i John-sika t_j yoma-nai no?
 what ACC John only read NEG Q
 (82) a. Dare-ga LGB-sika yoma-nai no?
 who NOM LGB only read NEG Q
 ‘Who read only LGB?’
 b. *LGB-sika_i dare-ga t_j yoma-nai no?
 LGB only who NOM read NEG Q

Moreover, as was the case with *naze*, the ban on *sika*-NPs c-commanding wh-words can be overcome by the presence of an additional wh-word with the same scope as the offending wh-word which is not c-commanded by the *sika*-NP:

Japanese (Tanaka (in preparation))

- (83) a. *John-wa [Mary-sika nani -o yoma-nai to] Sue -ni itta no?

John TOP Mary only what ACC read NEG that Sue DAT said Q

‘What did John tell Sue that only Mary read?’

- b. John-wa [Mary-sika nani -o yoma -nai -to] dare -ni itta no?

John TOP Mary only what ACC read NEG that who DAT said Q

‘Who did John tell that only Mary read what?’

Tanaka (in preparation) develops an approach to these facts in which *sika*-NPs and *wh*-words move to specifiers of NegP and CP, respectively. As he points out, there is evidence from the ordering of morphemes in the Japanese verb that CP dominates NegP; *nai* ‘NEG’ is closer to the verb stem than *no* ‘Q’. Translating Tanaka’s insight into our terms, we may see the ban on *sika*-NPs c-commanding *wh*-words and the requirement that multiple *wh*-movement paths cross in languages like Bulgarian and Chinese as instances of the same phenomenon:

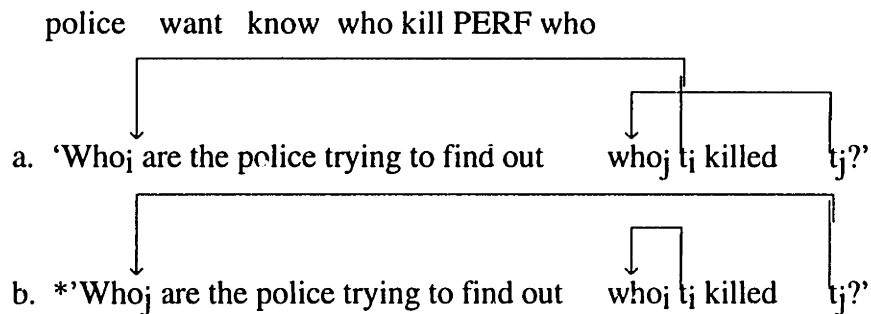
Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova,

Kamen Stefanov, p.c.)

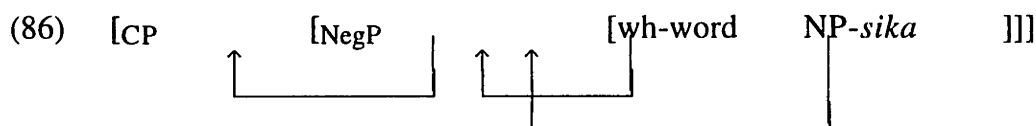
- (84) a.? Koj se opitvat da razberat kogo t e ubil t ?
 who SELF try to find-out whom AUX killed
- b.* Kogo se opitvat da razberat koj t e ubil t ?
 whom SELF try to find-out who AUX killed
-

Chinese (Lisa Cheng, Hooi Ling Soh, Wei-tien Dylan Tsai, p.c.)

(85) jingcha xiang-zhidao [shei sha -le shei]



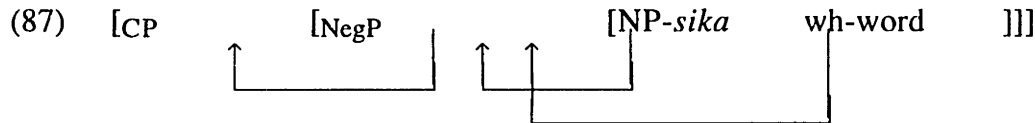
Recall the account of the facts in (84-85) developed in chapter 3, section 8. Bulgarian and Chinese are languages with multiple specifiers of CP, and these multiple specifiers are not equidistant; thus, the two wh-words move to multiple specifiers of the embedded C⁰, and when the higher C⁰ attracts a wh-word, it attracts the higher one. Tanaka's Japanese data may be seen in a similar light:



As in (84-85), Tanaka's data involve two attractors and two possible attracted elements, and Tanaka's observation is that, as in the Bulgarian and Chinese case, the underlying c-command relation between the two attracted elements must be preserved. To apply the account of the Bulgarian and Chinese facts to Tanaka's data, we will need to understand NegP and CP as both attracting by means of a single feature which both wh-words and *sika*-NPs possess, so that both kinds of NPs can be attracted in the syntax to either head. Furthermore, NegP and CP will need to be able (in principle) to have multiple specifiers²⁰. Thus, in (86), NegP attracts both the wh-word and the *sika*-NP, and CP then attracts the

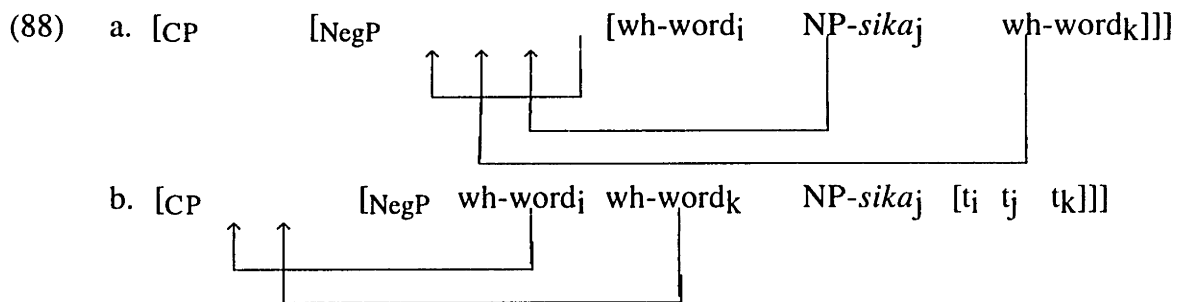
²⁰ Recall that in Chapter 2 I defended the idea that Japanese CP can only have a single specifier; in the terminology developed there, Japanese, like Serbo-Croatian, is an IP-absorption language. On the other hand, we have seen that Serbo-Croatian may also have multiple specifiers, in principle, as long as only a single specifier is occupied in the overt syntax. One possibility is that this property of Serbo-Croatian can be extended to the relevant Japanese cases. Another is that what I have labeled "CP" here is simply misnamed, and is in fact an IP-level projection. I will not try to choose between these alternatives here.

higher of NegP's two specifiers, the one occupied by the wh-word. At the end of the derivation, the wh-word is in Spec CP and the *sika*-NP is in Spec NegP, and the structure is therefore well-formed. If the positions of the two operators were reversed, however, an ill-formed structure would result:



At the end of the derivation in (87), the wh-word is in Spec NegP, and the *sika*-NP is in Spec CP. This is not a well-formed representation, as the operators are not in specifiers of the appropriate heads for interpretation.

On the other hand, if we add a third wh-word above the *sika*-NP, the sentence improves, as we have seen. This is again expected; the third wh-word will allow the second and third operators to “switch places” as they are attracted, resulting in a well-formed structure. The derivation of an example like (83b), then, would be as in (88):



In (88a) the higher wh-word is attracted first to NegP, rendering Neg⁰ impervious to Shortest and allowing it to attract the second wh-word next, and then the *sika*-NP last. As a result, the two wh-words are in the two higher specifiers of NegP, and can both be attracted by C⁰ without difficulty in (88b).

It is unclear whether this account of the distribution of *sika*-NPs can be extended to the Anti-Superiority effect for *naze*. We have seen that the distribution of *naze* is remarkably like that of a *sika*-NP; both are unable to c-command wh-words, and this

effect, in both cases, can be removed by the presence of an additional wh-word. We might entertain a theory, then, on which *naze*, like *sika*-NPs, must be interpreted in Spec NegP. One additional bit of evidence for this approach is noted by Miyagawa (1997b), who points out that *naze* is unlike other wh-words in being immune to the ban on wh-words being c-commanded by *sika*-NPs:

Japanese (Miyagawa 1997b, 10)

(89) a. Hanako-sika naze ik-anai no?

Hanako only why go NEG Q

‘Why is only Hanako going?’

b. *Hanako-sika doko -e ik-anai no?

Hanako only where to go NEG Q

‘Where is only Hanako going?’

If *naze* is unlike other wh-words in being interpretable in Spec NegP, then the contrast between (89a) and (89b) would follow straightforwardly; there is no need for *naze* to move into Spec CP.

There are at least two objections to an account of Anti-Superiority of this type, however. One is that it is unclear why *naze* should be interpreted in Spec NegP while other wh-words are interpreted in Spec CP. At the very least, this would seem to show that these projections are mislabelled. A second, possibly more serious objection has to do with the nature of the additional-wh effect for *sika*-NPs and for *naze*. We have seen that both Anti-Superiority effects and the ban on *sika*-NPs c-commanding wh-words can be obviated by the presence of an additional wh-word which is not c-commanded by the offending operator:

Japanese (Saito 1994a, 195-196)

- (90) a. *John -ga naze nani -o katta no?

John NOM why what ACC bought Q

‘What did John buy why?’

- b. Dare -ga naze nani -o katta no?

who NOM why what ACC bought Q

‘Who bought what why?’

Japanese (Tanaka (in preparation))

- (91) a. *John-wa [Mary-sika nani -o yoma-nai to] Sue -ni itta no?

John TOP Mary only what ACC read NEG that Sue DAT said Q

‘What did John tell Sue that only Mary read?’

- b. John-wa [Mary-sika nani -o yoma -nai -to] dare -ni itta no?

John TOP Mary only what ACC read NEG that who DAT said Q

‘Who did John tell that only Mary read what?’

However, there is an important difference between the saving effect for *naze* and the saving effect for *sika*-NPs. As (91) shows, the additional wh-word for a *sika*-NP need not be a clausemate of the *sika*-NP. For *naze*, however, the wh-words must be clausemates:

Japanese (Saito 1996, 27)

- (92) a. *Mary -ga [John -ga naze nani -o katta to] omotteru no?

Mary NOM John NOM why what ACC bought that thinks Q

‘What does Mary think [that John bought t why]?’

- b. Mary-ga [dare -ga naze nani -o katta to] omotteru no?

Mary NOM who NOM why what ACC bought that thinks Q

‘Who does Mary think [t bought what why]?’

- c. *Dare -ga [John -ga naze nani -o katta to] omotteru no?

who NOM John NOM why what ACC bought that thinks Q

‘Who thinks [that John bought what why]?’

The interesting contrast is the one between (91c) and (92c); here we can see that a non-clausemate additional *wh*-word can save an ill-formed construction with a *sika*-NP, but not one with *naze*. This would seem to indicate that the constraint which is responsible for Anti-Superiority constraints every intermediate CP, not merely the ultimate landing site of the *wh*-words. In (91c), we might say, the intermediate C^0 only attracts two *wh*-words, and must therefore attract them in a way which violates the requirement that *naze* be c-commanded by any other *wh*-words in Spec CP.

On the other hand, Maki (1994) notes that *naze* combined with another *wh*-word outside its clause is ill-formed in any case, even when no Anti-Superiority violation is involved:

Japanese (Maki 1994, 201)

- (93) a. [John -ga [Mary -ga naze kubininatta to] omotteiru ka] osiete kudasai
 John NOM Mary NOM why was-fired that think Q tell please
 ‘Please tell me [why John thinks [that Mary was fired t]]’
- b. ?*[Dare -ga [Mary -ga naze kubininatta to] omotteiru ka] osiete kudasai
 who NOM Mary NOM why was-fired that think Q tell please
 ‘Please tell me [who thinks [that Mary was fired why]]’

I have no account of the ill-formedness of (93b), but it seems possible that whatever rules out (93b) will rule out (92c) as well. To summarize, then, the idea that *naze* and *sika*-NPs are both interpreted in Spec NegP encounters serious but perhaps not insurmountable objections, and unifies the Anti-Superiority facts with the distribution of *sika*-NPs. I will have to leave the question of whether this unification is in fact legitimate to future research.

Tanaka (in progress) discusses another interaction between wh-words and a type of operator which gives substantial support to his approach. NPs marked with the emphatic suffix *-koso*, like *sika*-NPs, are restricted in their possible c-command relations with wh-words. However, the restriction in the case of *koso*-NPs is the exact opposite of the restriction on *sika*-NPs; *koso*-NPs must c-command all wh-words in their clause, while *sika*-NPs cannot c-command wh-words:

Japanese (Tanaka (in preparation))

- (94) a. John-koso nani -o yonda no?

John EMPH what ACC read Q

‘What did *John* read?’

- b. *Nani-oj John-koso t_j yonda no?

what ACC John EMPH read Q

- (95) a. *Dare-ga LGB-koso yonda no?

who NOM LGB EMPH read Q

‘Who read *LGB*?’

- b. LGB-kosoj dare-ga t_j yonda no?

LGB EMPH who NOM read Q

Compare the facts with *sika*-NPs ((81-82), repeated as (96-97)):

Japanese (Tanaka (in preparation))

- (96) a. *John-sika nani -o yoma-nai no?

John only what ACC read NEG Q

‘What did only John read?’

- b. Nani-oj John-sika t_j yoma-nai no?

what ACC John only read NEG Q

- (97) a. Dare-ga LGB-sika yoma-nai no?

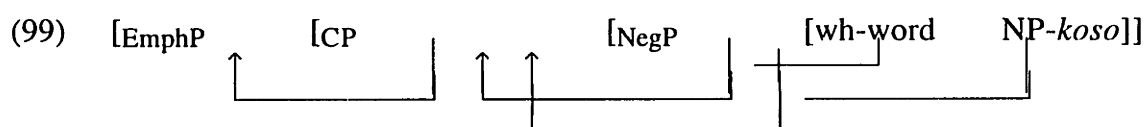
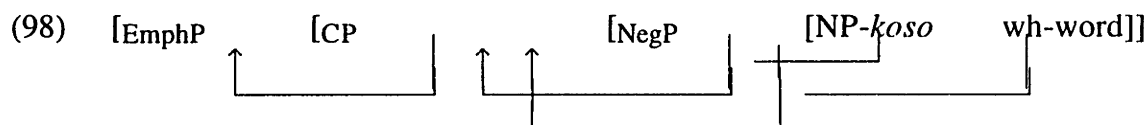
who NOM LGB only read NEG Q

‘Who read only *LGB*?’

- b. *LGB-sikaj dare-ga t_j yoma-nai no?

LGB only who NOM read NEG Q

Following Tanaka (in preparation), we may extend the analysis of the distribution of *sika*-NPs to *koso*-NPs by postulating an EmphP dominating CP, to which *koso*-NPs must move to be interpreted:



As desired, (98) will be a well-formed representation, in which the *koso*-NP moves into Spec EmphP, and the *wh*-word into Spec CP. (99) will be ill-formed, since the two operators will end the derivation in specifiers of the wrong heads; the *wh*-word is in Spec EmphP, and the *koso*-NP in Spec CP.

This approach makes a number of predictions, which appear to be borne out. One prediction, as Tanaka (in preparation) notes, is that *sika*-NPs will behave like *wh*-words as far as *koso*-NPs are concerned; that is, they will have to be c-commanded by any *koso*-NP in the sentence:

Japanese (Tanaka (in preparation))

- (100) a. John-koso LGB-sika yoma-nai

John EMPH LGB only read NEG

‘John read only LGB’

- b. *LGB-sika_i John-koso t_j yoma-nai

LGB only John EMPH read NEG

Another prediction, which cannot be tested, is that we should get an “additional-*koso* effect” for sentences in which *koso*-NPs are c-commanded by a *wh*-word, just as *wh*-

words which are c-commanded by *sika*-NPs can be saved by the presence of an additional wh-word. Because of the oddity of multiple *koso*-NPs in a sentence, judgments are unclear on this point.

Finally, we predict that a wh-word c-commanded by a *sika*-NP can be saved, not only by another wh-word, but by a *koso*-NP. Recall that the theory developed here is one in which all three of these operator types may move into specifiers of the three relevant heads; the only requirement is that operators and heads be appropriately matched by the end of the derivation. In any case in which there are three operators, the bottom two operators may “change places” for reasons having to do with the PMC, as we have seen. Thus, any operator whose ultimate destination is higher than that of *sika*-NPs--that is, either a wh-word or a *koso*-NP--should be able to rescue a wh-word c-commanded by a *sika*-NP. This prediction is borne out:

Japanese (Takako Aikawa, Shigeru Miyagawa, p.c.)

- (101) a. *John-wa [Mary-sika nani -o yoma-nai to] Sue -ni itta no?

John TOP Mary only what ACC read NEG that Sue DAT said Q

‘What did John tell Sue that only Mary read?’

- b. John-wa [Mary-sika nani -o yoma -nai -to] dare -ni itta no?

John TOP Mary only what ACC read NEG that who DAT said Q

‘Who did John tell that only Mary read what?’

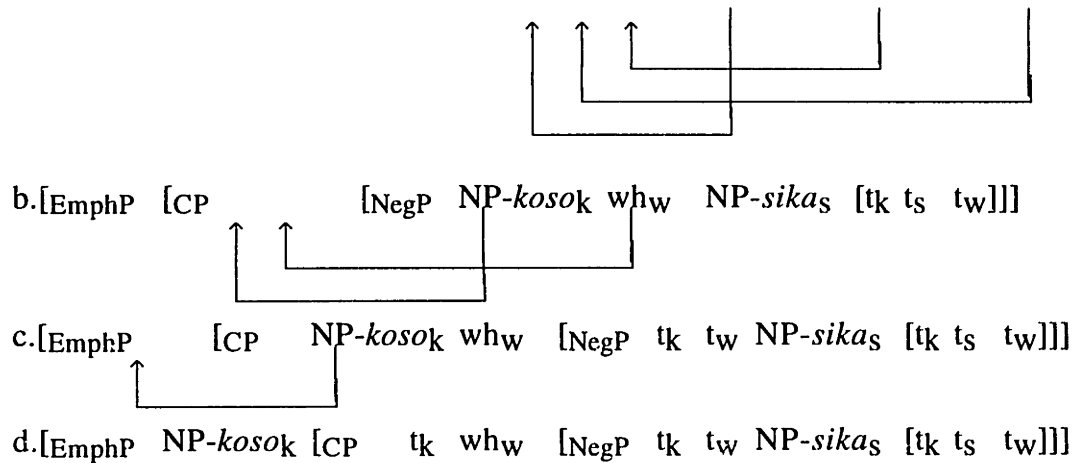
- c. ? John-wa [Mary-sika nani -o yoma-nai to] Sue -koso -ni itta no?

John TOP Mary only what ACC read NEG that Sue EMPH DAT said Q

‘What did John tell *Sue* that only Mary read?’

The contrast in (101a-b) is the additional-wh effect we saw before in (91a-b). (101c) is the interesting case; here a *koso*-NP has been substituted for the additional wh-word of (101b), and the resulting sentence is good. The derivation of (101c) would be as in (102):

- (102) a. [EmphP [CP [NegP [NP-*koso*_k NP-*sika*_S wh_w]]]



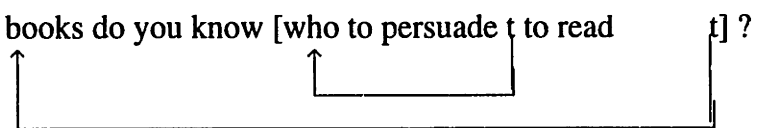
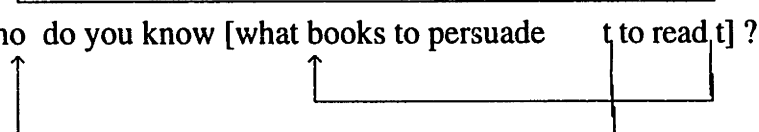
In (102a), Neg⁰ attracts all three operators, and because it is attracting three elements the second two may be attracted in any order; thus, the *sika*-NP and the *wh*-word change places. The other steps are all straightforward; in (102b), the occupants of the two topmost specifiers of NegP move to the multiple specifiers of CP, and in (102c), the *koso*-NP in the highest specifier of CP moves to Spec EmphP. At the end of the derivation all the operators are in the appropriate maximal projections.

We have seen, then, that Japanese exhibits signs of the Bulgarian phenomenon discovered by Bošković (1995a, to appear). In both languages, when an attractor attracts more than two elements, the elements after the first may be attracted in any order. This phenomenon may be attributed, on the account developed here, to the application of the PMC to Shortest; once an attractor has attracted one element in a manner which obeys Shortest, it is immune to Shortest for the rest of the derivation. To the extent that the account of the Bulgarian facts generalizes naturally to account for the Japanese facts, we are again led to an approach that postulates covert movement, as opposed to one in which *wh*-in-situ is interpreted in situ. The account just given makes crucial reference to the creation of dependencies at particular points in the derivation, and it is therefore not clear how such an account could be recreated in an approach that interpreted *wh*-in-situ by means of unselective binding, for instance.

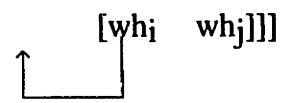
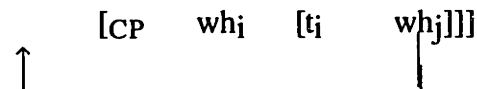
2.6.2.2 Path Containment Condition

In chapter 3 I discussed an insight of Kitahara (1994, 1997), who points out that the effects of Pesetsky's (1982) Path Containment Condition can be made to follow from Cyclicity and Shortest. The effects of the Path Containment Condition are exemplified in (103):

(Pesetsky 1982, 268)

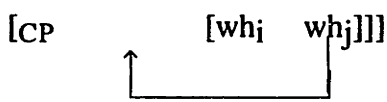

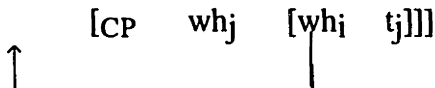
- (103) a. what books do you know [who to persuade t to read t] ?

 b. * who do you know [what books to persuade t to read t] ?


Following Kitahara, we can postulate the derivation in (104) for the well-formed (103a):

- (104) a. [CP [CP [wh_i wh_j]]]

 b. [CP [CP wh_i [t_i wh_j]]]
 c. [CP [CP wh_i [t_i wh_j]]]


In (104a), the lower C⁰ is Merged and attracts the closest available wh-word, obeying Shortest. In (104b), Merge continues until the matrix C⁰ is Merged, at which point it must again attract the closest available wh-word. In (104c), it attracts the lower of the two wh-words.

Depending on our assumptions, this last move involve a violation of Shortest on the part of the attractor, since wh_i is closer to the higher C⁰ than wh_j is. Alternatively, we might define Shortest so that this move does not violate it, since the higher wh-word is already in a checking relation with a C⁰ and thus might be ineligible for further checking.

- (106) a. 
- b. 
- c. 

In (106a), the newly-Merged embedded C^0 attracts the lower of the two *wh*-words, thus violating Shortest. The higher C^0 is then Merged in (106b), and attracts the remaining *wh*-word (possibly violating Shortest again, depending on our assumptions, as noted above).

Thus, the effects of the Path Containment Condition seem to follow fairly straightforwardly from basic assumptions, a desirable result. Let us consider the distribution of PCC effects in other languages, and see how this account fares.

2.6.2.2.1 The PCC in other languages

We have seen that in CP-absorption languages in which *wh*-movement all occurs on a single level, an anti-PCC effect is observed; intersecting *wh*-movement paths obligatorily cross:

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova,

Kamen Stefanov, p.c.)

- (107) a. Koj se opitvat da razberat kogo t e ubil t ?
 who SELF try to find-out whom AUX killed

- b. * Kogo se opitvat da razberat koi t e ubil t ?
 whom SELF try to find-out who AUX killed

- (108) a. Na kogo se opitvaš da razbereš kakvo dade Ivan t t ?
 to whom SELF you-try to find-out what gave Ivan

- b. * Kakvo se opitvaš da razbereš na kogo dade Ivan t t ?
 what SELF you-try to find-out to whom gave Ivan

Chinese (Hooi Ling Soh, Lisa Cheng, Wei-Tien Dylan Tsai, p.c.)

- (109) jingcha xiang-zhidao [shei sha -le shei]

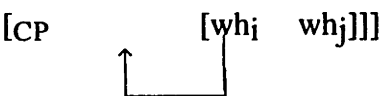
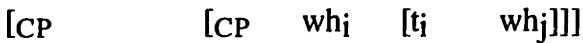
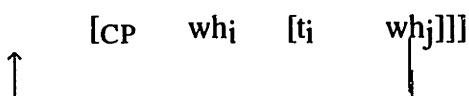
police want know who kill PERF who

- a. 'Who_i are the police trying to find out who_j t_j killed t_j?'
 t_j?

- b.* 'Who_j are the police trying to find out who_i t_i killed t_j?'
 t_j?

Why do anti-PCC effects appear in these languages? In chapter 3 I suggested that the answer should be tied to the availability of multiple specifiers of CP in these languages, in the following way. Assume a version of Shortest according to which the wh-movement to

the higher CP in a derivation with nested paths does involve a violation of Shortest on the part of an attractor; the matrix C^0 fails to attract the closest wh-word in these cases, which is the one already in the specifier of the lower CP. In English, this violation of Shortest by the attractor is unavoidable. Consider again the derivation that yields nested paths in English:

- (110) a. 
- b. 
- c. 

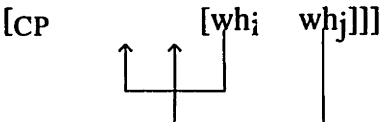

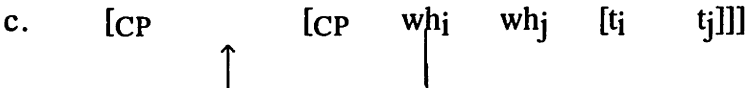
In (110c), the higher C^0 violates Shortest by attracting wh_j instead of wh_i . If it were to obey Shortest, the embedded C^0 would be left without a wh-word in its specifier in the overt syntax. This is apparently intolerable.

One reason such an operation might be intolerable would follow fairly naturally from the approach to successive-cyclic wh-movement suggested in chapter 4. The suggestion was that all complementizers in a given language have the same syntactic properties; in English, all complementizers have a single strong +wh feature. The difference between declarative and interrogative complementizers, on this view, is purely one of interpretation. A complementizer is interpreted as interrogative if it has a wh-word in its specifier at the end of the derivation.

In section 4.3 of chapter 4, I suggested that the computational system “realizes” that a wh-feature on a declarative complementizer is uninterpretable, and therefore deletes it, when the wh-word moves on to another complementizer. This assumption was useful in dealing with the properties of improper movement. In fact, the theory developed in chapter 4 demands that something of this kind should take place; a wh-word checking multiple

strong wh-features on distinct heads would be an ill-formed PF object. If this approach is on the right track, then attraction of the already-moved wh-word wh_i by the higher C^0 in an example like (110) will simply result in the embedded C^0 's being interpreted as declarative; its wh-feature will be viewed as uninterpretable and deleted by the computational system when the wh-word moves to the higher complementizer.

In languages like Bulgarian and Chinese, however, the embedded C^0 has two wh-words in its specifier at the point in the derivation corresponding to (110b), and the higher of the two may therefore be freely attracted; since the embedded complementizer will still have a wh-word in its specifier, its [+wh] feature will not be deleted. The derivation of the well-formed examples with crossing paths, above, proceeds as follows:

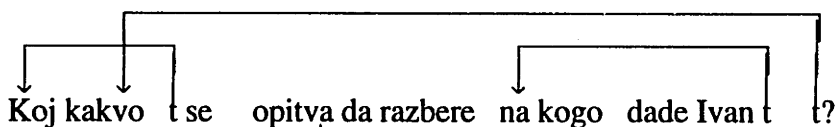
- (111) a. 
- b. 
- c. 

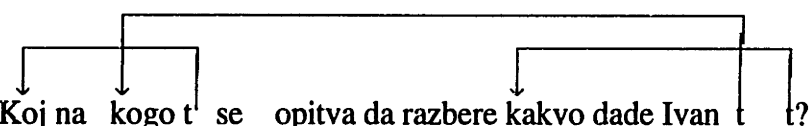
In (111a), both wh-words are attracted into multiple specifiers of the lower CP. The matrix CP is later Merged. In this case, the higher of the two wh-words can move freely into the specifier of the higher CP, since there will still be a wh-word left in the lower CP, allowing it to be interpreted. The crucial difference between English on the one hand and languages like Bulgarian and Chinese on the other, on this account, arises from the availability of multiple specifiers of CP in languages like Bulgarian and Chinese, which permit multiple wh-movement to all occur simultaneously. As a result, Shortest may be satisfied in these languages, since the highest wh-word is always available for further movement. In English, on the other hand, only one wh-word may move to each C^0 in the overt syntax,

and that wh-word must therefore stay in its position; the attractor thus unavoidably violates Shortest²¹.

One piece of evidence that Shortest is responsible for the anti-PCC effect in languages like Bulgarian and Chinese has to do with an additional-wh effect in these languages. PCC effects reemerge in these languages if an additional wh-word is added in a position above both of the wh-words in the lower clause:

Bulgarian (Roumyana Izvorski, p.c.)

- (112) a. 
 who what SELF try to find-out to whom gave Ivan
 'Who is trying to find out to whom Ivan gave what?'

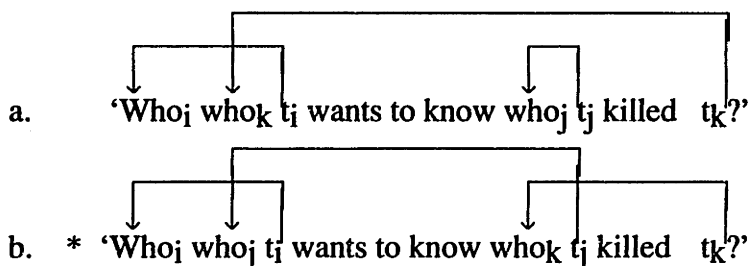
- b. * 
 who to whom SELF try to find-out what gave Ivan
 'Who is trying to find out what Ivan gave to whom?'

²¹ In principle, we could define the set of possible attractees such that Shortest would be violated by attraction of the lower wh-word in Bulgarian and Chinese, but not in English. That is, we could make Shortest "aware" of the fact that the single wh-word in the lower Spec CP is immovable in English, making attraction of the lower wh-word the only (and thus the Shortest) option. This account differs only very slightly from the one developed above. Given that even PCC-obeying structures in English are not perfect, it seems reasonable to appeal to a Shortest violation in the English case as well.

Chinese (Wei-tien Dylan Tsai, Hooi-Ling Soh, p.c.)

(113) Shei xiang-zhidao shei sha -le shei?

who want know who kill PERF who



This is very reminiscent of the data having to do with the effects of the PMC on Shortest as applied to the attractor, discussed in section 2.6.2.1 of this chapter. We have seen that when three wh-words are all attracted by a single attractor, the second and third may be attracted in any order:

Bulgarian (Bošković 1995a, 13-14)

(114) a. Kogo kakvo e pital Ivan

whom what AUX asked Ivan

'Who did Ivan ask what?'

b. ?*Kakvo kogo e pital Ivan

what whom AUX asked Ivan

(115) a. Koj kogo kakvo e pital

who whom what AUX asked

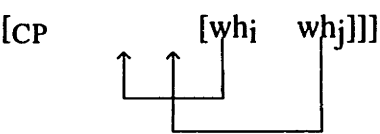
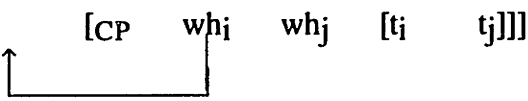
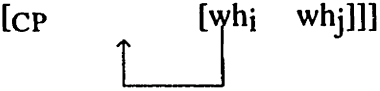
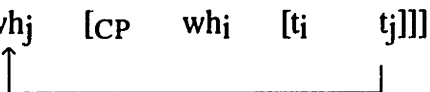
'Who asked whom what?'

b. Koj kakvo kogo e pital

The claim made above was that the optionality in (115) reflected the effect of the PMC on Shortest. Once *koj* has been attracted by C⁰, the PMC renders that C⁰ immune to the effects of Shortest for the rest of the derivation. C⁰ may now attract either *kogo* or *kakvo* next, and the result is an optionality in the ordering of the second and third wh-words.

The data in (112-113) are arguably similar; addition of a higher wh-word renders the matrix C⁰ immune to the effects of Shortest, and allows either of the two lower wh-words to be attracted, whereas without the additional wh-word only the higher of the two lower wh-words can be attracted. This is clearly not the whole story, however, since it simply predicts that nested and crossing paths for the two lower wh-words ought to be equally acceptable. Why is there a preference for nested paths in this case?

Consider the derivations involved in producing crossing and nested paths in languages like Bulgarian and Chinese (the additional wh-word in the nested case is omitted for the sake of clarity):

- (116) a. 
- b. [CP [CP wh_i wh_j [t_i t_j]]]
- c. 
- (117) a. 
- b. [CP [CP wh_i [t_i wh_j]]]
- c. 

One difference between the derivation with crossing paths in (116) and the one with nested paths in (117) has to do with the number of steps. In (116), the wh-word which eventually lands in the higher Spec CP moves there by means of two steps, while in (117) only one step is involved. The total number of “wh-movements” in (116), then, is higher than the

number in (117). It seems reasonable to appeal to a requirement of Fewest Steps in this case, which would prefer the derivation in (117) over that in (116) (for discussion of requirements of this type, cf. Chomsky 1991, Collins 1994, Oka 1995, Kitahara 1997, and references cited there):

(118) Given two well-formed derivations with the same Numeration, the derivation involving fewer operations is preferred.

(118) will give us the desired result (although it has a somewhat worrying global character)²². In the case without an additional-wh word, Shortest will require the higher of the two wh-words to be attracted in Bulgarian and Chinese; Fewest Steps will therefore have no effect. When the matrix C^0 is rendered immune to Shortest by the presence of an additional wh-word, both crossing and nested paths become possible in principle, and Fewest Steps prefers the derivation with nested paths.

In order for this approach to PCC effects to hold, as we saw, we must assume that the nested paths in English in fact violate Shortest; in a derivation with nested paths, when the higher C^0 attracts the lower of the two wh-words, it violates Shortest by failing to attract the wh-word which is already in the specifier of the lower C^0 . In principle, we might expect the PMC to allow attraction of the lower of the two wh-words, since the higher wh-word has already undergone a move which obeys Shortest. We apparently want to avoid this result.

One way of avoiding this result is actually implicit in the particular definition of the Attract relation assumed here, repeated as (119):

²² An alternative approach might be to claim that Fewest Steps is “more highly ranked” in English than it is in CP-absorption languages; this ranking might be used to explain both the lack of multiple overt wh-movement in English (on a derivational version of Fewest Steps, which would yield the same results as Procrastinate) and the presence of PMC effects even when Shortest Attract must be violated. I will not pursue this alternative further here.

(119) *Attract*

An attractor K attracts a feature F , creating a copy α' of an element α containing F , and Merging α' with K . The relations between α' , K , and F must all obey Shortest.

- (120) a. $\quad\quad\quad K \quad [\alpha \quad F \quad]$
 b. $[\alpha' \quad F' \quad] \quad K \quad [\alpha \quad F \quad]$

Consider an example like (121), where J and K are both attractors and α and β are words. (121) is the point in the derivation at which α has undergone movement to the specifier of K :

- (121) $J \quad [\alpha' \quad F' \quad] \quad K \quad [\alpha \quad F \quad] \quad [\beta \quad G \quad]$
-

If J attracts β , does it incur a violation of Shortest, as desired? There is some reason to believe that it does. The well-formed attraction of α creates relations between α' , K , and F which obey Shortest, and allows ignoring of these elements for purposes of evaluating other dependencies with respect to Shortest. The competing elements for attraction by J , however, are F' and G , neither of which have previously been affected by the PMC. We thus correctly predict that J violates Shortest when it attracts β instead of α , as long as we draw a distinction between α' and the feature F' contained in it (and between F and its copy F').

The particular definition of *Attract* used here, then, creates a relation of which the head and tail are of different “types”, the tail being a feature and the head a complex structure containing multiple features. As a result, movement will never render the head or tail of a chain invisible for purposes of Shortest, since the kinds of things which are attracted are not the kinds of things which are heads of dependencies created by *Attract*²³.

²³ A similar move with respect to the attractor will be necessary to avoid unpleasant consequences for the Head Movement Constraint; Shortest will have to constrain the relation between the attracting feature and

The only objects which the PMC will make invisible to the effects of Shortest will be attractors. This is a desirable result, as we have seen. Another case in which this result will be a welcome one is that of successive-cyclic movement; we would not want a well-formed movement to an intermediate specifier to license a Shortest-violating movement to a higher specifier.

The approach taken here assumes a single constraint, Shortest, which constrains relations between several different elements. It is crucial, on this account, that the head of a movement chain is not the moved feature but a copy of a larger structure containing the feature targeted by Attract. Thus, further attraction of features out of the head of a well-formed chain is still subject to Shortest; the part of the head of a chain which obeys Shortest is not the same as the part of the tail. Another approach might involve several formally distinct constraints, Shortest Attract and Shortest Move, of which the first constrains the relation between the attractor and the tail of a chain and the second the relation between the head and the tail. This would have the same empirical result as the approach sketched here; the head of a well-formed movement chain would still be subject to Shortest Attract, since it would have already obeyed Shortest Move but not Shortest Attract (which would only constrain relations between heads and tails). I have no strong arguments for choosing between these approaches, and will continue assume the first, without argument.

A third approach might claim that the only part of Shortest which is necessary as a constraint is Shortest Attract (that is, the requirement that an attractor attract the closest available mover). What I have here called the effects of Shortest on the relation between the head and tail of the chain (what we can think of as Shortest Move, as opposed to Shortest Attract) might be built into the definition of Attract, and thus made impervious to

the attracted feature. Otherwise, well-formed movement into the specifier of a head would licensing HMC-violating movement past that head.

the effects of the PMC. We have not yet seen any cases in which the relation between the head and tail of a chain is saved from Shortest by the PMC. We will now turn to such a case.

Let us consider the situation in IP-absorption languages in which all movement occurs on a single level. Here speakers differ; for some speakers, both crossing paths and nested paths are impossible, and two clausemate wh-words must always take the same scope. For those speakers who do allow distinct scopes for clausemate wh-words, the preference is for crossing paths:

Serbo-Croatian (Rudin 1988, 459, Željko Bošković, Milan Mihaljevic, p.c.)

- (122) a. * Šta si me pitao ko t može da uradi t?
 what AUX-2sg me asked who can to do
 ‘What have you asked me who can do?’

- b. ?* Ko si me pitao šta t može da uradi t?
 who AUX-2sg me asked what can to do
 ‘Who have you asked me what can do?’

Japanese (Takako Aikawa, Satoshi Oku, Shigeru Miyagawa, p.c.; cf. also

Nishigauchi (1990, 33), Saito (1994a, 198), Grewendorf and Sabel (1996, 57))

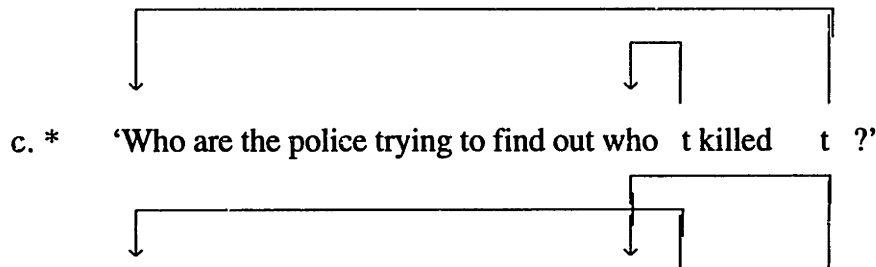
(123) Keesatu-wa [dare -ga dare -o korosita ka] sirabeteiru no?

police TOP who NOM who ACC killed Q are-investigating Q

a. 'Are the police trying to find out who killed who?'

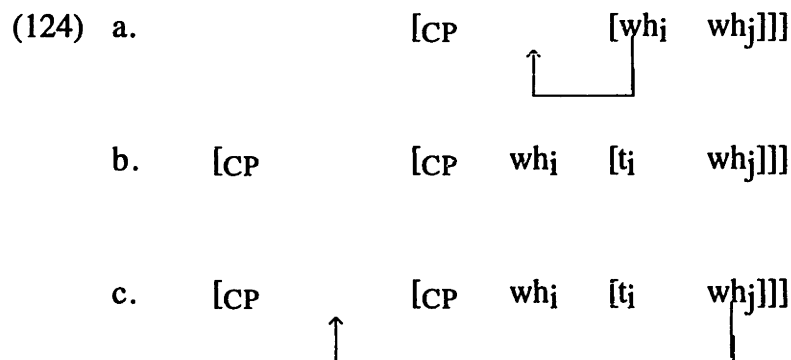
b. ??'For which murderer x and which victim y are the police investigating whether

x murdered y?'



d. ?? 'Who are the police trying to find out who t killed t?'

By hypothesis, these are languages, like English in the overt syntax, in which C^0 can only project a single specifier. The question, then, is why they do not behave like English with respect to PCC effects. Consider again the derivation of nested wh-movement paths in English:



Recall that the status of the final move in (124c) depends on our definition of Shortest, and of the sets of possible movers and landing sites. Depending on our assumptions, both the

higher C^0 and the head and tail of the resulting movement chain might in principle be in violation of Shortest here. The attractor is attracting a wh-word which is not the highest one, and the wh-word is moving past a potential landing site. We saw evidence from Chinese and Bulgarian suggesting that the higher C^0 does in fact violate Shortest in this case, at least in those languages, since PCC effects emerge in those languages just in the case in which the higher C^0 is rendered immune to Shortest by the presence of a higher wh-word. What about the wh-word?

There is reason to believe that even if the relation between wh_j and its trace in (124c) violated Shortest, the PMC would rescue it. Consider again the definition of Shortest:

(125) *Shortest*

A pair P of elements $\{\alpha, \beta\}$ obeys Shortest iff there is no well-formed pair P' which can be created by substituting γ for either α or β , and the set of nodes c-commanded by one element of P' and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominating the other.

In (124c), movement of wh_j violates Shortest by virtue of passing the embedded C^0 . Shortest applies to the pair $\{\alpha, \beta\}$, where α is the copy of wh_j in the specifier of the matrix C^0 and β is the attracted wh-feature in wh_j . The α in this pair could be replaced with a γ , a copy of wh_j in the specifier of the embedded C^0 , and the pair $\{\gamma, \beta\}$ is “shorter” than the pair $\{\alpha, \beta\}$.

However, the embedded C^0 has participated in a well-formed Attract relation which obeys Shortest, and can therefore be ignored when evaluating other relations with respect to Shortest. If we ignore the embedded C^0 , there are no places in the structure where a copy of wh_j could be placed which would create a well-formed dependency and would be closer to the wh-feature in wh_j 's base position than the copy in the specifier of the matrix

C^0 . Thus, the relation between *wh_j* and its trace does not in fact violate Shortest in English; it is only the matrix C^0 which violates Shortest in this case.

Now let us consider the IP-absorption languages again. These are languages in which *wh*-words can be interpreted by adjunction to IP-level projections. Such languages are distinguished, among other things, by a lack of Superiority effects for local *wh*-movement:

Serbo-Croatian (Bošković 1995c, 5-6)

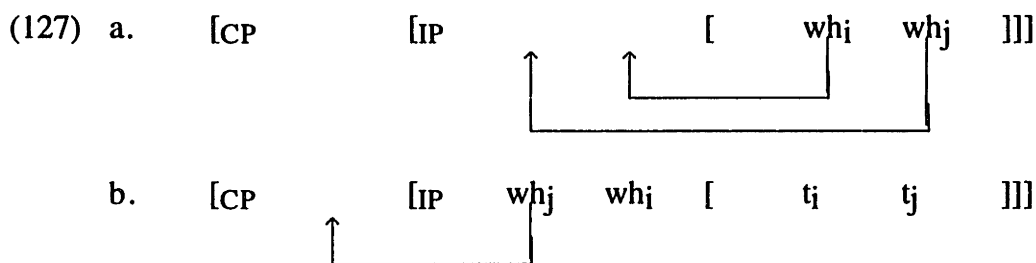
- (126) a. Ko je koga vidjeo?
 who AUX whom seen
 ‘Who saw whom?’
 b. Koga je ko vidjeo?

If the theory of multiple specifier ordering outlined in chapter 3 is correct, this freedom of ordering is indicative of the presence of multiple potential attractors for *wh*-words; movement to multiple specifiers of a single head yields strict Superiority effects. Thus, IP-absorption languages differ from English in that there are always multiple potential landing sites for *wh*-movement. Well-formed *wh*-movement to the embedded Spec CP, in this case, will make C^0 invisible for purposes of Shortest, but not the various IP-level positions, of which an arbitrary number are available. The equivalent of the movement in (124c) in an IP-absorption language, then, unavoidably involves a violation of Shortest not only on the part of the matrix attracting C^0 , as in English, but also on the part of the moving *wh*-word. We therefore expect this derivation to be worse in an IP-absorption language than it is in English, and it is.

In the English PCC-obeying cases, then, we have a case in which the PMC improves the status of the relation between the head and tail of a chain, by rendering invisible the intermediate C^0 which could in principle provide a closer landing site. The

crucial difference between a language like English and a language like Serbo-Croatian has to do with the greater number of available landing sites in the latter type of language. Thus, we apparently do want the locality requirement on the relation between the head and tail of a chain to be a constraint to which the PMC can in principle apply. As we have seen, the PMC's effects on this locality requirement are more "limited" in a sense than the effects of the PMC on attractors. The PMC cannot, for instance, render the head of a chain immune to Shortest for the rest of the derivation. In the theory developed above, this is because the head of a chain is never the kind of object which can be attracted and become the tail of a chain; a chain has as its tail a feature and as its head a copy of a larger domain which contains that feature. On the other hand, the PMC can improve the status of the relation between the head and tail of a chain by making potential landing sites for this relation invisible. In English, the PMC can have this effect, thus improving the status of nested paths, since there is only a single attractor responsible for *wh*-movement in English. In an IP-absorption language, because there are several attractors responsible for *wh*-movement, the PMC cannot make them all invisible, and nested paths are correspondingly worse than they are in English.

Let us consider the derivations which are possible in IP-absorption languages. The derivations which yield identical scope for both *wh*-words are comparatively trivial. When both *wh*-words move locally, they are both attracted into IP-level projections, and then one *wh*-word is attracted to Spec CP:



Here “IP” stands for a number of IP-level projections, which permit Superiority violations to occur (see chapter 2, section 3.3 for discussion). The derivation in (127) yields one such Superiority violation, although obedience to Superiority is also possible. Shortest is obeyed throughout the derivation in this case.

Now let us consider a case in which both *wh*-words move long-distance. Recall that we have evidence from the IP-absorption language Hungarian that the IP projections are not used as intermediate landing sites for successive-cyclic movement. Only the IP projections in the clause in which a *wh*-word takes scope are possible landing sites. The evidence has to do with the behavior of preverbal particles in Hungarian. These particles obligatorily follow the verb when *wh*-movement to an IP projection takes place:

Hungarian (adapted from Kiss 1994, 21, 37)

(128) a. Mari felhívta Jánost

Mary PREV-called John

‘Mary called up John’

b. Ki hívta fel Jánost?

who called PREV John

‘Who called up John?’

Long-distance *wh*-movement in Hungarian, however, does not trigger particle inversion on verbs along the path of movement; in (129), the particles *ki* and *be* remain in their preverbal positions, as Horvath (1985) points out:

Hungarian

(129) János melyik fiúnakj gondolta hogy Péter kijelentette

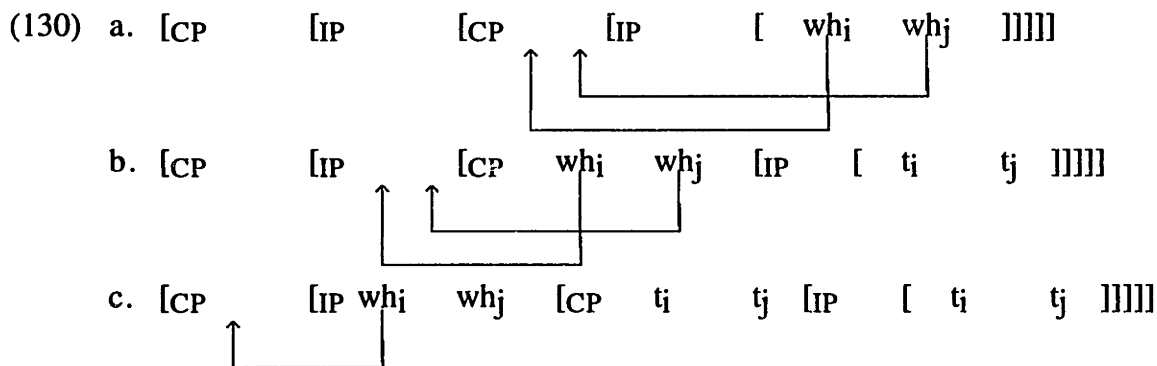
John which boy-to thought that Peter PREV-reported

hogy a házigazda már bemutatta Marit t_j?

that the host already PREV-showed Mary-ACC

'To which boy did John think Peter declared that the host had already introduced Mary?'

Multiple long-distance movement in IP-absorption languages, then, apparently proceeds as follows:



Here the two *wh*-words undergo a series of attractions to multiple specifiers of single heads, first moving into the specifiers of the embedded C⁰, and then to the matrix IP, and finally the higher *wh*-word is attracted into the matrix Spec CP. There are no Shortest violations in this derivation; the embedded IP in this case, being in a non-interrogative clause, is not a possible landing site, and the moves in (130a) therefore do not violate Shortest.

Now let us consider derivations in which the two *wh*-words take different scopes; one *wh*-word moves locally, and another moves long-distance. Here it will be convenient to divide the derivations into two classes. We have seen that *wh*-movement to Spec CP may use IP as an intermediate landing site; this was taken to be the reason for the lack of


local Superiority effects in IP-absorption languages. On the other hand, we have no data bearing on the question of whether movement to IP is actually obligatory for *wh*-words which are to move on to CP. In principle, one might expect it to be possible to skip IP and move directly to CP. Let us first consider derivations, however, in which this option (if it is an option) is not taken, and local *wh*-movement to CP proceeds via IP.

Such derivations will never allow two *wh*-words to take distinct scopes. Consider the base structure for such a derivation, given in (131):

(131) [CP [IP [CP [IP [*wh_i* *wh_j*]]]]]]


Recall that we are currently considering derivations in which the *wh*-word which is to take local scope moves into the lowest IP as its first step. It will not matter which of the two *wh*-words undergoes this move; let us assume, for simplicity's sake, that it is the higher of the two:

(132) [CP [IP [CP [IP [*wh_i* *wh_j*]]]]]]



Next the embedded C⁰ is merged and begins attracting *wh*-words. The *wh*-word which was attracted into the lower IP (*wh_i*) must be attracted first into Spec CP, since it is now the higher of the two *wh*-words:

(133) [CP [IP [CP [IP *wh_i* [*t_i* *wh_j*]]]]]]



Because *wh_i* is the one which moved into IP, it is the *wh*-word which will receive local scope; recall that a *wh*-word may only move into the IP of the clause in which it bears scope. The only *wh*-word which can move at this point in the derivation, then, is *wh_j*, the *wh*-word which is attempting to take matrix scope. All the potential moves which might be available to *wh_j* are ill-formed, however. It cannot move into the embedded IP, not only because this move would be anti-cyclic but because movement into the embedded IP is inconsistent with taking matrix scope, as we have seen. If it moves directly into the matrix



IP or CP, both it and its attractor violate Shortest. Skipping over the potential landing sites in the embedded IP-level projections qualifies as a Shortest violation on the part of the moving wh-word, and cannot be improved by the PMC, as noted earlier (this case crucially differs from the multiple long-distance movement case discussed above in that the embedded clause is interrogative, and the IP-level projections are thus in principle capable of being landing sites for wh-movement). Moreover, the attractor which attracts wh_j into the matrix clause from its present position violates Shortest, since wh_j is a closer potential mover. Finally, wh_j might move first into the embedded CP, and then into the matrix clause. The first step of this move will be a violation of Shortest by the wh-word, since it skips the embedded IP. Since the embedded CP already has one specifier occupied by wh_i, movement of wh_j will be forced by Shortest to be to a lower specifier. Further movement into the matrix clause will therefore violate Shortest on the part of the attractor, since wh_i is a higher possible attractee than wh_j even after both wh-words have moved to specifiers of the embedded CP. The chain created by movement of wh_j into the matrix clause, then, will unavoidably involve violations of Shortest both by the attractor and by the wh-word.

Generally, any derivation in which one of the two wh-words begins by moving into an IP-level projection in the embedded clause will be ill-formed. Because the lower IP is “active” as a potential landing site, whichever wh-word is moving to get wide scope will unavoidably skip a potential landing site (namely, the lower IP) in doing so. If the derivation begins by movement of the wh-word which is to receive narrow scope into the embedded IP, this wh-word will always c-command the wh-word which is to receive wide scope, and any attractor attracting the latter wh-word will therefore violate Shortest as well, since the wh-word which takes narrow scope will always be a potential closer attractee.

Both the attractor and the mover for the longer move in these derivations therefore violate Shortest.

Let us now consider derivations in which no *wh*-word moves into the local IP-level projection. Recall that we have no evidence that such a derivation is possible at all. It is clear that IP-absorption languages allow movement to a local Spec CP via an IP-level projection; what is not clear is whether the intermediate step of landing in an IP-level projection is required or merely an option. For some speakers, the only readings for sentences involving multiple clausemate *wh*-words are the ones which we have seen already to be unproblematic, in which both *wh*-words have the same scope. It may be that for these speakers, at least, derivations in which local movement does not proceed via IP are ill-formed.

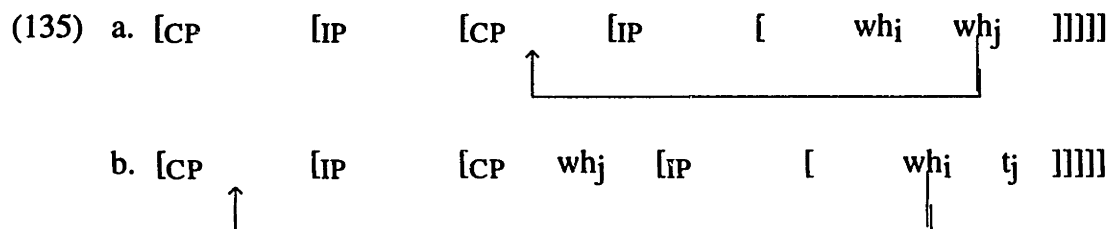
Bearing this in mind, let us move on to consider the derivations in question. There are four derivations to consider. One is the derivation that is comparatively well-formed in English; the lower C^0 attracts the higher of the two *wh*-words, and then the matrix C^0 (or I^0 ; the difference will be unimportant in what follows) attracts the lower:

- (134) a. [CP [IP [CP [IP [*wh_i* *wh_j*]]]]]

 b. [CP [IP [CP *wh_i* [IP [*t_i* *wh_j*]]]]]


As was noted above, this derivation is less well-formed in an IP-absorption language than it is in English. In English, (134b) involves a violation of Shortest only by the matrix C^0 ; *wh_j* does not violate Shortest, since the PMC marks the embedded C^0 as invisible for purposes of Shortest. In IP-absorption languages, both the matrix C^0 and *wh_j* violate Shortest in (134b). The embedded IP, as an IP in an interrogative clause, is a potential

landing site for wh-movement, and movement past it therefore violates Shortest. The move in (134a) also violates Shortest, for the same reason.

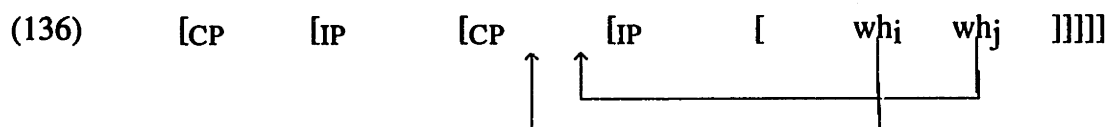
For completeness' sake, we should also consider a comparable derivation with crossing paths, given in (135):



(135) is clearly even worse than (134); both of the moves, in this case, violate Shortest both for the attractors and for the moving wh-words.

Now let us consider derivations using multiple specifiers of the embedded CP. We saw in section 1.8 of the preceding chapter that even IP-absorption languages may have multiple specifiers of CP at some point in the derivation, as long as no more than a single specifier remains at the end of the derivation. In chapter 4 I claimed that this indicated that all the wh-features on C⁰ in these languages except one are weak, and thus incompatible with a specifier occupied by an overt NP at PF.

The derivation should begin, then, as in (136):



(136) is also the beginning of the well-formed derivation beginning with this representation in CP-absorption languages. In an IP-absorption language, (136) is more problematic than it would be in a CP-absorption language; both of the moving wh-words in (136) violate Shortest, since they both skip the embedded IP. The attractor, however, does not violate Shortest in either case. Next, one of the wh-words must move into the higher CP; as in a CP-absorption language, Shortest requires it to be the higher of the two wh-words:

- (137) [CP [IP [CP wh_i wh_j [IP [t_i t_j]]]]]
-

The best derivation in an IP-absorption language which yields distinct scopes for clausemate wh-words, then, is the same as the best derivation in a CP-absorption language; both wh-words move directly to multiple specifiers of CP, after which the higher wh-word moves into the higher clause. This is the only derivation in which there is no chain in which both the attractor and the head of the chain violate Shortest. The derivation yields crossing paths, as desired.

We have seen that crossing paths in IP-absorption languages are available only for some speakers; there are other speakers who reject altogether sentences in which clausemate wh-words do not have the same scope. The account given above makes even the “best” derivation ill-formed in a variety of ways, as we should hope. Both of the moving wh-words in (136), as noted, violate Shortest, which is not the case in a CP-absorption language. The wh-word which takes local scope skips the local IP-level projection, an operation which cannot be independently demonstrated to be possible; we know that movement to Spec CP via IP is an option, but we do not know whether it is the only option. The derivation exhibits a “look-ahead” property of a type which seems to be disfavored in syntax, in two respects. The failure of the wh-word which takes local scope to move into IP is crucial for the well-formedness of the derivation, as we have seen, since it renders the movement of the other wh-word directly into CP a violation of Shortest only for the wh-word, rather than for both the wh-word and the attractor. Also, when the two wh-words land in the embedded Spec CP, the first wh-word which lands there (the one which, after both wh-words have moved, is in the higher specifier) must check a weak feature, rather than the strong feature which would permit it to remain in situ. If it were to check the strong feature, it would have to remain in situ, and the other, lower wh-word

would have to be attracted into the matrix clause; the attractor would thereby violate Shortest. Thus, the computational system must “foresee” that this wh-word will move on. No such foreknowledge is involved in the CP-absorption case, for instance, in which all of the wh-words in the specifiers of CP are checking the same kind of feature, and all of them could in principle remain in their positions if no features were introduced to attract them further. It is presumably this “look-ahead” quality which rules out this derivation in English, which (on the theory developed in chapter 4) should also in principle allow multiple specifiers of CP during the derivation, as long as they do not persist to its end. In IP-absorption languages, however, there are no better alternatives; in particular, the derivation used in English is completely ruled out by Shortest, which is violated by both the moving wh-word and the attractor, rather than only by the attractor as in English.

In this section I have compared a number of possible derivations in different languages, and have tried to show that the “best” derivations are the ones which are in fact found. The PCC effect in English, on the account given above, follows roughly as it is derived by Kitahara (1994, 1997). Given that only one landing site per wh-word is available in the overt syntax, a derivation in which each wh-movement tries to avoid violation of Shortest as much as possible will yield nested paths. The derivation which gives nested paths involves only a single violation of Shortest, this being incurred when the longer of the two moves takes place over the intervening wh-word in Spec CP of the embedded clause. This is apparently regarded as preferable to moving a single wh-word through both CPs; I have suggested that this has to do with a general ban on looking ahead in the derivation, which prevents the computation from “seeing” that the lower CP will be able to attract a second wh-word at LF.

In languages in which all wh-movement occurs on a single level, there are more landing sites available at once. As a result, it is easier for an attractor to obey Shortest without depriving the lower CP of its wh-word and harder for a moving wh-word to avoid

the effects of Shortest via the PMC. In CP-absorption languages, Shortest can be obeyed perfectly throughout the derivation; both *wh*-words can move to the embedded CP, and the higher of the two can then move further into the higher CP. The comparable derivation is ruled out in English by the ban on removing the only specifier of the embedded CP. This ban fails to apply in the CP-absorption case, since the embedded CP has multiple specifiers at the point in the derivation at which the higher C^0 needs to attract a *wh*-word. If the higher C^0 is rendered immune to Shortest by the presence of an additional *wh*-word, then both the derivation with crossing paths and the one with nested paths become in principle possible, and Fewest Steps prefers the one with nested paths.

Finally, in an IP-absorption language, the multiplicity of possible landing sites makes it impossible for the PMC to render them all invisible for purposes of Shortest Move. As a result, the nested-paths derivation, which in English yields a Shortest violation only by the matrix attractor, violates Shortest both for the attractor and for the moving *wh*-word. The only derivation which does not violate Shortest in this way and which gives the two *wh*-words distinct scopal positions is the one which is used in CP-absorption languages. This derivation involves considerably more “looking ahead” in IP-absorption languages than it does in CP-absorption languages, however, and is unsurprisingly less universally available.

In IP-absorption languages, derivations in which the *wh*-words always move as soon as possible to the closest available landing sites, checking strong features whenever possible and as early in the derivation as possible, will give multiple clausemate *wh*-words the same scope. This is not true in CP-absorption languages, or in English, where multiple clausemate *wh*-words may have distinct scopes without any need to “delay” performing an immediately available operation in order to make a later operation possible. The only case outside of the IP-absorption languages which seems to crucially involve global evaluation of multiple derivations, in the account just given, is the emergence of PMC effects in CP-

absorption languages when an additional wh-word renders the higher attractor immune to Shortest. I attributed this effect to Fewest Steps, a global condition.

2.6.2.2.2 Volvo-sentences

There is at least one empirical difference between the Path Containment Condition developed in Pesetsky (1982) and the similar effect which is derived here. Pesetsky claims that in cases of nested paths, only the outer path can contain a second variable which is c-commanded by the heads of both chains but c-commands neither of the tails. He reports the following contrast (Pesetsky (1982 (426-7))):

(138) a. ?This Volvo is one car OP_i that I know who_j to persuade

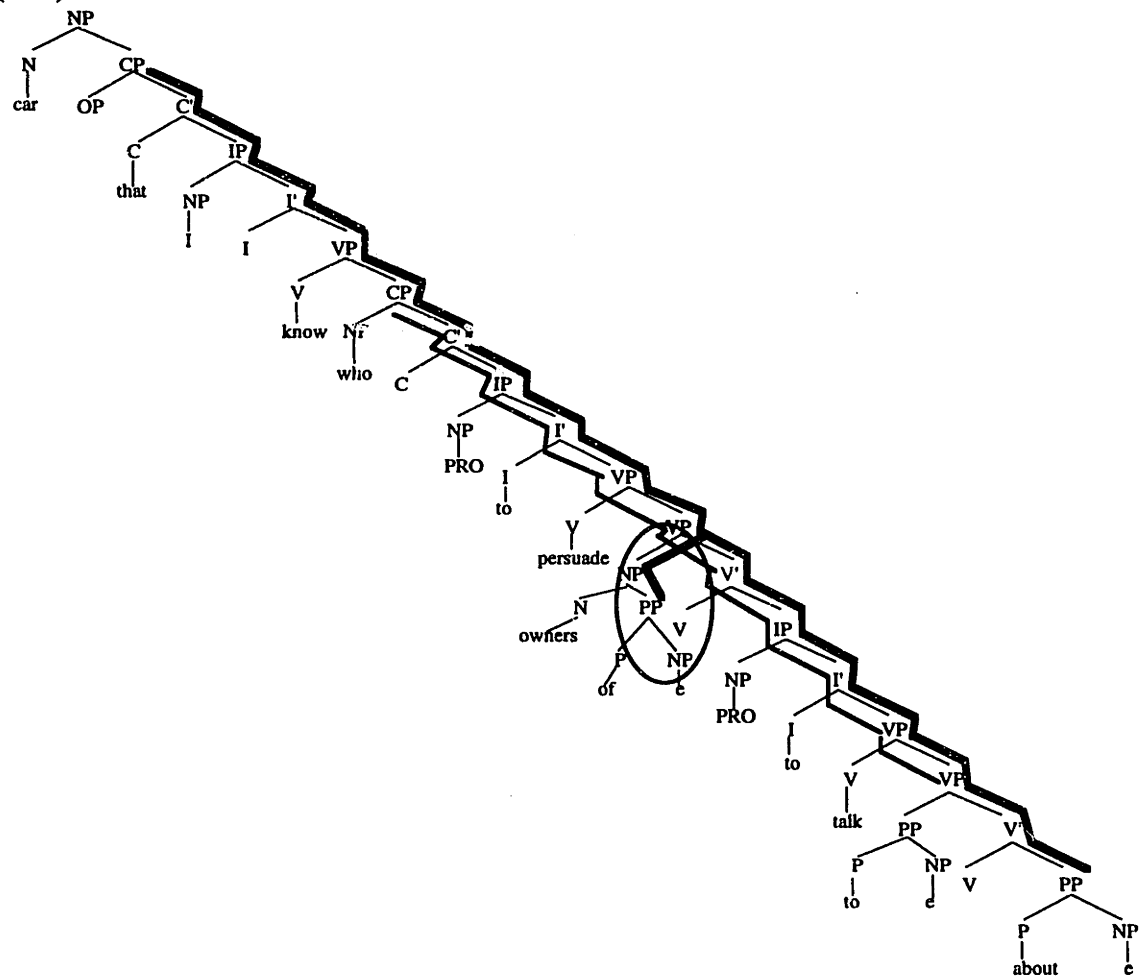
[owners of e_j] to talk to e_j about e'_i

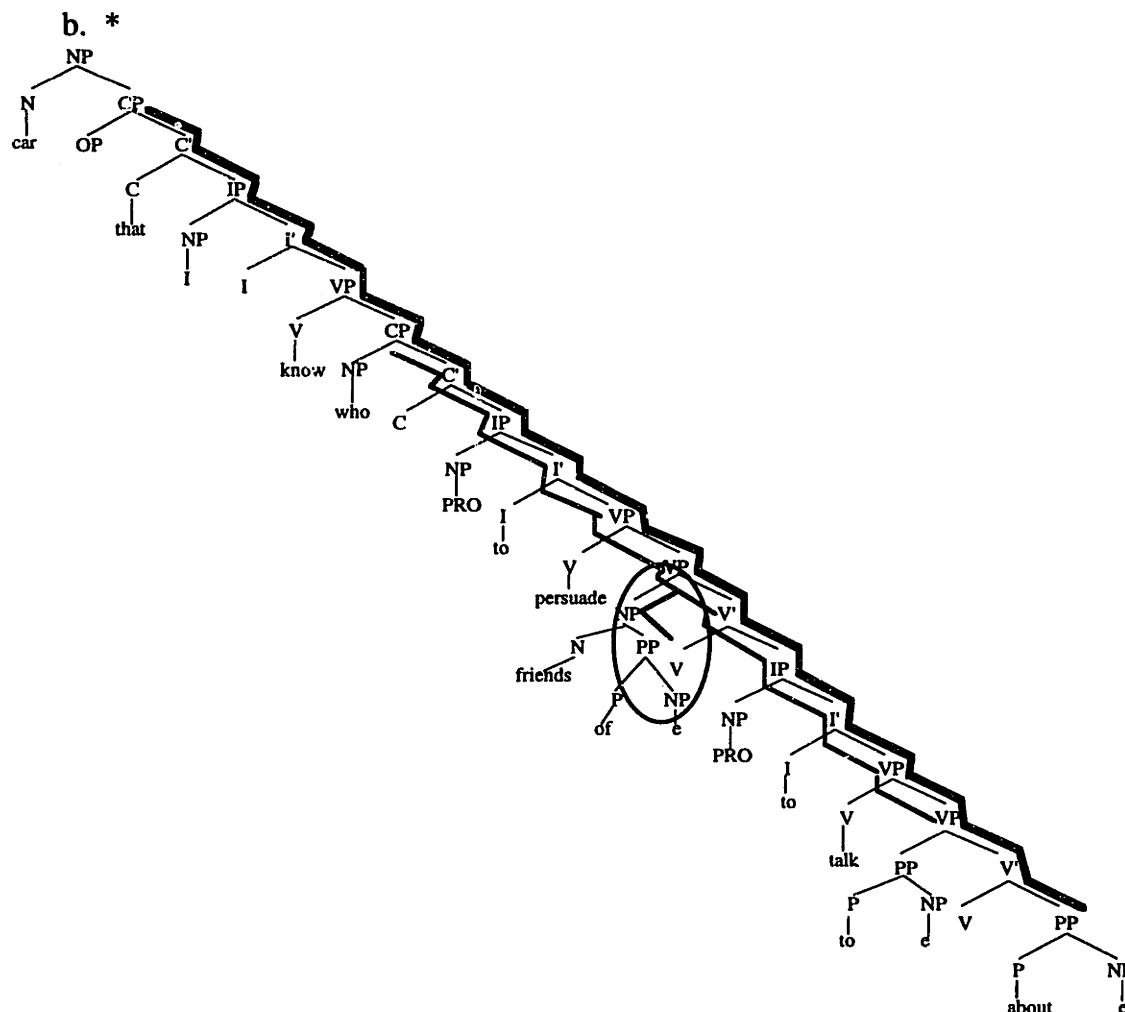
b. *This Volvo is one car OP_i that I know who_j to persuade

[friends of e_j] to talk to e'_j about e_i

This contrast follows from Pesetsky's (1982) theory. His Path Containment Condition requires that when two paths intersect, the syntactic nodes crossed by one of the paths must be a subset of the nodes crossed by the other. Cases like (131) involve a set of nodes (those contained in the object of *persuade*) which are crossed by only one of the intersecting paths (that is, only one of the two binders binds a variable embedded within the object of *persuade*). If these nodes are associated with the the path headed by *who*--that is, with the path whose nodes must be a subset of those of the other path--neither path will be completely contained in the other, and the Path Containment Condition will be violated. If the nodes are associated with the path headed by *OP*, on the other hand, the PCC is obeyed:

(139) a.





Thus, the contrast in (138) is predicted by Pesetsky's (1982) theory.

On the other hand, no such contrast is expected on the theory developed here.

Consider again the sentences in (138), repeated as (140):

(140) a. ?This Volvo is one car OP_i that I know who_j to persuade

[owners of e_i] to talk to e_j about e'_j

b. *This Volvo is one car OP_i that I know who_j to persuade

[friends of e_j] to talk to e'_j about e_i

The theory as we have developed it so far predicts that both of these sentences should be equally well-formed. All of the dependencies involved maximally obey Shortest, some of them via the PMC. In (140a), the relation between who_j and e_j obeys Shortest, as there

are no elements which could participate in this move along the path of the dependency (recall that e_i does not interfere, since it does not c-command e_j); this renders the attractor responsible for movement of *who_j* invisible for Shortest, and the relations between Op_i and its variables then satisfy Shortest modulo the PMC (although the attractor involved does violate Shortest, as we have seen). In (140b), the relations between *who_j* and its two variables both satisfy Shortest, and the PMC thus allows the relation between Op_i and its variable to obey Shortest as well. This theory thus fails to predict the contrast reported by Pesetsky between (140a) and (140b).

In fact, however, it is not clear that this contrast is a syntactically conditioned one.

Consider (141-142):

- (141) a. ?This Volvo is one car Op_i that I know *who_j* to persuade
[owners of e_j] to talk to e_j about e'_i
b. *This Volvo is one car Op_i that I know *who_j* to persuade
[friends of e_j] to talk to e'_j about e_i
- (142) a. *This is something Op_i that I know which grad student _{i} to
persuade [owners of e_j] to talk to e_j about e'_i
b. ?This is something Op_i that I know which grad student _{i} to
persuade [friends of e_j] to talk to e_j about e'_i

The sentences in (142) are structurally identical in the relevant respects to those in (141), yet the judgments are reversed. The fact that the PMC fails to predict the contrast in (141), then, is not an argument against it; apparently, there is nothing intrinsically ill-formed about the structure in (141b). As far as the syntax is concerned, either binder may bind the variable contained in the object of *persuade*, just as the PMC predicts.

What does account for the contrasts in (141-142)? One possibility is that the relevant factor is a processing strategy which prefers whenever possible to associate gaps with a binder which is particularly conversationally salient. In (141), the two possible

binders are the wh-word *who* and a relative operator which is associated with *this Volvo*. If we understand the latter as being more “prominent” in some way which is relevant to processing, the preference for using it to bind the extra variable might be explained. Similarly, in (142), the two binders are *which grad student* and an operator associated with *something*. Again, since we are given more specific information about the former binder, we should perhaps not be surprised to find that the reading in which it binds the extra variable is preferred.

Pesetsky (1982) also discusses a second class of Volvo-sentences which contain a total of four variables (from Pesetsky (1982, 430-1)):

- (143) a. ??This Volvo is one car OP_i that I know who_j
 [people that talk to e_j about e_i] can persuade e'_j to buy e'_i
 b. *This Volvo is one car OP_i that I know who_j
 [people that talk about e_i to e_j] can persuade e'_j to buy e'_i

Here, as Pesetsky points out, the Path Containment Condition must effectively be obeyed both in the subordinate clause complement of *persuade* and in the relative clause modifying *people*. Depending on how we interpret Shortest, the PMC will lead us to expect this effect as well; here, we apparently need to understand Shortest as being satisfied in cases where several variables are bound by a single binder only if all the binder-variable links satisfy Shortest. That is, Shortest might be restated as in (144):

(144) *Shortest*

A pair P of elements $\{\alpha, \beta\}$, where α and β are (possibly singleton) sets of coindexed elements, obeys Shortest iff there is no well-formed pair P' which can be created by substituting γ for either α or β , and the set of nodes c-commanded by one element of P' and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominating the other.

In (143a), then, the relations between *who*_j and its variables both satisfy Shortest, thus rendering *who* ignorable for Shortest and allowing construction of variable-binding links that cross the position of *who*. In (143b), on the other hand, the links between the binders and their variables within the relative clause both violate Shortest, and the structure is ill-formed. Changing the binders has no effect in this case, which is to be expected, as this particular phenomenon is genuinely syntactic in nature:

- (145) a. ??This is something OP_i that I know which grad student_j
 [people that talk to e_j about e_i] can persuade e'_j to buy e'_i
 b. *This is something OP_i that I know which grad student_j
 [people that talk about e_j to e_i] can persuade e'_j to buy e'_i

This phenomenon bears an obvious similarity to the behavior of Subjacency discussed above; as Chomsky 1986 notes, Subjacency must be obeyed both in the main clause and by the parasitic gap:

- (146) a. *Who_j did John hire t_j without asking who had interviewed t_j?
 b. *Who_j did John ask before hiring t_j who had interviewed t_j?

In section 2.5.2, I suggested that the ill-formedness of the sentences in (146) might be handled by stating Subjacency as a constraint on a +wh complementizer's relations to all wh-words with a given index, rather than simply to a given wh-word:

- (147) a. An attractor must trigger a dependency consisting of a (possibly singleton) set of well-formed links which share an index.
 b. A well-formed link must consist of a head α and a tail β such that there is no γ , γ a member of some set of barriers (including wh-islands, complex noun phrases, adjuncts, and subjects), such that α c-commands γ and γ dominates β .

As in the case of Shortest, then, the suggestion is that a set of coindexed dependencies ought to undergo the constraints on dependencies as a unit (as in Kayne's (1983a) notion of G-projection), rather than applying the constraints to each dependency separately²⁴.

Another possibility, following Chomsky 1986, would be to claim that parasitic gaps always involve a second operator base-generated in the position of the parasitic gap. The fact that this operator must obey Shortest, then, is not surprising. Such a claim obviously involves giving up on the account developed here of the Connectedness effects. The behavior of the Path Containment Condition provides us with an argument against this approach. Consider (146b), repeated as (148):

(148) *This Volvo is one car OP_i that I know who_j

[OP'_i OP'_j people that talk about e'_i to e'_j] can persuade e_j to buy e_i

We are now considering an approach according to which sentences like (141) contain four operators, rather than two. The operators OP'_i and OP'_j inside the subject NP, which are responsible for the creation of the parasitic gaps, violate the Path Containment Condition, and the structure is therefore ill-formed. If we develop an account of this kind, however, we will need to add a stipulation to ensure that the parasitic-gap-operators do indeed violate the PCC. There is no obvious reason for this to be so; that is, there is no reason for (148) to have the structure in (149a) rather than that in (149b):

(149) a. This Volvo is one car OP_i that I know who_j

[OP'_i OP'_j people that talk about e'_i to e'_j] can persuade e_j to buy e_i

b. This Volvo is one car OP_i that I know who_j

[OP'_j OP'_i people that talk about e'_i to e'_j] can persuade e_j to buy e_i

²⁴ This might be implemented either by understanding Subjacency and Shortest Move as representational constraints (but see section 2.5.1 for an argument against this), or by assuming a form of across-the-board movement in parasitic gap constructions, so that derivational versions of Subjacency and Shortest Move can apply to the movement relations connecting a single binder to its multiple variables.

The PCC is violated in (149a), but not in (149b). An account along these lines must apparently stipulate that the parasitic-gap operators must enter into the same hierarchical relation to each other as the operators on which they are parasitic. On the account developed here, on the other hand, (148) has only two operators, rather than four. The fact that the operators binding the *e'* variables and the operators binding the *e* variables must be in the same hierarchical relation to each other is now unsurprising, as these operators are identical. The particular redefinition of Shortest developed in this section, and the parallel definition of Subjacency developed above, thus seem well-motivated.

3 Some possible further expansions

In this section I will briefly discuss some other facts which appear to be cases of PMC effects. I will be unable to give full accounts of the phenomena in question, however.

3.1 Stylistic Fronting

Icelandic exhibits a phenomenon known as “stylistic fronting” (cf. Platzack 1987, Maling 1990, Rögnvaldsson and Thráinsson 1990, Jónsson 1991, Poole 1996 for discussion). Various kinds of heads in Icelandic can front past the tensed auxiliary, just in case the subject has undergone some kind of movement. Fronting can affect participles, predicate adjectives, or certain kinds of adverbs²⁵, including negation:

²⁵It is not clear to me at this point why adverbs should behave this way, on the common assumption that adverbial heads do not participate in head-movement. Perhaps this is (rather weak) evidence that the common assumption is wrong.

- (150) a. Hún benti á ýmsa roskna trésmiði í bænum,
 she pointed to various old carpenters in town
 sem byrjaði höfðu t; með engum efnum.
 that begun had with nothing
- b. Þetta er glæpamaðurinn sem ekki hefur t; verið dæmdur.
 this is criminal-the that not had been convicted
- c. Hann fann mynd sem vandlega hafði t; verið falin.
 he found picture that carefully had been hidden

Stylistic Fronting is only possible if the subject has undergone some kind of movement; it cannot appear in its in-situ position, as in (150a), but can be extracted, as in (150b), or left in a lower position in the tree, as in (150c), a possibility which is only open to indefinites. Stylistic fronting is also possible with impersonal passives; here the subject is phonologically null, so its position is difficult to determine:

- (151) a. * Hún benti á bæinn, þar sem byrjað; höfðu t;
 she pointed to town where that begun had
 trésmiðirnir með engum efnum.
 carpenters-the with nothing
- b. Hún benti á ýmsa roskna trésmiði í bænum,
 she pointed to various old carpenters in town
 sem byrjað; höfðu t; með engum efnum.
 that begun had with nothing
- c. Hún benti á bæinn, þar sem byrjað; höfðu t;
 she pointed to town where that begun had
 nokkrir trésmiðirnir með engum efnum.
 some carpenters with nothing
- d. Það var hætt að rigna þegar komið; var t; þangað.
 it was stopped to rain when arrived was thither

Thus, it would appear that well-formed movement of the subject relaxes somewhat the conditions on head-movement, licensing a head-movement which would be ill-formed in isolation.

If a sentence contains multiple heads that could in principle undergo Stylistic Fronting, only the highest can be fronted (Jónsson 1991, 5):

- (152) a. Þetta er aðferð sem ekki; hefur t; verið reynd á mönnum.
 this is method that not has been tried on people
- b. * Þetta er aðferð sem reynd; hefur ekki verið t; á mönnum.
 this is method that tried has not been on people

In other words, extraction of the subject seems to render the inflected auxiliary transparent for head-movement, but not any of the lower heads. In terms of the theory being

developed here, movement of the subject apparently is not rendering an attractor immune to Shortest; if that were the case, both of the examples in (152) would be well-formed. Rather, the inflected auxiliary is being made invisible, either as a potential attractee or as a potential landing site for movement of the lower head.

Exactly how the inflected auxiliary is rendered invisible is somewhat mysterious, however. Thus far, the elements which are rendered invisible by the PMC have been either direct participants in well-formed movement relations (in particular, attractors, or landing sites) or islands in a local relation with a well-formed movement. This is not in any obvious sense a case of either of these. One possible approach would be to extend the notion of “participant” in an attraction relation, capitalizing on the specifier-head relation between the moved subject and the auxiliary which is skipped. This approach would involve saying that the attraction relation which causes movement of the subject is also in effect an attraction relation with the head of which the subject is a specifier. By virtue of having participated in a well-formed attraction relation, the auxiliary is now invisible for purposes of Shortest, and can be skipped by movement of other heads.

3.2 Holmberg’s Generalization

Holmberg’s (1986) Generalization is essentially the converse of the Stylistic Fronting case just discussed. Here movement of the verb (corresponding to Y^0 , above) licenses movement of the object past some obstacle (perhaps the base-position of the subject, as in Chomsky (1993)) (Icelandic, from Jonas and Bobaljik (1993:93)):

- (153) a. * Jólasveinarnir hafa búðinginnj ekki borðuðu t_j.

Christmas-trolls have pudding-the not eaten

'The Christmas-trolls haven't eaten the pudding'

- b. Jólasveinarnir borðuðu; búðinginnj ekki t_j t_j.

Christmas-trolls ate pudding-the not

'The Christmas-trolls didn't eat the pudding'

That is, well-formed head-movement can apparently license an XP-movement which would be ill-formed in isolation; this is reminiscent of the Stylistic Fronting case, in which well-formed XP-movement licenses overly long head-movement. Theories in which head-movement renders the specifier of the tail of the head-movement chain transparent for Shortest Move include Chomsky's (1993) notion of Equidistance, and Baker's (1988) Government Transparency Corollary. We might try to derive this effect via the PMC. As in the Stylistic Fronting case, the account might crucially refer to properties of the specifier-head relation, making the Attract relation with the moved head a relation not only with the head but with its specifier as well. We should expect such an operation to make the specifier of the moved head transparent for purposes of Shortest, allowing attraction past this specifier, as desired.

3.3 Lower-wh effects

In section 2.6 of this chapter we considered a number of additional-wh effects, in which an ill-formed wh-movement could be rendered well-formed by the addition of another wh-word. In some cases, the additional wh-word c-commanded the offending wh-word; in others, it was c-commanded or dominated by an offending island. The former

cases were taken to be cases in which an attractor is rendered impervious to a constraint, while the latter were cases of islands being made transparent.

Kayne (1983a) and Pesetsky (1982) discuss cases of additional-wh effects in which the additional wh-word is c-commanded by the offending wh-word's base position:

(154) a. ?*What books do you expect who to give t to Bill?

b. What books do you expect who to give t to whom?

In (154b), the additional wh-phrase *to whom* improves the status of a Superiority violation. Watanabe (1991) and Saito (1994a) discuss similar data for Japanese Anti-Superiority, which also exhibits an additional-wh effect with lower wh-words:

Japanese (Saito 1994a, 231)

(155) a. *Naze dare-ga soko-ni itta no?

why who NOM there to went Q

'Who went there why?'

b. ?Naze dare-ga doko-ni itta no?

why who NOM where to went Q

'Who went where why?'

Interestingly, this effect appears to be a property of covert movement; that is, it is only in languages in which some or all wh-movement is covert that this effect appears. In Bulgarian, as we have seen, the second and third of three wh-words may be attracted in either order:

Bulgarian (Bošković 1995a, 13-14)

(156) a. Kogo kakvo e pital Ivan

whom what AUX asked Ivan

‘Who did Ivan ask what?’

b. ?*Kakvo kogo e pital Ivan

what whom AUX asked Ivan

(157) a. Koj kogo kakvo e pital

who whom what AUX asked

‘Who asked whom what?’

b. Koj kakvo kogo e pital

Thus, the presence of the higher wh-word *koj* in (157) eliminates the Superiority effect seen in (156) with regard to placement of the other two wh-words. However, no lower-wh effect appears in this case; the first and second wh-words are still strictly ordered:

Bulgarian (Roumyana Izvorski, p.c.)

(158) * Kogo koj kakvo e pital

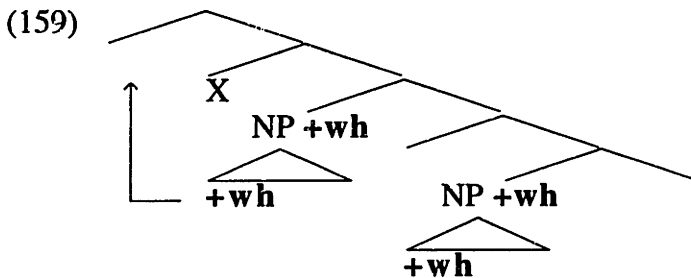
whom who what AUX asked

‘Who asked whom what?’

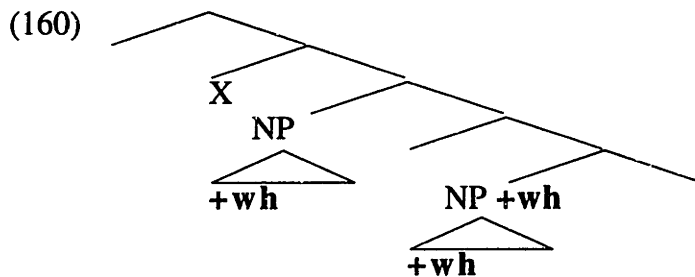
We clearly do not want to say that the lower wh-word is rendering the attractor immune to Shortest; on the version of the PMC developed here, this would force us to the conclusion that covert movement of the lower wh-word precedes Superiority-violating overt movement, and this conclusion is inconsistent with our previous assumptions about the relative timing of overt and covert movement. We are thus forced to the conclusion that lower-wh effects involve the elimination of some kind of island. Essentially following Kayne (1983a), then, we draw a parallel between lower-wh effects and the conditions on parasitic gap formation; both phenomena involve an additional wh-word in a position c-

commanded by the offending island, which renders the island transparent for extraction. There are a number of possible accounts which would have this property. I will simply sketch one here.

Suppose that the *wh*-feature contained in a *wh*-word or phrase typically percolates or is projected up to the highest node in the NP (the label for this node is unimportant; I will refer to it as NP). This has the result that the *wh*-feature of a *wh*-NP has the same c-command domain as the NP itself. If the notion of “closest possible attractee” relevant for Shortest makes crucial reference to the c-command domain of features, then Attract will always single out that *wh*-feature which is not c-commanded by any other *wh*-features:



Suppose further that this percolation of the *wh*-feature to NP is in principle optional. In place of the structure in (159), then, a multiple-*wh* construction could have the structure in (160):



In (160), the *wh*-features are not in a c-command relation, and Shortest should be satisfied by attraction of either.

Furthermore, the *wh*-feature in the higher *wh*-word in (160) is more deeply embedded than the highest *wh*-feature in the higher *wh*-word in (159). Suppose that the

feature is in fact so deeply embedded that it is in an island, from which it cannot escape unless another wh-word renders the island transparent. This is the contribution of the lower wh-word, which makes the NP transparent for wh-feature extraction as it passes by it in the covert syntax.

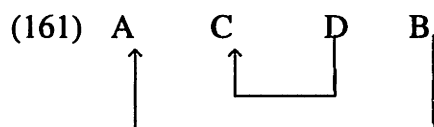
Many questions arise, which I will have to leave to further work. For instance, why can the second wh-word in (160) not render the island NP transparent by its movement in the overt syntax? Why is a third wh-word necessary? One possible answer to this question was developed in section 2.6.1.2 of this chapter, where we saw another case in which an overt wh-movement cannot render an island transparent for covert wh-movement. I suggested that the effects of the PMC might be temporary, vanishing once the well-formed wh-word is no longer in a local relation with the island. This answer drives us to an account in which movement of the “skipped” wh-word actually does occur covertly (and not, for instance, overtly but invisibly). Another question, which I will have to leave for future work, is why this phenomenon does not appear in Bulgarian.

4. Is the PMC recursive?

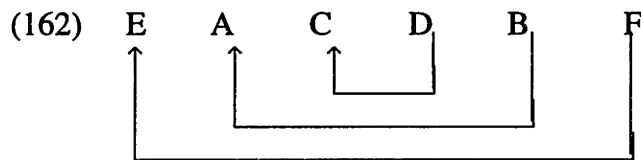
The basic effect of the PMC is to make dependencies that would be ill-formed in isolation well-formed in the presence of a well-formed dependency. A natural question arises as to whether this process is recursive: that is, does a dependency which is rendered well-formed by virtue of the PMC trigger the PMC for other dependencies? The answer would appear to be yes, as I will try to argue in the next few sections.

4.1 Path Containment Condition

Consider the PMC-based explanation of the PCC. According to the story developed here, the PMC improves the grammaticality of the movements in (161):



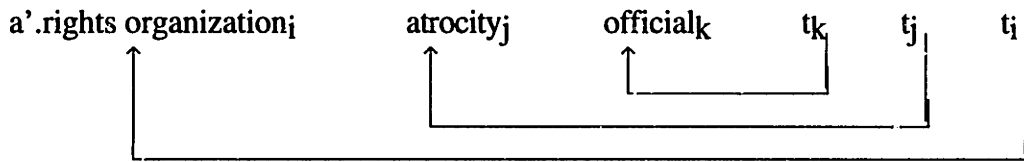
Here movement from D to C has made the landing site C invisible for purposes of Shortest, and movement from B to A therefore fails to violate Shortest. The link between A and B, in other words, is well-formed only by virtue of the PMC. Thus, this is a case in which we can investigate the question of whether the PMC is recursive. We can do this by adding a third movement path with a head above A and a tail below B:



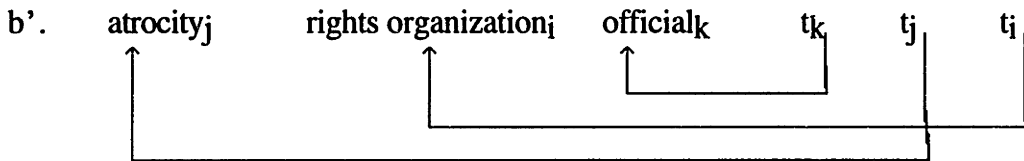
If the PMC is recursive, the well-formed path between A and B will render the landing site A invisible for Shortest, just as the path between C and D made the landing site C invisible, and the structure in (162) should be well-formed. If the PMC is not recursive, the move from E to F should be a Shortest violation; the link between A and B, which obeys Shortest only by virtue of the PMC, will not trigger the PMC and render itself invisible, and triple nested paths should be ill-formed.

In fact, there does seem to be a contrast between (163a) (with the structure in (163a')) and (163b) (with the structure in (163b')):

- (163) a. ?Amnesty International is one human rights organization [O_i that I wonder
[which atrocity_j this official would be easiest [O_k to talk to t_k about t_j on
behalf of t_i]]]



- b. *The annexation of East Timor is one atrocity [O_j that I wonder which human
rights organization_i this official would be easiest [O_k to talk to t_k about t_j on
behalf of t_i]]]



If there is in fact a contrast between (163a) and (163b), this argues in favor of a recursive PMC; if the PMC is recursive, (163b) will involve violation of Shortest which will not be present in (163a), as desired.

4.2 *that*-trace effects

Another argument for a recursive PMC can be constructed from the behavior of *that*-trace effects. Consider again the PMC-based account of the facts in (6), repeated as (164):

- (164) a. *Who_j do you think t'_j that t_j left?
b. Who_j do you think t'_j t_j left?
c. Who_j do you think t''_j that John said t'_j t_j left?

The account was based on a version of the principle in (165):

- (165) Certain traces must be antecedent-governed, and *that* blocks antecedent-government.

We assumed earlier, following Lasnik and Saito (1984), that all of the traces in (164) are of the type requiring antecedent government (in their terms, none of these traces are lexically governed). Antecedent-government of t by t' in (164a) is blocked by the presence of *that*, following (165), and the sentence is therefore ill-formed. In (164b), on the other hand, t' can antecedent-govern t , and *who* antecedent-governs t' ; the sentence is therefore well-formed. In a case like (164c), the relation between t' and t obeys (165), and as a result t' and t are ignorable with respect to (165). The relationship between t'' and t' thus effectively obeys (165), despite the fact that no antecedent-government is possible; t' participates in a dependency that obeys (165) and therefore need not participate in another to be well-formed. Thus, the relation between t'' and t' obeys (165) only because of the PMC. The well-formedness of (166), then, would seem to be another argument for a recursive PMC:

(166) Who_i do you think t''' _i that John said t'' _i that Mary thinks t' _i t _i left?

Here, the link between t' and t obeys (165), as before, and the link between t'' and t' obeys (165) because of the PMC²⁶. The link between t''' and t'' in fact disobeys (165), so the PMC is apparently capable of saving it. Thus, the link between t'' and t' , which obeys (165) because of the PMC, must trigger the PMC with respect to itself, rendering t'' invisible for purposes of evaluating the structure with respect to (165). In other words, the PMC is apparently recursive.

²⁶ Actually, strictly speaking, the PMC fails to apply to the link between t'' and t' , because t' is ignored. This is a rather serious problem; clearly, we do not always want to say that a constraint is obeyed when it fails to apply to a given dependency (for example, we do not want to say that well-formed head-movements can render an ill-formed link between an anaphor and its antecedent well-formed). There has to be some difference between a dependency to which a constraint genuinely does not apply and a dependency which "evades" a constraint as a result of the PMC. This might be an argument against the notion of "ignoring" parts of the structure; rather, we may need to understand the PMC as marking parts of the structure as having already obeyed a constraint (so that the constraint still applies, but is automatically satisfied).

4.3 Parasitic Gaps

Chomsky (1986) discusses a class of parasitic gaps which are licensed by parasitic gaps:

(167) Who_i did John ask t_i [whether we should talk to t'_i [before we hire t''_i]]

In (167), the dependency between *who* and t' violates a wh-island, but is licensed by the well-formed dependency between *who* and t; the PMC allows us to ignore the wh-island, which is along the path of the well-formed dependency, and renders the dependency between *who* and t' well-formed. The dependency between *who* and t'' violates an Adjunct Island, and will therefore also need help from the PMC. Here the saving dependency is apparently the one between *who* and t'. Note that the adjunct island is not along the path of the dependency between *who* and t; in the absence of t', the structure in (167) is ill-formed²⁷:

(168) *Who_i did John ask t_i [whether we should talk to Bill [before we hire t''_i]]

Thus, the relation between *who* and t'' is apparently saved by the dependency between *who* and t' in (167); that is, the dependency between *who* and t', which is saved by the PMC, also triggers the PMC. Again, the PMC appears to be recursive.

5. Conclusion

We have seen that we can unify a number of apparently diverse phenomena under the rubric of Minimal Compliance. It may be a general property of human language that constraints need not be satisfied perfectly in all parts of a given structure for that structure to be well-formed. Recognition of the Principle of Minimal Compliance as an independent

²⁷ (168) contrasts with (i):

(i) Who_i did John ask t_i [whether we should talk to his_i mother [before we hire t'_i]]

Here the dependency between *who* and the resumptive pronoun *his* appears to license the true parasitic gap t'_i. On this account, the relation between the wh-word and the resumptive pronoun must be regarded as a dependency which is subject to island constraints but always passes them, and is thus able to obviate the effects of islands for real gaps. This status for resumptive pronouns is reminiscent of the behavior of Japanese wh-words other than *naze* 'why' in section 2.6.1.2, which can obviate the effects of islands for *naze* by being placed inside the same island.

fact about human language thus promises to simplify considerably our statements about the permissible relations among syntactic elements.

If the theory sketched above is right, we are also driven by acceptance of the PMC to a particular point of view about the structure of the grammar. In particular, we must apparently view the grammar in derivational rather than representational terms; this seems to yield the best analysis of certain facts about Subjacency, as we have seen. More generally, PMC effects appear to be sensitive to the order in which syntactic operations take place, which suggests that they might be useful in investigating the extent to which syntactic phenomena crucially rely on properties of the derivation. Furthermore, as developed above, the PMC entails a view of the grammar in which it contains particular constraints that must be checked. The fact that the PMC interacts in different ways with different constraints (see sections 2.1 and 2.6, in particular, for some discussion of this) suggests that these constraints are in fact formally distinct in some way. I have encoded this distinction here in the statements of the constraints themselves; one can imagine other ways of attempting to encode it (for instance, by having more general constraints which have different effects on different structures, as suggested in footnote 14 and pursued to some extent in the unification of Shortest Move and Shortest Attract as a single constraint Shortest), but I will have to leave this problem, along with many others, for future research²⁸.

²⁸A reviewer of an article based on this chapter suggests that notions like the PMC might be relevant in other areas of the grammar as well. For instance, as the reviewer notes, Chomsky (1964, 1986:39) suggests that processes may only be able to apply once to a given domain, on the basis of the contrast in (i):

- (i) a.* Who did you wonder what John [gave t t]?
- b.?? To whom did you wonder what John [gave t t]?

Chomsky notes that (ia) is worse than (ib), and suggests that this might be due to the fact that (ia) involves multiple instances of the same category (NP) being extracted from the same phrase (VP) by the same process (wh-movement). If there is a general principle that processes can only apply once to a given domain, this might well subsume the PMC as it has been developed here. Another phenomenon which might fall under this general heading is the notion of feature checking, whereby a process makes a portion of the structure “invisible” to the computational component.

Chapter Six: Conclusion

In chapter 1 I discussed a number of theories about the architecture of the grammar. The thesis has largely been an effort to discover evidence that would help us to choose among these theories. One theory, the one which I assumed and which seems to most accurately capture the facts of multiple wh-movement across languages, is the “classic” theory which postulates both overt and covert wh-movement, and in which all overt movement precedes all covert movement; this is the theory defended in Huang (1982) and Lasnik and Saito (1984). As I pointed out, this theory has the property (unique among the theories discussed) that languages which move all their wh-words and languages which leave all wh-words in situ should have a similar status with respect to diagnostics for the order of operations in the derivation; these are the languages in which all movement takes place in a single component of the derivation. They contrast in this regard with English, which does some movement in the overt syntax and some movement in the covert syntax.

A second theory, proposed by Cole and Hermon (1994), Tsai (1994), and Reinhart (1995), denies the existence of covert movement, and argues that wh-in-situ is interpreted via a strategy which does not involve movement at all. I tried to argue against theories of

this type in chapter 2, showing that languages which leave all *wh*-words in situ may be divided into the two classes established by Rudin (1988) for multiple overt movement languages; I referred to these classes as IP-absorption and CP-absorption languages.

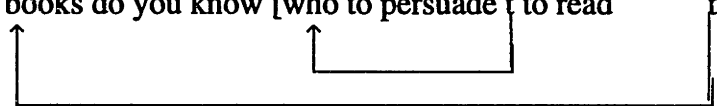
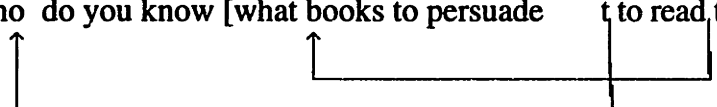
The third theory of the overt/covert distinction which I discussed in chapter 1 was that proposed by Bobaljik (1995), Brody (1995b), Groat and O'Neil (1996), and Pesetsky (to appear). This theory claims that the overt/covert distinction is a purely phonological one, having to do with which part of a chain is pronounced. In chapter 4 I discussed a theory of well-formed PF objects which is certainly consistent with theories of this kind; I argued that feature strength serves as a "signpost" instructing PF to pronounce the copy in a chain associated with the strong feature. These theories differ from the first class of theories mentioned above in that they fail to draw a syntactic distinction between multiple overt *wh*-movement languages and multiple *wh*-in-situ languages on the one hand and "mixed" languages like English on the other; the timing of *wh*-movement is the same in all three classes of languages. I have argued that there are in fact similarities between languages like Bulgarian and languages like Japanese which English does not share, and that these similarities are best explained in terms of the timing of *wh*-movement.

Finally, a fourth theory, that proposed by Watanabe (1992), is similar to the third theory just mentioned in that it allows for "overt" movement which appears for phonological reasons to be covert. Watanabe postulates this kind of movement in the course of claiming that the timing of *wh*-movement in English and Japanese is the same, the only difference between them being that overt *wh*-movement in Japanese is invisible for phonological reasons.

I have tried to show that there are several phenomena which group multiple overt *wh*-movement languages and multiple *wh*-in-situ languages together to the exclusion of "mixed" languages of the English type, and that the best account of these phenomena makes reference to the timing of *wh*-movement. One of these phenomena was the behavior

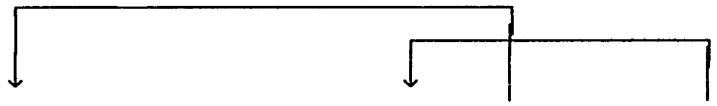

of Path Containment Condition effects in the various languages. English exhibits the Path Containment Condition, as is well known:

(Pesetsky 1982, 268)


- (1) a. what books do you know [who to persuade t to read t] ?

 b. * who do you know [what books to persuade t to read t] ?


Languages which do all their wh-movement at a single point in the derivation, on the other hand, exhibit anti-PCC effects. This is clearest in the case of what I have called the CP-absorption languages, where all speakers prefer crossing paths to nested paths for non-D-linked wh-words:

Bulgarian (Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, Kamen Stefanov, p.c.)

- (2) a. Koj se opitvat da razberat kogo t e ubil t ?
 who SELF try to find-out whom AUX killed

 b.* Kogo se opitvat da razberat koj t e ubil t ?
 whom SELF try to find-out who AUX killed


Chinese (Hooi Ling Soh, Lisa Cheng, Wei-Tien Dylan Tsai, p.c.)

- (3) jingcha xiang-zhidao [shei sha -le shei]
 police want know who kill PERF who

 'Who_j are the police trying to find out who_j t_i killed t_j?'
 (Note: The original image has a typo 'tj' which has been corrected to 't_j' for consistency with linguistic notation.)

In IP-absorption languages, many speakers do not allow clausemate wh-words to receive distinct scopes at all, but those who do also exhibit anti-PCC effects:

Serbo-Croatian (Rudin 1988, 459; Željko Bošković, Milan Mihaljevic, p.c.)

- (4) a. *Šta si me pitao ko može da uradi?

what AUX-2s me asked who can to do

‘What have you asked me who can do?’

- b. ?*Ko si me pitao šta može da uradi?

who AUX-2s me asked what can to do

‘Who have you asked me what can do?’

Japanese (Takako Aikawa, Satoshi Oku, Shigeru Miyagawa, p.c.)

- (5) Keesatu-wa [dare -ga dare -o korosita ka] sirabeteiru no?

police TOP who NOM who ACC killed Q are-investigating Q

- a. * ‘Who are the police trying to find out who t killed t?’

- b. ?? ‘Who are the police trying to find out who t killed t?’

I argued in section 2.6.2.2.1 of chapter 5 that the contrast between English on the one hand and languages like Bulgarian, Chinese, Japanese, and Serbo-Croatian on the other has to do with the fact that at the point in the derivation of these structures at which the higher C^0 is merged and must attract a wh-feature, the lower C^0 in English only contains a single wh-word, while in the other languages it may contain more than one wh-word. I suggested that attraction of the single wh-specifier of the lower C^0 would effectively render the lower C^0 declarative, causing its strong wh-feature to vanish after checking. In the languages other than English, on the other hand, the higher C^0 can attract a wh-word from the lower C^0 without depriving it entirely of wh-words.

A second contrast between English and the other languages had to do with a phenomenon I referred to as parasitic wh-movement, discussed in section 2.6.1.1 of chapter 5. In languages which do all their wh-movement at a single point in the derivation, there is a strategy for obviating islands which is unavailable in English; a wh-word may render an island which is along its path transparent for purposes of a second wh-extraction out of the island. Interestingly, this strategy is consistent with derivations in which the well-formed wh-movement lands in its scopal position after the ill-formed wh-movement:

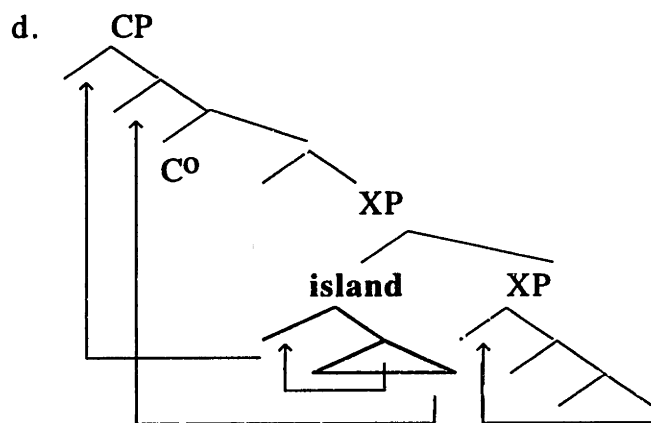
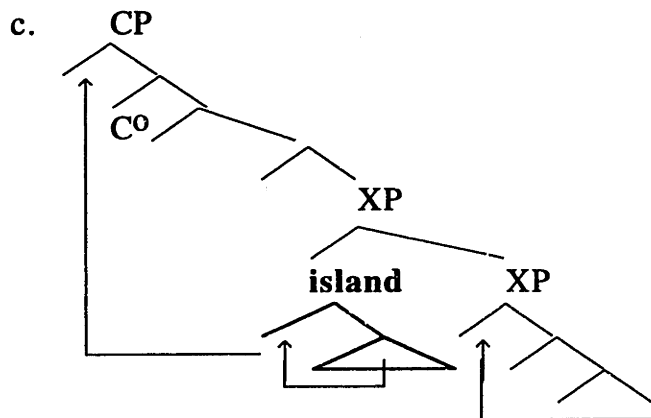
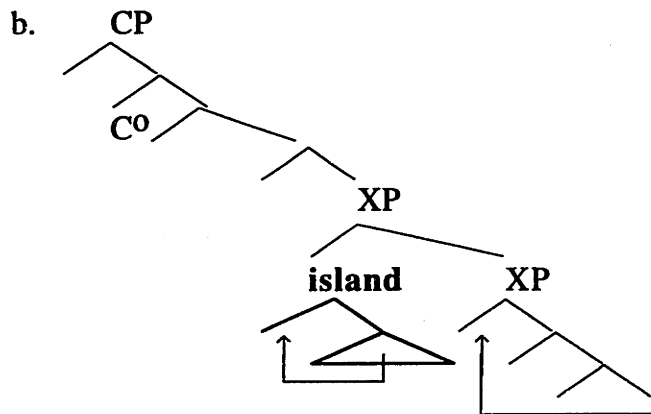
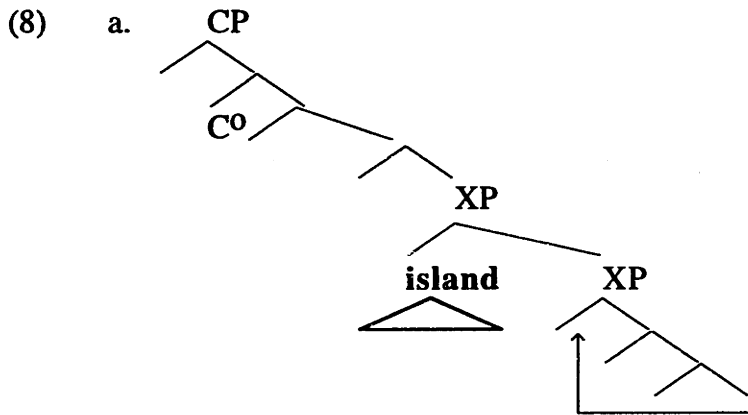
Bulgarian (Roumyana Izvorski, p.c.)

- (6) a. *Kakvoj kazva tozi služitel na [žurnalistite, kojto razsledvat tj],
 what tells this official to the-journalists who investigate
 če komunistite sa zabludili redaktorite im?
 that the-communists AUX deceived the-editors their
 ‘What does this official tell journalists who are investigating that the communists
 have deceived their editors?’
- b. ?Kakvoj kogož kazva tozi služitel na [žurnalistite, kojto razsledvat tj],
 what who tells this official to the-journalists who investigate
 če komunistite sa zabludili t_k?
 that the-communists AUX deceived
 ‘What does this official tell journalists who are investigating that the communists
 have deceived who?’

In (6b), the well-formed wh-movement of *kogo* lands in a position to the right of the ill-formed movement of *kakvo*. According to the theory developed in chapter 3, this means that *kakvo* has landed in Spec CP before *kogo* has. Such derivations are not available in English, where covert movement cannot license overt movement, arguably because covert movement follows overt movement in the derivation:

- (7) a. *Which car did you persuade [the man who bought t] to sell the hubcaps?
 b. *Which car did you persuade [the man who bought t] to sell which hubcaps?

To account for the contrast between languages like Bulgarian and Japanese on the one hand and English on the other, I made use of theories of wh-movement like those of Chomsky (1986), Saito and Fukui (1996), Agbayani (1997), and Fox (1997), among others, in which wh-movement proceeds via successive-cyclic adjunction to some or all of the maximal projections which dominate the extraction site and are c-commanded by the eventual landing site. What distinguishes Bulgarian and Japanese from English, on this view, is the availability of derivations in which the well-formed wh-movement stops in a position sufficiently local to the island to render it transparent, after which wh-movement can proceed out of the island and “skip” the stopped, well-formed wh-movement, landing in the landing site before the well-formed wh-movement does:



This derivation is unavailable in English, since only one *wh*-movement to Spec CP can take place in the overt syntax; if movement out of the island is to land in Spec CP before the well-formed movement does, then the well-formed movement cannot begin until after the movement out of the island has ended. Thus, the parasitic *wh*-movement strategy is unavailable in English.

To summarize, then, we have seen several phenomena which group languages of the Japanese type and languages of the Bulgarian type to the exclusion of languages like the English type. I have offered accounts of these phenomena which make crucial reference to the timing of *wh*-movement. To the extent that these accounts are convincing, they argue strongly in favor of a derivational theory of the grammar, and in particular for a theory of the derivation which permits both overt and covert movement, and in which all overt movement must precede all covert movement.

Bibliography

- Agbayani, Brian. 1997. Category raising, adjunction, and minimality. *UCI Working Papers in Linguistics* 3, 1-26.
- Alexiadou, Artemis, and Elena Anagnostopoulou. 1997. Postverbal subjects. Paper read at the 13th Comparative Germanic Syntax Workshop.
- Aoun, Joseph. 1985a. *Generalized binding: the syntax and logical form of wh-interrogatives*. Foris, Dordrecht.
- Aoun, Joseph. 1985b. *A grammar of anaphora*. MIT Press, Cambridge.
- Aoun, Joseph, and Yen-hui Audrey Li. 1993. On some differences between Chinese and Japanese wh-elements. *Linguistic Inquiry* 24:365-72.
- Baek, Yoo-Kyung Judy. In progress. Locality Condition on A-Chain and Dual Function of Finite Tense. Doctoral dissertation, MIT.
- Baker, C. Lee. 1970. Notes on the description of English questions: the role of an abstract question morpheme. *Foundations of Language* 6:197-219.
- Baker, Mark. 1988. *Incorporation: a theory of grammatical function changing*. The University of Chicago Press, Chicago.
- Baker, Mark. 1996. *The polysynthesis parameter*. Oxford University Press, Oxford.
- Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language Semantics* 4:1-56.
- Beck, Sigrid, and Shin-Sook Kim. 1996. On wh- and operator scope in Korean. *Arbeitspapiere des SFB 340: Bericht Nr. 73*.

- Berinstein, Ava. 1985. *Evidence for multiattachment in K'ekchi Mayan*. Garland Publishing, Inc., New York.
- Bittner, Maria. 1994. *Case, scope, and binding*. Kluwer, Dordrecht.
- Bobaljik, Jonathan. 1993. Ergativity and ergative unergatives. in *Papers on Case and Agreement I: MIT Working Papers in Linguistics 19*, edited by Colin Phillips and Jonathan Bobaljik. MITWPL, Cambridge.
- Bobaljik, Jonathan. 1995. Morphosyntax: the syntax of verbal inflection. Doctoral dissertation, MIT.
- Bobaljik, Jonathan, and Andrew Carnie. 1996. A Minimalist approach to some problems of Irish word order. Robert Borsley and Ian Roberts, eds., *The syntax of the Celtic languages: a comparative perspective*. 223-240. Cambridge University Press, Cambridge.
- Bobaljik, Jonathan, and Dianne Jonas. 1996. Subject positions and the role of TP. *Linguistic Inquiry* 27:195-236.
- Bolinger, Dwight. 1978. Asking more than one thing at a time. Hiz, Henry. *Questions*. 107-150. D. Reidel, Dordrecht.
- Borsley, Robert, and Janig Stephens. 1989. Agreement and the position of subjects in Breton. *Natural Language and Linguistic Theory* 7:407-428.
- Bošković, Željko. 1995a. On certain violations of the superiority condition, AgrO, and economy of derivation. ms., University of Connecticut.
- Bošković, Željko. 1995b. Principles of economy in nonfinite complementation. Doctoral dissertation, University of Connecticut.
- Bošković, Željko. 1995c. Superiority effects with multiple wh-fronting in Serbo-Croatian. Ms., University of Connecticut.
- Bošković, Željko. 1997. Sometimes in Spec CP, sometimes in-situ. ms., University of Connecticut.
- Bošković, Željko. To appear. Multiple wh-fronting and economy of derivation. To appear in *Proceedings of WCCFL 16*.
- Brandi, Luciana, and Patricia Cordin. 1989. Two Italian dialects and the null subject parameter. 111-142 in *The null subject parameter*, edited by Osvaldo Jaeggli and Ken Safir. Kluwer, Dordrecht.
- Branigan, Philip. 1992. Subjects and complementizers. Doctoral dissertation, MIT.
- Bresnan, Joan, and Sam Mchombo. 1987. Topic, pronoun, and agreement in Chichew^oa. *Language* 63:741-782.

- Brody, Michael. 1995a. Focus in Hungarian and bare checking theory. *Tübingen Working Papers in Linguistics*.
- Brody, Michael. 1995b. *Lexico-logical form: a radically minimalist theory*. MIT Press, Cambridge, Mass.
- Browning, Marguerite. 1987. Null operator constructions. Doctoral dissertation, MIT.
- Campos, Héctor. 1997. On subject extraction and the antiagreement effect in Romance. *Linguistic Inquiry* 28:92-119.
- Carstens, Vicki. 1985. Wh-movement in Yoruba. *Studies in African linguistics* 9:39-44.
- Carstens, Vicki. 1987. Extraction asymmetries in Yoruba. David Odden, ed., *Current approaches to African linguistics* 4, 61-72. Foris, Dordrecht.
- Čavar, Damir. 1996. On cliticization in Croatian: syntax or prosody? ms., Universität Potsdam.
- Cheng, Lisa Lai-Shen. 1991. On the typology of wh-questions. Doctoral dissertation, MIT.
- Cho, Eun. 1996. Neg fronting without a NegP. ms., Cornell University.
- Chomsky, Noam. 1964. *Current issues in linguistic theory*. the Hague: Mouton.
- Chomsky, Noam. 1973. Conditions on transformations. Stephen Anderson and Paul Kiparsky, eds., *A festschrift for Morris Halle*, 232-286. New York, Holt, Reinhart and Winston.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Foris, Dordrecht.
- Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. MIT Press, Cambridge, MA.
- Chomsky, Noam. 1986. *Barriers*. MIT Press, Cambridge, MA.
- Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In *Principles and parameters in comparative grammar*, edited by Robert Freidin, 417-454. MIT Press, Cambridge, MA.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, ed. Ken Hale and Samuel Jay Keyser, 1-52. MIT Press, Cambridge, MA.
- Chomsky, Noam. 1995. *The minimalist program*. MIT Press, Cambridge, MA.
- Christensen, Kirsti Koch. 1982. On multiple filler-gap constructions in Norwegian. Elisabet Engdahl and Eva Ejerhed, eds., *Readings on unbounded dependencies in Scandinavian languages*, 77-98. Almqvist and Wiksell International, Stockholm.

- Chung, Sandra. 1982. Unbounded dependencies in Chamorro grammar. *Linguistic Inquiry* 13:39-77.
- Chung, Sandra. 1983. The ECP and government in Chamorro. *Natural Language and Linguistic Theory* 1:207-244.
- Chung, Sandra. 1989. On the notion "null anaphor" in Chamorro. Osvaldo Jaeggli and Ken Safir, eds., *The Null Subject Parameter*, 143-184. Kluwer, Dordrecht.
- Chung, Sandra. 1994a. Phrase structure, agreement, and proper government. Ms., University of California at Santa Cruz.
- Chung, Sandra. 1994b. Wh-agreement and "referentiality" in Chamorro. *Linguistic Inquiry* 25:1-44.
- Cinque, Guglielmo. 1990. *Types of A-bar dependencies*. MIT Press, Cambridge, Mass.
- Clements, George. 1984. Binding domains in Kikuyu. *Studies in the Linguistic Sciences* 14:37-56.
- Cole, Peter, and Gabriella Hermon. 1994. Is there LF wh-movement? *Linguistic Inquiry* 25:239-62.
- Cole, Peter, and Gabriella Hermon. 1997. The typology of wh-movement: wh-questions in Malay. ms., University of Delaware.
- Collins, Chris. 1994. Economy of derivation and the Generalized Proper Binding Condition. *Linguistic Inquiry* 25:45-61.
- Collins, Chris, and Höskuldur Thráinsson. 1993. Object shift in double object constructions and the theory of Case. 131-174 in MITWPL 19: Papers on Case and Agreement II, edited by Colin Phillips. MITWPL, Cambridge, MA.
- Comorovski, Ileana. 1986. Multiple wh-movement in Romanian. *Linguistic Inquiry* 17:171-177.
- Comrie, Bernard. 1979. Degrees of ergativity: some Chukchee evidence. 219-240 in *Ergativity: towards a theory of grammatical relations*, edited by Frans Plank. Academic Press, London.
- Cooreman, Ann. 1987. *Transitivity and discourse continuity in Chamorro narratives*. Berlin, Mouton de Gruyter.
- Craig, Colette. 1977. *The structure of Jacalteco*. University of Texas Press, Austin.
- Craig, Colette. 1979. The antipassive and Jacalteco. *Papers in Mayan Linguistics*, edited by Laura Martin. 139-164.
- Cresti, Diana. 1995. Extraction and reconstruction. *Natural Language Semantics* 3:79-122.

- Culicover, Peter. 1993. Evidence against ECP accounts of the *that-t* effect. *Linguistic Inquiry* 24:557-561.
- Davies, William, and Luis Enrique Sam-Colop. 1990. K'iche' and the structure of antipassive. *Language* 66:522-549.
- Dekydtspotter, Laurent. 1996. Fine-grained logical form. *Proceedings of ESCOL '96*.
- Déprez, Viviane. 1988. Stylistic inversion and verb movement. *Proceedings of ESCOL '88*.
- Déprez, Viviane. 1994. Agreement, specificity and impersonal constructions. ms., Rutgers.
- Diesing, Molly. 1992. *Indefinites*. MIT Press, Cambridge.
- den Dikken, Marcel. 1992. Particles. Doctoral dissertation, Rijksuniversiteit Leiden.
- Dixon, Robert M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge University Press.
- Dobrovie-Sorin, Carmen. 1993. What does QR raise? Jan Koster and Henk van Riemsdijk, eds., *The syntax of Romanian: comparative studies in Romance*, 242-77. Mouton de Gruyter, Berlin.
- Dukes, Michael. 1993. On the status of Chamorro wh-agreement. *Proceedings of WCCFL 11*.
- Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. Elisabet Engdahl and Eva Ejerhed, eds., *Readings on unbounded dependencies in Scandinavian languages*, 151-174. Almqvist and Wiksell International, Stockholm.
- Engdahl, Elisabet. 1984. Subject gaps: an asymmetry between questions and relatives in Norwegian. *Proceedings of NELS 14*, 64-80.
- Engdahl, Elisabet. 1985. Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics* 23:3-44.
- Erteschik-Shir, Nomi. 1982. Extractability in Danish and the pragmatic principle of dominance. Elisabet Engdahl and Eva Ejerhed, eds., *Readings on unbounded dependencies in Scandinavian languages*, 175-192. Almqvist and Wiksell International, Stockholm.
- Fanselow, Gisbert. 1997. MLC and strict cyclicity. Paper presented at the Workshop on the Minimal Link Condition, Potsdam.
- Fiengo, Robert, and Heasun Kim. 1980. Binding and control in Korean: structural restrictions on anaphora in a non-configurational language. *Journal of Linguistic Research* 1:59-73.

- Fox, Danny. 1997. Reconstruction, binding theory and the interpretation of chains. ms., MIT.
- Frampton, John. 1990. The fine structure of wh-movement and the proper formulation of the ECP. ms., Northeastern University.
- Franks, Steven. 1997. South Slavic clitic placement is *still* syntactic. Paper presented at Penn Linguistics Colloquium.
- Georgopoulos, Carol. 1985. Variables in Palauan syntax. *Natural Language and Linguistic Theory* 3:59-94.
- Georgopoulos, Carol. 1991. *Syntactic variables: resumptive pronouns and A-bar binding in Palauan*. Kluwer, Dordrecht.
- Gerdts, Donna. 1980. Antipassives and causatives in Halkomelem. *Proceedings of the Berkeley Linguistics Society* 6.
- Gerdts, Donna. 1988. A nominal hierarchy in Halkomelem clausal organization. *Anthropological Linguistics* 30:20-36.
- Grewendorf, Günther, and Joachim Sabel. 1996. Multiple specifiers and the theory of adjunction: on scrambling in German and Japanese. *Sprachwissenschaft in Frankfurt Arbeitspapier* 16.
- Groat, Erich, and John O'Neil. 1996. Spell-out at the LF interface. *Minimal ideas*, edited by Werner Abraham, Samuel Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart, 113-139. John Benjamins, Amsterdam.
- Grohmann, Kleanthes. 1997. German superiority. ms., University of Maryland.
- Guilfoyle, Eithne, Henrietta Hung, and Lisa Travis. 1992. Spec of IP and Spec of VP: two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375-414.
- Haegeman, Liliane. 1995. *The syntax of negation*. Cambridge University Press, Cambridge.
- Haegeman, Liliane. 1996. The typology of syntactic positions: L-relatedness and the A/A'-position. 141-165 in *Minimal ideas*, edited by Werner Abraham, Samuel Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart. John Benjamins, Amsterdam.
- Hagstrom, Paul. 1995. Negation, focus, and *do*-support in Korean. ms., MIT.
- Hale, Ken, and Luciana Storto. 1996. Agreement and spurious antipassives. Ms., MIT.
- Harley, Heidi. 1995. Subjects, events, and licensing. Doctoral dissertation, MIT.
- Hasegawa, Nobuko. 1984. On the so-called "zero pronouns" in Japanese. *The Linguistic Review* 4:289-341.

- Hendrick, Randall. 1988. *Anaphora in Celtic and universal grammar*. Kluwer, Dordrecht.
- Holmberg, Anders. 1986. Word order and syntactic features in the Scandinavian languages and English. Doctoral dissertation, University of Stockholm.
- Holmberg, Anders. 1996. The true nature of Holmberg's generalization. Handout from talk given at NELS 27.
- Holmberg, Anders, and Christer Platzack. 1995. *The role of inflection in Scandinavian syntax*. Oxford University Press, Oxford.
- Hornstein, Norbert. 1995. *Logical form: from GB to Minimalism*. Blackwell, Cambridge, Mass.
- Horvath, Julia. 1986. *FOCUS in the theory of grammar and the syntax of Hungarian*. Foris, Dordrecht.
- Howard, I., and A. Niekawa Howard. 1976. Semantics of the Japanese passive. Masatoshi Shibatani, ed., *Syntax and semantics 5: Japanese generative grammar*. Academic Press, New York.
- Huang, James. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT.
- Izvorski, Roumyana. 1995. WH-movement and focus-movement in Bulgarian. *Proceedings of CONSOLE 2*.
- Jaeggli, Osvaldo. 1984. Subject extraction and the null subject parameter. *Proceedings of NELS 14*, 132-153.
- Jayaseelan, K. A. 1990. Incomplete VP deletion and gapping. *Linguistic Analysis* 20:64-81.
- Johnson, Kyle. 1991. Object positions. *Natural Language and Linguistic Theory* 9:577-636.
- Johnson, Kyle. 1994. Bridging the gap. ms., University of Massachusetts at Amherst.
- Johnson, Kyle. 1997. In search of the English middle field. ms., University of Massachusetts at Amherst.
- Jonas, Dianne, and Jonathan Bobaljik. 1993. Specs for subjects: the role of TP in Icelandic. In *Papers on Case and Agreement I: MIT Working Papers in Linguistics 18*, ed. Jonathan Bobaljik and Colin Phillips, 59-98.
- Jónsson, Jóhannes. 1991. Stylistic fronting in Icelandic. *Working Papers in Scandinavian Syntax* 48:1-44.
- Kayne, Richard. 1972. Subject inversion in French interrogatives. Jean Casagrande and Bohdan Saciuk, eds., *Generative studies in Romance languages*, 70-126. Newbury House, Rowley, Mass.

- Kayne, Richard. 1983a. Connectedness. *Linguistic Inquiry* 14:223-250.
- Kayne, Richard. 1983b. *Connectedness and binary branching*. Foris, Dordrecht.
- Kayne, Richard. 1985. Principles of particle constructions. *Grammatical representation*, edited by Jacqueline Guéron, Hans-Georg Obenauer, and Jean-Yves Pollock, 101-140. Foris: Dordrecht.
- Kayne, Richard. 1989. Facets of past participle agreement in Romance. In *Dialect variation in the theory of grammar*, ed. Paola Beninca, 85-103. Foris, Dordrecht.
- Kayne, Richard, and Jean-Yves Pollock. 1979. Stylistic inversion, successive cyclicity, and move NP in French. *Linguistic Inquiry* 9:595-621.
- Kim, Shin-Sook, and Wolfgang Sternefeld. 1997. Superiority vs. cross over. Paper presented at the Workshop on the Minimal Link Condition, Potsdam.
- Kiss, Katalin. 1987. *Configurationality in Hungarian*. Reidel, Dordrecht.
- Kiss, Katalin. 1994. Sentence structure and word order. *Syntax and semantics 27: The syntactic structure of Hungarian*, ed. by Ferenc Kiefer & Katalin Kiss, 1-90. Academic Press, New York.
- Kitahara, Hisatsugu. 1994. Target- α : a unified theory of movement and structure-building. Doctoral dissertation, Harvard University.
- Kitahara, Hisatsugu. 1997. *Elementary operations and optimal derivations*. MIT Press, Cambridge.
- Koizumi, Masatoshi. 1993. Object agreement and the split VP hypothesis. *MITWPL 18: Papers on Case and Agreement I*, edited by Colin Phillips and Jonathan Bobaljik, 99-148.
- Koizumi, Masatoshi. 1995. Phrase structure in Minimalist syntax. Doctoral dissertation, MIT.
- Koopman, Hilda. 1983. Control from COMP and comparative syntax. *The Linguistic Review* 2:365-391.
- Koopman, Hilda. 1984. *The syntax of verbs: from verb movement rules in the Kru languages to universal grammar*. Foris, Dordrecht.
- Koopman, Hilda, and Dominique Sportiche. 1982. Variables and the bijection principle. *The Linguistic Review* 2:139-160.
- Kroch, Anthony. 1989. Amount quantification, referentiality, and long wh-movement. Ms., University of Pennsylvania.
- Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: Center for the Study of Language and Information.

- Kuno, Susumu. 1973. *The structure of the Japanese language*. MIT Press, Cambridge, MA.
- Kuno, Susumu. 1980. A (further) note on Tonoike's intra-subjectivization hypothesis. Yukio Otsu and Ann Farmer, eds., *MITWPL 2: Theoretical issues in Japanese linguistics*, 171-184.
- Kuroda, Sige-Yuki. 1971. Remarks on the notion of subject with reference to words like also, even, or only, Part II, In *Annual Bulletin* 4:127-152. Logopedics and Phoniatrics Research Institute, University of Tokyo. (Reprinted in *Papers in Japanese Linguistics* 11, 157-202.)
- Laka, Itziar. 1992. Negation in syntax: on the nature of functional categories and projections. Doctoral dissertation, MIT.
- Larsen, Thomas. 1987. The syntactic status of ergativity in Quiché. *Lingua* 71:33-59.
- Lasnik, Howard. 1995. A note on pseudogapping. Rob Pensalfini and Hiroyuki Ura, eds., *MITWPL 27: Papers on minimalist syntax*. MITWPL, Cambridge, MA.
- Lasnik, Howard, and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15:235-289.
- Lasnik, Howard, and Mamoru Saito. 1991. On the subject of infinitives. *Chicago Linguistics Society* 27.
- Lawal, Nike. 1987. Yoruba relativisation and the continuous segment principle. *Studies in African Linguistics* 18:67-80.
- Legendre, Géraldine, Colin Wilson, Paul Smolensky, Kristin Homer, and William Raymond. 1994. Optimality and wh-extraction. ms., Johns Hopkins University and University of Colorado.
- Lengyel, Thomas. 1978. Ergativity, aspect and related perplexities of Ixil-Maya. *Papers in Mayan linguistics* 2, edited by Nora England, 78-91.
- Longobardi, Giuseppe. 1987. Extraction from NP and the proper notion of head government. Alessandra Giorgi and Giuseppe Longobardi, eds., *The syntax of noun phrases*, 57-112, Cambridge University Press, Cambridge.
- MacLachlan, Anna. 1995. Aspects of ergativity in Tagalog. Doctoral dissertation, McGill University.
- MacLachlan, Anna, and Masanori Nakamura. 1994. Case-checking and specificity in Tagalog. ms., McGill University.
- Mahajan, Anoop. 1992. The A/A-bar distinction and movement theory. Doctoral dissertation, MIT.
- Maki, Hideki. 1994. Anti-anti-superiority. *Proceedings of ESCOL '94*.

- Marácz, Laszlo. 1989. Asymmetries in Hungarian. Doctoral dissertation, Rijksuniversiteit Groningen.
- Maling, Joan. 1990. Inversion in embedded clauses in Modern Icelandic. In *Syntax and semantics 24: Modern Icelandic syntax*, ed. Joan Maling and Annie Zaenen, 71-91. Academic Press, San Diego.
- Manzini, Maria Rita. 1991. *Locality*. MIT Press, Cambridge, Mass.
- Marantz, Alec. 1991. Case and licensing. *Proceedings of ESCOL '91*.
- Massam, Diane. 1985. Case theory and the projection principle. Doctoral dissertation, MIT.
- Merchant, Jason. 1996. Deviance from identity under ellipsis: the case of (multiple) sluicing. Handout from University of Groningen colloquium.
- McCloskey, James. 1990. Resumptive pronouns, A-bar binding and levels of representation in Irish. Randall Hendrick, ed., *Syntax and semantics 23: the syntax and semantics of modern Celtic languages*, 199-248. Academic Press, San Diego.
- McCloskey, James, and Ken Hale. 1984. On the syntax of person-number inflection in Modern Irish. *Natural Language and Linguistic Theory* 1:487-533.
- McGinnis, Martha. To appear. Reflexive external arguments and lethal ambiguity. To appear in *Proceedings of WCCFL 16*.
- Miyagawa, Shigeru. 1989. *Syntax and semantics 22: structure and case marking in Japanese*. Academic Press, San Diego.
- Miyagawa, Shigeru. 1994. Nonconfigurationality within a configurational structure. ms., MIT.
- Miyagawa, Shigeru. 1997a. Against optional scrambling. *Linguistic Inquiry* 28:1-26.
- Miyagawa, Shigeru. 1997b. On the nature of wh-scope. Ms., MIT.
- Moltmann, Friederike. 1995. Exception sentences and polyadic quantification. *Linguistics and Philosophy* 18:223-280.
- Mondloch, James. 1981. Voice in Quiché-Maya. Doctoral dissertation, SUNY-Albany.
- Mueller, Gereon. 1996. Optimality Theory in syntax II: parallel movement. ms., University of Tübingen.
- Mulders, Iris. 1996. Multiple checking and mirrored specifiers: the structure of CP. ms., Utrecht University.
- Mulders, Iris. To appear. Mirrored specifiers. To appear in *Linguistics in the Netherlands* 1997.

- Murasugi, Kumiko. 1992. Crossing and nested paths: NP movement in accusative and ergative languages. Doctoral dissertation, MIT.
- Nakamura, Masanori. 1994. An economy account of wh-extraction in Tagalog. *Proceedings of WCCFL 12*.
- Nakamura, Masanori. 1995. Reference set, Minimal Link Condition, and parameterization. Paper presented at the Workshop on Optimality in Syntax, MIT, Cambridge, MA.
- Nakamura, Masanori. 1996. Economy of chain formation. Doctoral dissertation, McGill University.
- Nakamura, Masanori. To appear. Wh-disagreement. To appear in *Proceedings of SCIL 7*.
- Neeleman, Ad. 1994. Scrambling as a D-structure phenomenon. 387-429 in *Studies on scrambling*, edited by Norbert Corver and Henk van Riemsdijk. Mouton de Gruyter, Berlin.
- Neijt, Anneke. 1979. *Gapping: a contribution to sentence grammar*. Foris, Dordrecht.
- Nishigauchi, Taisuke. 1990. *Quantification in the theory of grammar*. Kluwer, Dordrecht.
- Nishigauchi, Taisuke. To appear. 'Multiple sluicing' in Japanese and the functional nature of wh-phrases. To appear in *Journal of East Asian Linguistics*.
- Noonan, Máire. To appear. Functional architecture and wh-movement: Irish as a case in point. To appear in Eithne Guilfoyle, ed., *Celtic syntax* (special issue of *Canadian Journal of Linguistics*).
- Obenauer, Hans-Georg. 1992. L'interprétation des structures *wh* et l'accord du participe passé, to appear in Hans-Georg Obenauer and Anne Zribi-Hertz, eds., *Structures de la phrase et théorie du liage*, Presses Universitaires de Vincennes.
- Ochi, Masao. 1996. On the nature of relativization in Japanese. Paper presented at CONSOLE 5.
- Ochi, Masao. 1997. Move or attract? Paper presented at WCCFL 16.
- Oka, Toshifusa. 1995. Fewest steps and island sensitivity. *MITWPL 27: Papers on Minimalist syntax*, edited by Rob Pensalfini and Hiroyuki Ura. MITWPL, Cambridge, MA.
- Ouhalla, Jamal. 1993. Subject-extraction, negation and the anti-agreement effect. *Natural Language and Linguistic Theory* 11:477-518.
- Perlmutter, David. 1971. *Deep and surface structure constraints in syntax*. Holt, Rinehart, and Winston, New York.
- Pesetsky, David. 1982. Paths and categories. Doctoral dissertation, MIT.

- Pesetsky, David. 1984. Extraction domains and a surprising subject/object asymmetry. *GLOW newsletter* 12:58-60.
- Pesetsky, David. 1987. Wh-in-situ: movement and unselective binding. In Eric Reuland and Alice ter Meulen, eds., *The representation of (in)definiteness*, 98-129. MIT Press, Cambridge.
- Pesetsky, David. 1989. Language-particular processes and the earliness principle. ms., MIT.
- Pesetsky, David. 1991. Zero syntax II. Ms., MIT.
- Pesetsky, David. To appear. Some optimality principles of sentence pronunciation. To appear in Pilar Barbosa et al, eds., *Is the best good enough?*, MIT Working Papers in Linguistics and MIT Press, Cambridge, MA.
- Phillips, Colin. 1993. Conditions on agreement in Yimas. In Jonathan Bobaljik and Colin Phillips, eds., *MITWPL 18: Papers in case and agreement I*, 173-212.
- Phillips, Colin. 1996. Ergative subjects. *Theoretical approaches to empirical questions*, edited by Clifford Burgess, Katarzyna Dziwerek, and Donna Gerds, 341-357. CSLI, Stanford.
- Phillips, Colin. To appear. Disagreement between adults and children. To appear in Amaya Mendikoetxea and Miriam Uribe-Etxebarria, eds., *Theoretical issues in the morphology-syntax interface*. San Sebastian: ASJU.
- Platzack, Christer. 1987. The Scandinavian languages and the null subject parameter. *Natural Language and Linguistic Theory* 5:377-401.
- Pollock, Jean-Yves. 1981. On case and impersonal constructions. Robert May and Jan Koster, eds., *Levels of syntactic representation*, 219-252. Foris, Dordrecht.
- Poole, Geoffrey. 1996. Optional movement in the Minimalist Program. In *Minimal ideas*, ed. Werner Abraham, Samuel Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart, 199-216. John Benjamins, Philadelphia.
- Postal, Paul. 1974. *On raising: one rule of English grammar and its theoretical implications*. MIT Press, Cambridge, Mass.
- Raghibdoust, Shahla. 1994. Multiple wh-fronting in Persian. *Cahiers Linguistiques d'Ottawa* 21:27-58.
- Reinhart, Tanya. 1979. A second COMP position. *Theory of markedness in generative grammar*, ed. by Adriana Belletti et al, 517-557.
- Reinhart, Tanya. 1995. Interface strategies. OTS Working Papers, Utrecht.
- Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657-720.
- Richards, Norvin. 1990. Tagalog subjects. ms., Cornell University.

- Richards, Norvin. 1993. Tagalog and the typology of scrambling. Honors thesis, Cornell University.
- Richards, Norvin. 1995a. Towards a feature-movement theory of long-distance anaphora. Paper presented at the Workshop on Optionality, OTS/Utrecht University.
- Richards, Norvin. 1995b. Towards a theory of head-binding. Ms., MIT.
- Richards, Norvin. 1996. Subjacency forever. To appear in *Proceedings of CONSOLE 5*.
- Richards, Norvin. To appear a. Another look at Tagalog subjects. To appear in *Formal Issues in Austronesian Linguistics*, edited by Ileana Paul, Vivianne Phillips, and Lisa Travis. Kluwer, Dordrecht.
- Richards, Norvin. To appear b. Shortest moves to (anti-)superiority. To appear in *Proceedings of WCCFL 16*.
- Richards, Norvin. In press. The principle of minimal compliance. To appear in *Linguistic Inquiry*.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Foris, Dordrecht.
- Rizzi, Luigi. 1986. On chain formation. Hagit Borer, ed., *Syntax and semantics 19: the syntax of pronominal clitics*. New York, Academic Press.
- Rizzi, Luigi. 1990. *Relativized minimality*. MIT Press, Cambridge, Mass.
- Rögnvaldsson, Eiríkur, and Höskuldur Thráinsson. 1990. On Icelandic word order once more. In *Syntax and semantics 24: Modern Icelandic syntax*, ed. Joan Maling and Annie Zaenen, 3-40. San Diego: Academic Press, Inc.
- Rudin, Catherine. 1985. *Aspects of Bulgarian syntax: complementizers and wh constructions*. Slavica Publishers, Inc., Columbus, Ohio.
- Rudin, Catherine. 1988. On multiple questions and multiple wh-fronting. *Natural Language and Linguistic Theory* 6:445-501.
- Saddy, Douglas. 1991. Wh-scope mechanisms in Bahasa Indonesia. *MITWPL 15: More papers on wh-movement*, edited by Lisa Cheng and Hamida Demirdash.
- Sag, Ivan. 1976. Deletion and logical form. Doctoral dissertation, MIT.
- Saito, Mamoru. 1982. Scrambling, topicalization, and strong crossover. ms., MIT.
- Saito, Mamoru. 1985. Some asymmetries in Japanese and their theoretical implications. Doctoral dissertation, MIT.
- Saito, Mamoru. 1989. Scrambling as semantically vacuous A'-movement. 182-200 in Mark Baltin and Anthony Kroch, eds., *Alternative conceptions of phrase structure*, University of Chicago Press, Chicago.

- Saito, Mamoru. 1992. Long-distance scrambling in Japanese. *Journal of East Asian Linguistics* 1:69-118.
- Saito, Mamoru. 1994a. Additional-wh effects and the adjunction site theory. *Journal of East Asian Linguistics* 3:195-240.
- Saito, Mamoru. 1994b. Improper adjunction. *MITWPL 24: Formal Approaches to Japanese Linguistics 1*, edited by Masatoshi Koizumi and Hiroyuki Ura.
- Saito, Mamoru. 1996. Barriers for A-adjunction. *Japanese/Korean Linguistics* 5.
- Saito, Mamoru, and Naoki Fukui. 1996. Order in the theory of phrase structure and movement. ms., Nanzan University and University of California, Irvine.
- Sauerland, Uli. 1995. Sluicing and islands. ms., MIT.
- Sauerland, Uli. 1996a. The interpretability of scrambling. ms., MIT.
- Sauerland, Uli. 1996b. Guess how. *Proceedings of CONSOLE 4*.
- Schachter, Paul. 1976. The subject in Philippine languages: actor, topic, actor-topic, or none of the above. In Charles Li, ed., *Subject and Topic*, 491-518. New York: Academic Press.
- Schachter, Paul. 1996. The subject in Tagalog: still none of the above. *UCLA Occasional Papers in Linguistics* 15.
- Schafer, Robin. 1995. Negation and verb second in Breton. *Natural Language and Linguistic Theory* 13:135-172.
- Schneider-Zioga, Patricia. 1995. Specifier/head agreement in Kinande. *Cahiers Linguistiques d'Ottawa* 23:67-93.
- Schneider-Zioga, Patricia. 1996. An argument in favor of Agreement Phrase. *Proceedings of ESCOL '96*.
- Smith-Stark, Thom. 1978. The Mayan antipassive: some facts and fictions. *Papers in Mayan linguistics*, edited by Nora England, 169-181.
- Snyder, William. 1992. Chain-formation and crossover. ms., MIT.
- Sohn, Keun-Won. 1994. Adjunction to argument, free ride, and a minimalist program. *MITWPL 24: formal approaches to Japanese linguistics 1*, edited by Masatoshi Koizumi and Hiroyuki Ura. Cambridge, MA.
- Sonaiya, Remi. 1989. Wh-movement and proper government in Yoruba. Paul Newman and Robert Botne, eds., *Current approaches to African linguistics* 5, 109-126. Foris, Dordrecht.
- Spencer, Andrew. 1991. *Morphological theory: an introduction to word structure in generative grammar*. Oxford: Basil Blackwell.

- Stjepanović, Sandra. 1997. Clitics and VP ellipsis in Serbo-Croatian. Handout of talk given at Formal Approaches to Slavic Linguistics 6.
- Stowell, Tim. 1986. Null antecedents and proper government. *Proceedings of NELS 16*.
- Stump, Gregory. 1984. Agreement vs. incorporation in Breton. *Natural Language and Linguistic Theory* 2:289-348.
- Stump, Gregory. 1989. Further remarks on Breton agreement. *Natural Language and Linguistic Theory* 7:429-472.
- Szabolcsi, Anna, and Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1:235-284.
- Tada, Hiroaki. 1993. A/A-bar partition in derivation. Doctoral dissertation, MIT.
- Takahashi, Daiko. 1993. Movement of wh-phrases in Japanese. *Natural Language and Linguistic Theory* 11:655-678.
- Takahashi, Daiko. 1994a. Minimality of movement. Doctoral dissertation, University of Connecticut.
- Takahashi, Daiko. 1994b. Sluicing in Japanese. *Journal of East Asian Linguistics* 3:265-300.
- Takahashi, Daiko. 1997. Move-F and null operator movement. *The Linguistic Review* 14:181-196.
- Tanaka, Hidekazu. In progress. Conditions on logical form derivations and representations. Doctoral dissertation, McGill University.
- Taraldsen, Knut Tarald. 1986. *Som* and the binding theory. 149-184 in *Topics in Scandinavian syntax*, edited by Lars Hellan and Kirsti Koch Christensen. Dordrecht, D. Reidel.
- Thráinsson, Höskuldur. 1979. *On complementation in Icelandic*. New York: Garland.
- Topping, Donald. 1973. *Chamorro reference grammar*. Honolulu, University Press of Hawaii.
- Trechsel, Frank. 1993. Quiché focus constructions. *Lingua* 91:33-78.
- Tsai, Wei-Tien Dylan. 1994. On economizing the theory of A-bar dependencies. Doctoral dissertation, MIT.
- Ura, Hiroyuki. 1993. On feature-checking for wh-traces. *MITWPL 18: Papers on case and agreement I*, edited by Jonathan Bobaljik and Colin Phillips, 243-280.
- Ura, Hiroyuki. 1996. Multiple feature-checking: a theory of grammatical function splitting. Doctoral dissertation, MIT.

- Valois, David, and Fernande Dupuis. 1992. On the status of (verbal) traces in French: the case of stylistic inversion. Paul Hirschbühler and Konrad Koerner, eds, *Romance languages and modern linguistic theory*, 325-338. John Benjamins, Amsterdam.
- Veraart, Fleur. 1996. On the distribution of Dutch reflexives. Masters thesis, MIT.
- Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. Oxford University Press, Oxford.
- Voskuil, Jan. 1996. *Comparative morphology: verb taxonomy in Indonesian, Tagalog, and Dutch*. Doctoral dissertation, Holland Institute of Generative Linguistics.
- Watanabe, Akira. 1992. Subjacency and S-structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1:255-291.
- Watanabe, Akira. 1996. *Case absorption and wh-agreement*. Kluwer, Dordrecht.
- Watanabe, Akira. 1997. Absorption as feature checking. ms., Kanda University of International Studies.
- Wiltschko, Martina. 1997. Superiority in German. Paper presented at WCCFL 16.
- de Wind, Maarten. 1997. Attract F and accusative case-checking in French Stylistic Inversion. Paper presented at WCCFL 16.
- Yatsushiro, Kazuko. 1996. On the unaccusative construction and nominative case licensing. ms., University of Connecticut.
- Yip, Moira, Joan Maling, and Ray Jackendoff. 1987. Case in tiers. *Language* 63:217-250.
- Zaenen, Annie, Joan Maling, and Höskuldur Thráinsson. 1985. Case and grammatical functions: the Icelandic passive. *Natural Language and Linguistic Theory* 3:441-483.