MINIMALISM IN SYNTACTIC DERIVATION

by

Toshifusa Oka

B.A., Meiji Gakuin University
(1984)

M.A., University of Tsukuba
(1987)

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

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September, 1993

© Toshifusa Oka, 1993
All rights reserved

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

Signature of Author

Department of Linguistics and Philosophy
July 22, 1993

Certified by

Noam Chomsky
Thesis Supervisor

Accepted by

Wayne O'Neil
Chair, Departmental Committee
MINIMALISM IN SYNTACTIC DERIVATION

by

Toshifusa Oka

Abstract

This thesis proposes a theory of how syntactic derivations are constrained. Within the Minimalist Program of Chomsky 1992, linguistic expressions are optimal realizations of interface conditions. Optimal derivations have minimal cost. The cost of a derivation is determined by principles of Economy.

A novel formulation of the system of Economy principles is proposed. For example, the Shortest Steps principle of Chomsky (1992) is replaced by the Shallowness principle, which requires that operations must be the shallowest in the phrase markers, and other principles such as the Procrastinate principle are redefined. It is also proposed that Economy principles are ordered in a certain way. Furthermore, a system of feature-checking is proposed, generalizing abstract Case to PP complements and adjuncts.

As a consequence, wh-movement phenomena that have been treated as independent phenomena, including the superiority, nesting, and ECP effects, are explained within a system based purely on Economy basis, crosslinguistic and intralinguistic variations being derived from parameterized morphological properties.

Thesis Supervisor: Noam Chomsky
Title: Institute Professor
Acknowledgement

My dissertation year has been enjoyable thanks to people I love and respect.

For their help and support I would like to thank the members of my dissertation committee first: Noam Chomsky, Ken Hale, and Howard Lasnik.

I greatly appreciate the manner Noam acted as my supervisor. He treated me as a linguist and kept giving me sincere and serious criticisms, which stimulated me to think deeper and aim higher. And when I was really in need, he didn’t fail to suggest a way out. Through conversations with him I have learned a lot about how we should do linguistics, which I believe will remain to be the most important lessons to me as long as I try to be a linguist.

I enjoyed every single conversation with Ken. I admire his insight in language and linguistic theory. And with his warm heart he always makes me feel happy. Without his gentle smile, Building 20 would be a totally different place.

I was very lucky to be able to work with Howard. He is so generous to share his precious time with me that I could bring him almost every ideas that just occured to me. Without his suggestions and encouragement, I would never have reached where I am.

Secondly, I would like to thank Doug Jones, Shigeru Miyagawa, Mamoru Saito, and Yuji Takano for repeatedly having discussions with me, which helped develop ideas.

I am also thankful for their help in various manners to Jun Abe, Jonathan Bobaljik, Chris Collins, Naoki Fukui, Masatoshi Koizumi, Colin Philips, Hiroaki Tada, Hubert Truckenbrodt, Wei-tien Tsai, Hiroyuki Ura, and Akira Watanabe.

I would like to take this opportunity to thank people who made my life at/around MIT possible and pleasant for four years.

I am much obliged to the faculty members besides the ones I mentioned above: François Dell, Morris Halle, Jim Harris, Irene Heim, Jim Higginbotham, Michael Kenstowicz, Jay Keyser, Utpal Lahiri, Alec Maranz, Wayne O’Neil, David Pesetsky, Ken Wexler. Special thanks go to Jay for his warm encouragements.

I am also thankful to Marilyn Goodrich, Nancy Peters, Rachel Pearl, Marilyn Silva, Monique Tasker, Wendy Weber, and Jamie Young for keeping my life in the department more comfortable.

I was lucky to be among my classmates: Tom Green, Doug Jones, Sigal Uziel, and Akira Watanabe. The first year would have been much more difficult for me, if they were not with me. I will never forget the birthday
party for me which they suddenly started in the middle of Wayne’s survey class in my first year. Thank you! Special thanks go to Doug for his constant support. His friendship is one of the most important things I have got here.

I would like to thank other fellow students too: Pilar Barbosa, Jonathan Bobljik, Eularia Bonet, Phil Branigan, Tony Bures, Andrew Carnie, Lisa Cheng, Chris Collins, Diana Cresti, Hamida Demirdache, Jill Gaulding, Heidi Harley, Michael Hegarty, Mika Hoffman, Sabine Iatridou, Bill Ilsdari, Peter Ihionu, Shawn Kern, Utpal Lahiri, Itzier Laka, Harry Leder, Yafei Li, Masatoshi Koizumi, Miori Kubo, Paul Law, Anoop Mahajan, Seth Minkoff, Friederike Moltmann, Kumiko Murasugi, Renate Musan, Rolf Noyer, Orin Percus, Colin Philips, Michelle Sigler, Hiroaki Tada, Chris Tancredi, Carol Tenny, Hubert Tuckenbrodt, Wei-tien Tsai, Hiroyuki Ura, and Bao Zhiming. Special thanks go to Hamida for being always kind to me.

I would also like to thank other people at/around MIT, especially Piroska Csuri, Mattew Holman, Márti and Laci Maracz, Javier Ormazabal, and Myriam Uribe-Etxeharria.

I wouldn’t have survived here without Japanese around me: Jun Abe, Noriko and Masao Fukuhara, Naoki Fukui, Katsunori Fukuyasu, Ken Hinomizu, Hiroto Hoshi, Masayuki Ikeuchi, Toshimi and Toyoshi Kaito, Haruko Kato, Norimi Kimura, Hisatsugu Kitahara, Masatoshi Koizumi, Miori Kubo, Shigeru Miyagawa, Yoji Miyamoto, Masayuki Ohkado, Masayuki Oishi, Kohei, Ikuko and Koji Ono, Mamoru Saito, Shizue and Minoru Shimbo, Hiroaki Tada, Michie and Yuji Takano, Akemi and Shoichi Tanaka, Kazuko and Shigeo Tonoike, Masaaki Tsutsui, Asako Uchibori, Taira Uchida, Hiroyuki Ura, Saeko Urushibara, Akira Watanabe, and Noriaki Yusa.

I am much obliged to Sakio Saito and Shigeo Tonoike at Meiji Gakuin University, where I met linguistics. And I am also thankful to Minoru Nakau and other people who form the Tsukuba linguistics community. The liberal and lively atmosphere at Tsukuba helped me a lot to decide to become a linguist.

Finally, I can never thank my mother Fusako and my wife Madoka enough for their love and support. I dedicate this thesis to them.
# Table of Contents

1. Introduction. .......................................................... 9
2. Shallowness. ......................................................... 11
3. Superiority and Nesting. ........................................... 16
4. An Economy system. ................................................ 29
   4.1. A reformulation of Economy principles. ............... 29
   4.2. Comparison domains. ........................................ 39
5. Impossible extractions. ............................................ 49
   5.1. Adjunct extraction. .......................................... 49
   5.2. Other impossible extractions. ............................. 61
6. Strict Cycle. .......................................................... 88
7. A-over-A. ............................................................. 102
0. Introduction.

In this dissertation we will present an articulated theory of the Economy of Derivation basically within the minimalist framework of Chomsky (1992). Given an extended system of feature-checking, an organized set of Economy principles will provide a unified explanation for various wh-movement phenomena that have been subject to unrelated considerations.

In section 1 we will propose an Economy principle, which we call Shallowness. In section 2 we will see how Shallowness works, considering the superiority and nesting effects exemplified below:

(1) a. who did you persuade t [ to buy what ]
   b. ??what did you persuade who [ to buy t ]
(2) a. ??what did you decide [ who [ to persuade t [ to buy t ]]]
   b. *who did you decide [ what [ to persuade t [ to buy t ]]]

These two phenomena will be given parallel accounts. In section 3 we will reformulate other Economy principles. And we will consider how

---

1. The "pure superiority" violation, which is found in (1b) is not as bad as the "standard superiority" violation, which is found in the following:

(i) *what did who buy t

It is likely that additional factors are involved here, including a well-known degrading effect which is generally observed when a nominative wh-phrase is in situ.

Some speakers seem to accept both of (1b) and (i). I suspect that such speakers can analyze single-word wh-phrases in situ as if they were which-phrases, which Pesetsky (1987) calls “D-linked.”

(ii) a. which student did you decide [ which book [ to persuade t [ to buy t ]]]
   b. which book did which student buy t
We assume that Hornestein and Välimäki’s (1990) analysis of complex \textit{wh}-phrases is basically correct. We will discuss this matter later.

Example (2a) is not perfect. There has not been an agreement on the acceptability of examples with a \textit{wh}-island violation. For some speakers the \textit{wh}-island constraint is so severe that they even judge (2a) as worse than (1b), which involves a superiority violation. Generally, the effect of a \textit{wh}-island violation depends on two factors. One is the sort of the extracted element: \textit{which}-phrases and relative pronouns/operators are easier to extract than single-word \textit{wh}-phrases. Thus, even speakers to whom (2a) is unacceptable find a much sharper contrast in the following:

(iii) a. which book did you decide [ which boy to persuade t_j to buy t_i ]
     b. * which boy did you decide [ which book to persuade t_i to buy t_j ]

(iiiia) is fairly acceptable while (iiib) is still totally unacceptable. Another factor is the tense of the \textit{wh}-clause: infinitival \textit{wh}-clauses more readily allow extraction than tensed \textit{wh}-phrases, with variations among languages. What should be noted is that whenever there is a contrast the crossing is worse than the nesting. This contrast is one of the phenomena that I would like to mainly consider in this dissertation. The subjacency effect, which I use as a descriptive term to refer to phenomena such as the degrading of acceptability found in examples like (2a), is not intended to be explained in this dissertation.

One might also think that the total unacceptability of examples like (2b), which involves a linear crossing, is a garden path effect rather than involving a syntactic factor. However, this conjecture is simply wrong. Consider the following pair of examples:

(i) a. ? what do you wonder [ who he persuaded friends of t_i [ to do t_j at the party ]]
     b. ? who do you wonder [ what he persuaded friends of t_i [ to do t_j at the party ]]

There is no clear contrast found here: (ib) is as acceptable as (1va), though the former, like (2b), involves a linear crossing. The acceptability difference between (ib) and (2b) should be attributed to their structural difference. It is easy to define a hierarchical relation in terms of structural notions in such a way that in (2b) the two \textit{wh}-phrases in their original positions are in that relation but not in (ib).
competing derivations are compared with each other. In section 4 we will propose a mechanism of nonstructural Case-checking so as to account for some ECP phenomena on the same basis as the superiority and nesting effects. A typical case is the impossibility of adjunct extraction from a wh-island, which is exemplified below:\(^3\)

(3) a. ?what do you wonder [ whether [ John fixed t yesterday ]]
   b. *when do you wonder [ whether [ John fixed it t ]]

We will further discuss the unextractability of nonreferential phrases from a wh-island and the complementizer-trace phenomena. In section 5 we will argue that Strict Cycle and successive cyclicity can be deduced from Economy considerations. In section 6 we will further discuss some related phenomena.

---

Similarly, the superiority effect exemplified in (1), should be also described in structural terms (Cf. Fiengo (1980), Fiengo, Huang, Lasnik, and Reinhart (1988), Hornstein and Weinberg (1990)):

(ii) a. ? whom did you persuade friends of t [ to buy what ]
   b. ? what did you persuade friends of whom [ to buy t ]

Note also that although Pesetsky (1982) and Lasnik and Saito (1992) give a structural explanation for the nesting and superiority effects and the superiority effect, respectively, their formulations, unlike the one I will present in this dissertation, fail to capture the above facts.

\(^3\) As has been pointed out in the literature that the argument-adjunct asymmetry with regard to the extractability from a wh-island becomes clearer with why and how instead of when and where. See Huang (1982) and Murasugi and Saito (1992), among others. Leaving open where this difference comes from, our discussions concerning adjuncts in this work are not affected by the choice of the adjunct we use for exposition.
1. Shallowness.

We propose the following general principle, which plays a significant role to determine derivational cost:

(4) Shallowness
   An operation must be the shallowest.

This principle covers Move-\(\alpha\) operations and feature-checking operations. Let us restrict ourselves to Move-\(\alpha\) operations until we discuss the Strict Cycle.

With regard to Shallowness no distinction is made between singular and binary operations or between substitution and adjunction. The domain of alternatives from which the shallowest operation is chosen is restricted in a particular manner, which we will return to below.

The depth of a Move-\(\alpha\) operation is defined as follows:

(5) Depth
   The depth of a Move-\(\alpha\) operation affecting \(\alpha\) is the union of the depth of \(\alpha\) in the input of the operation and the depth of \(\alpha\) in the output, where the depth of \(\alpha\) is the set of the maximal projections which dominate \(\alpha\).

Consider the following Move-\(\alpha\) operation, for example, represented in a schematic tree diagram:

\footnote{We follow Chomsky's (1992) definitions of domination and containment throughout the present dissertation.}
The operation proceeds in the manner described in Chomsky (1992). \( \alpha \) is selected as the affectee and \( X' \) is selected as the target, which is extended to \( X'' \) so that \( \Delta \) is inserted in the Spec position. \( \alpha \) is substituted for \( \Delta \), leaving a trace in the original position. The depth of \( \alpha \) in the input is \{\( Y'' \)\} and that in the output is \{\( X'' \)\}. Therefore, the depth of this operation is \{\( Y'', X'' \)\}.

Take a binary operation, which selects two separate phrase markers as its affectee and target so as to combine them to yield a single phrase marker. In this case the depth of the affectee in the input is \( \varnothing \), and therefore the depth of the operation is equivalent to the depth of the affectee in the output, according to the definition in (5).

There may be cases where the target is a phrase marker contained in a larger phrase marker. In other words, a target may be a subphrase marker of a maximal phrase marker that is a constituent of the input structure. A covert raising of an object to Spec AGRoP provides such an example. Consider the following:
The affectee is α, which has the depth \{Z'', Y'', X''\} in the input. The target is Y', which is extended to Y'' to have Δ. As a consequence, two Y'' nodes appear in the output structure. Let us assume that these two Y'' nodes represent a one-category maximal projection of Y.\(^5\) Thus the depth of α in the output is \{Y'', X''\}. The depth of this operation is therefore \{Z'', Y'', X''\}. In cases like this, where the phrase marker remains unextended, the depth of the operation is equivalent to that of the affectee in the input.\(^6\)

Next consider adjunction operations such as the following:

\(^5\) This assumption is not trivial graph-theoretically, though it is set-theoretically (see Lasnik and Kupin (1977)). It may be possible to assume that the two nodes in question are two segments of one category as we will soon come to assume concerning adjunction.

\(^6\) Under Chomsky’s (1992) usage, the target of the operation in question is the entire phrase marker rooted by X''.

-14-
In the input the depth of the affectee $\alpha$ is $\{Y" , X"\}$. In the output two $X"$ nodes have been created by adjunction. They represent a two-segment category, which is a maximal projection of $X$, given the segment-category distinction proposed by May (1985) and adopted by Chomsky (1986, 1992), among others. Thus, the depth of the affectee in the output is $\emptyset$, since the maximal projection $X"$ does not dominate the affectee, the affectee being regarded as the newly created copy of $\alpha$. Therefore, the depth of the operation is equivalent to the depth of the affectee in the input, namely $\{Y" , X"\}$.

We further propose that the depths of two elements/operations are compared with each other in the following way:

\begin{equation}
\alpha \text{ is shallower than } \beta \text{ if and only if the depth of } \alpha \text{ is properly included in the depth of } \beta.
\end{equation}

Consider the following, for example:
Since the set that is the depth of $\alpha$ is a proper subset of the set that is the depth of $\beta$ and the set that is the depth of $\gamma$, $\alpha$ is shallower than $\beta$ and $\gamma$ since there is no inclusion relation between the depth of $\beta$ and the depth of $\gamma$, neither of $\beta$ and $\gamma$ is shallower or deeper than the other. The comparison of Move-$\alpha$ operations in depth proceeds along the same line.


Let us start with a discussion of the superiority and nesting effects exemplified in (1) and (2):

(1) a. who did you persuade $t$ [ to buy what ]
   b. *what did you persuade who [ to buy $t$ ]

(2) a. *what did you decide [ who [ to persuade $t$ [ to buy $t$ ] ] ]
   b. *who did you decide [ what [ to persuade $t$ [ to buy $t$ ] ] ]

-16-
If we assume that (a) and (b) in each case are yielded by competing derivations, then a possibility will be made open that the contrast can be accounted for under Economy considerations. One of the fundamental Economy principle of Chomsky (1989,1992) is Fewest Steps, which requires that the derivation with the fewest steps blocks all other competing derivations. To appeal to this principle, we have to be able to make (a) and (b) different from each other in terms of the number of derivational steps in (1) and (2). Shallowness will provide us with a necessary implement for this purpose. Let us see how it works.

First consider the superiority case. The examples in (1) contain two wh-phrases. Both wh-phrases have a wh-feature to get checked by a feature of C. One wh-phrase is overtly raised to Spec CP. We assume that the other wh-phrase is covertly raised and adjoined to Spec CP.

A possible derivation for (1a) is illustrated in the following, irrelevant details being omitted for the sake of discussion:

\[(11) \quad \text{LF} \quad [2] \quad [1] \quad \text{CP who [IP t [CP what] [CP who [IP t [CP what]]]}}\]

In overt syntax the wh-phrase who is raised to Spec CP, and the wh-phrase what remains in situ. To get its wh-feature checked, what is raised in covert syntax, adjoining to who. Similarly, (1b) will have the following derivation:
Here the *wh*-phrase *what* is overtly raised and the *wh*-phrase *who* is raised at LF.

These two derivations are not distinguished from each other in terms of the number of steps. However, derivation (12) violates Shallowness, assuming that if *who* is overtly raised instead of *what*, it will provide an alternative operation to the overt movement of *what* [1] in (12). For the movement in question is shallower than [1], since every maximal projection dominating *who* dominates *what* but not every maximal projection dominating *what* dominates *who*. Thus, the movement of *who* would be shallower than the movement of *what*.

The following derivation for (1b) satisfies Shallowness:

\[(12)\]

```
[CP what [IP who [CP t]
```

In overt syntax *what* is first adjoined to a maximal projection which is as shallow as *who*, then raised to the Spec of the matrix CP. For concreteness, let us assume that the maximal projection which the first Move-α operation targets is the complement CP, having at some point of derivation the following VP structure, which involves a Larsonian recursion:
(14)

```
(14)  
  VP       
  /       /  
 you    V'    
  /       |      
 persuade VP  
  /       /  
 who    V'    
  /      /    
 tv    CP     
         /     
        .... what ....
```
Once *what* is adjoined to the complement CP, it is as shallow as *who* since there is no maximal projection intervening between them.\(^7\) Thus the second movement of *what* \(^2\) is as shallow as a hypothetical overt raising of *who*.\(^8\)

A question arises here how the first movement of *what* in (13) satisfies Shallowness. We do not want it to be the case that the movement of *who* to the matrix Spec CP still counts as an alternative operation, since a derivation involving this operation would block the wanted derivation under Shallowness before Fewest Steps is brought into play.

\(^7\) If the CP is inserted in Spec VP, we will have the following VP structure:

(i)

```
(1) VP
    +---+---+
    |   |   |
you V'
    +---+---+
    |   |   |
persuade VP
    +---+---+
    |   |   |
who V'
    +---+---+
    |   |   |
tv VP
    +---+---+
    |   |   |
CP V'
    +---+
       .... what ....
```

In this case it is the lowest VP rather than CP that is as shallow as *whom* so as to serve as the target of the intermediate adjunction.

Another possible way to get as shallow as *whom* is to adjoin to it. If this option is valid, the present discussion will not be essentially affected.

\(^8\) Under the Strict Cycle, which we will return to later, *whom* must be inserted in its original position before the raising of *what* to the matrix Spec CP. However, *what* is in principle allowed to be adjoined to the intermediate CP either before the insertion of *whom* or after it in the present system, which does not assume Form-Chain and therefore does not require that \(^1\) is immediately followed by \(^2\). For a related discussion, see Collins (1992), who takes a Form-Chain approach.
Let us now propose the following conditions to determine the domain of alternative operations from which the shallowest one is chosen:

(15) Two Move-\(\alpha\) operations are alternative to each other if and only if
   a. they have the same input structure,
   b. they are movements to the same position, and
   c. they are motivated by the feature-checking operations applied in the same position.

Following Chomsky (1992), we assume that the structure at each point in a derivation is a set of phrase markers. A subphrase marker \(\alpha\) in a phrase marker \(K\) is in the same position as a subphrase marker \(\beta\) in a phrase marker \(K'\) if and only if \(K\) and \(K'\) are identical except for \(\alpha\) and \(\beta\). A Move-\(\alpha\) operation \(M\) is motivated by a feature-checking operation \(C\) if and only if \(C\) is the first feature-checking operation that \(M\) is connected to. The binary relation "is-connected-to" is a transitive and antisymmetric one that is defined on operations: an operation \(O\) is connected to \(O'\) if and only if \(O\) precedes \(O'\) in the derivation and \(O'\) affects (move or feature-check) the element that has undergone \(O\).

The (first) feature-checking operation that the intermediate adjunction of \textit{what} in (13) is connected to is the feature-checking on \textit{what} in the matrix Spec CP. The one-step raising of \textit{who} to Spec CP would also induce a feature-checking operation there. Thus (15c) as well as (15a) is met. But the two Move-\(\alpha\) operations are movements to different positions, and therefore they are not alternative operations.

Now we are in a position to compare derivation (11) for (1a) and derivation (13) for (1b) in terms of the number of derivational steps.
Suppose that it is a single Move-\(\alpha\) operation that counts as one step. Under this assumption derivation (11) is one step shorter and therefore more economical than derivation (13), so that the latter is ruled out in favor of the former. Note that (11) and (13) yield different LF structures. We will consider below how the comparison domain of derivations is determined.

Let us now turn to the nesting effect which is exemplified in (2):

(2)  

a. \(\?\)what did you decide [ who [ to persuade t [ to buy t ]]]

b. *who did you decide [ what [ to persuade t [ to buy t ]]]

Here we would like to make two assumptions to account for the nesting effect in a parallel way with the superiority effect. Suppose first that the derivations proceed in such a cyclic way that the raising of a \(wh\)-phrase to the lower Spec CP precedes the one to the higher Spec CP.\(^9\) Suppose further that not only a \(wh\)-phrase in situ but also a \(wh\)-phrase in Spec CP is regarded as interfering with the raising of another \(wh\)-phrase, a position which was taken by Chomsky (1973) in discussing the superiority condition.\(^10\) The first assumption actually suffices to make us develop a parallel analysis for the nesting effect. However, the second assumption is motivated by the fact that it enable us to explain ECP phenomena on the basis of the same notions employed to explain superiority and nesting effects. We will see this later.

The assumption in question is straightforwardly derived by the assumption that the \(wh\)-feature of a \(wh\)-phrase in Spec CP does not disappear

\(^9\) We will discuss the Strict Cycle later. The assumption in the text is sufficient for the present purpose.

\(^10\) This position was reconsidered in Noam Chomsky’s class (1990 Fall). Cheng and Demirdash (1990) independently argue for a similar position.
after its checking, as opposed to Chomsky's (1992) position, and behaves to Shallowness indistinguishably from the unchecked wh-feature of a wh-phrase in situ.

Let us see how these assumptions work. The following derivation yields the nesting example (2a):

(16) ![Diagram](attachment:chart.png)

In (16) who is first raised to the lower Spec CP, and then what is raised to the higher Spec CP. However, what cannot reach its surface position in a single step, since there is an intervening wh-phrase, namely who in the lower Spec CP. The raising of who from the lower Spec CP to the higher one, in which its wh-feature is again checked, counts as a shallower alternative operation. Thus, what must take an additional step to get as shallow as the lower Spec CP before it raised to the higher CP. It must be adjoined to the lower IP for that purpose.

Although a wh-phrase in Spec CP has the ability of undergoing further movement and wh-checking, a derivation in which this ability is actually exercised to give rise to double checking will be ruled out under Economy considerations. We will come to this matter soon.\footnote{We would have a similar situation if the IP-adjoined position were in the checking domain of C. The raising of what from the IP-adjoined position to the higher Spec CP in (16), for example, would induce two wh-checking operations. Thus we are forced to assume that the IP-adjoined position is inside the complement domain of C and therefore outside its checking domain.}

Chomsky (1992) defines the complement domain of a head in terms of domination by its complement. Thus an adjunct to the complement is not in
The following derivation yields the crossing example (2b):

(17) \[\text{[CP who [IP [CP what [IP t [CP t]]]]]}\]

When *what* is raised, it will have to cross over *who*. This leads to one extra step.\(^{12}\) Suppose that derivations (16) and (17) compete with each other under Fewest Steps. The former is chosen over the latter as a more economical derivation. Thus we have seen how the superiority and nesting effects are derived in parallel way from Shallowness and Fewest Steps.\(^{13}\)

the complement domain but in the checking domain. If the complement domain is defined in terms of containment, an adjunct to the complement is excluded from the complement domain and enters into the checking domain. Chomsky (class lectures 1991 Fall) left it open which definition should be taken, probably due to the lack of deciding argument. In the present system it is obligatory to adopt the latter definition.

\(^{12}\) For the purpose of the present discussion, we assume that *what* is adjoined to CP to get as shallow as *who*, just as in derivation (13), a superiority violation case. However, it is actually the case that *who* is raised to its Case-checking position, namely Spec AGRoP, prior to the raising to Spec CP, so that the intermediate adjunction of *what* should be not to the CP but to the maximal projection which is as shallow as Spec AGRoP, namely the complement of AGRo, which we will later propose is ASPP.

\(^{13}\) We pointed out in the footnote 2 that the nesting and superiority effects are not observed in the two *wh*-phrases involved are not originaly in some hierarchical relation. Consider the examples presented there:

(i) a. ? what \(j\) do you wonder [ who \(i\) he persuaded friends of \(t_j\) [ to do \(t_j\) at the party ]]
   b. ? who \(i\) do you wonder [ what \(j\) he persuaded friends of \(t_j\) [ to do \(t_j\) at the party ]]

(ii) a. ? whom did you persuade friends of \(t\) [ to buy what ]
    b. ? what did you persuade friends of whom [ to buy \(t\) ]

This fact is expected under our analysis with Shallowness. We have defined the depth of an element/operation as a set of dominating maximal
Unlike us, Chomsky (1992) assumes that a feature disappears as soon as it is checked in the course of derivation. Similarly, a feature which checks another feature automatically disappear once it plays its own role. He assumes further that checking features are strong or weak, and that strong features are visible and weak ones are invisible at PF. Assuming checking features are all illegitimate PF-objects, strong ones must do their work to disappear by SPELL-OUT in order for the derivations to converge at PF.

In the present system it must be alternatively assumed that a feature is maintained through the derivation and that unnecessary and illegitimate features are deleted at the levels of PF and LF under the Economy of representation. If a PF-illegitimate feature is strong, then it must be deleted there. It is reasonable to assume that a feature can be deleted if only if it has performed its own function, if any. To be deleted at PF, a checking feature must finish checking by SPELL-OUT.

Noam Chomsky (personal communication) has pointed out that a question here is how it is indicated at PF whether a feature has performed its function, which would never arise in Chomsky's system. Although a device for that purpose is readily supplied, it will introduce an additional complexity in the theory.14

The problem is actually more general. An assumption implicit in the above discussion, which will be necessary in discussions below, is that it

---

14 The same device will be also needed for Greed, an Economy principle. Later we will give our own formulation of Greed and show that it helps explain ECP effects better.
must be indicated whether an element has undergone a particular Move-α operation. This is not a trivial assumption under the present system, which takes Move-α rather than Form-Chain as transformation. A minor device that registers the results of Move-α and checking operations is needed to make the entire system work.

We are not just importing something new here. Rather we are trading. Once the registering device in question is incorporated, it open a way to explain the superiority, nesting and ECP effects on the same Economy principles so as to dispense with the assumptions that otherwise would be needed but not independently motivated. There is no reason to think that we are now making a losing bargain.

Before we leave this section, let us consider some assumptions hidden in the discussion above. First, we have been ignoring traces. Take derivation (16), for example. The intermediate IP-adjunction of what ([2]) crosses over the trace of who. Under the copy theory of movement, a trace is simply a full copy of the moved element, so that they should include the same set of features. If it is possible to move a trace, then the IP-adjunction of who should count as an alternative operation to movement [2]. Thus Shallowness will force what to first get adjoined to the most deeply embedded CP just as in (17), so that the derivation will involve the same number of steps as (17), making it impossible to distinguish between the derivations in terms of the number of steps, a result that we should be able to avoid.

Let us propose the following condition:
(18) An element can be affected by an operation if and only if it is the highest among the copies that are linked with each other by Move-α.

This condition is applied to Move-α and feature-checking operations. Given a structure as the input of an operation, the affectee has a form of phrase marker, which may or may not be a subphrase marker of a larger phrase marker. The root category of a phrase marker α, call it $C_\alpha$, is the category to which the root/top node of α belongs as one of its segments in the maximal phrase marker containing α. A phrase marker α is higher than a phrase marker β if and only if every segment dominating $C_\alpha$ dominates $C_\beta$, and some segment dominating $C_\beta$ does not dominate $C_\alpha$.  

When what is IP-adjointed in (16), the trace copy of who is lower than the copy in Spec CP, so that it cannot be affected by a Move-α operation. Thus the IP-adjunction of what does not have an alternative operation, satisfying Shallowness.

Condition (18) is not a simple restatement of the idea that traces are unmovable. It does not prohibit a trace from being moved unless it has been left by a raising operation. And it covers feature-checking operations as well as Move-α operations. Let us see some consequences relevant in the present discussion.

Take derivation (16) again. The adjunction of what to the lower IP ([2]) would violate Shallowness if the lowering of who from Spec CP to adjoin to the IP counted as an alternative operation, since the latter is shallower than the former.

---

15 This definition of height will have to be modified so as to deal with head movement and binding.
However, Condition (18) allows no further Move-α operation to affect *who* in the IP-adjointed position, since the trace copy is higher in Spec CP. That is, the IP-adjunction of *who* cannot be connected to and therefore cannot be motivated by any feature-checking operation, in particular the feature-checking operation applied in the matrix Spec CP. Given (15c), the lowering IP-adjunction of *who* in question does not count as an alternative operation to any movement, in particular the IP-adjunction of *what* ([2]) in (16).

Condition (18) straightforwardly derives another implicit assumption: no element can be adjoined to itself. Take derivation (11), for example:

\[(11) \quad \text{[CP} \quad \text{[IP} \quad \text{t} \quad \text{[CP} \quad \text{what} \quad \text{[1]} \quad \text{LF} \quad \text{[2]} \quad \text{[CP} \quad \text{who} \quad \text{IP} \quad \text{t} \quad \text{[CP} \quad \text{what}\]
\]

Here *what* is adjoined to *who* in Spec CP ([2]). If adjunction of *who* to itself counted as a alternative operation to [2], then Shallowness would be violated. However, Condition (18), which covers feature-checking operations, prohibits the adjoined *who* from being affected by a feature-checking operation, since the trace copy in Spec CP is higher under our definition of height. Thus the adjunction of *who* to itself, which is a lowering in our sense, is not motivated by a wh-checking operation, so that it does not count as an alternative to [2].

Derivations which involve lowering operations such as the two cases just discussed are ruled out by Greed, an Economy principle of Chomsky (1992), which captures the idea that an element undergoes movements only
to get its own requirements satisfied. We will below give a definition of this principle.

4. An Economy system.

4.1. A reformation of Economy principles.

Note that we have different concepts of Economy from those shared by recent works on this matter such as Chomsky (1989, 1992, class lectures), Chomsky and Lasnik (1991), Collins (1992).

First of all, we assume that it is not the Form-Chain operation but the Move-α operation that counts as one step in a derivation, contrary to Chomsky’s (1992) position. Only under this assumption are the superiority and nesting effects provided with an explanation based on Fewest Steps. For good and bad cases cannot be distinguished in terms of the number of chain-formations.  

Chomsky (1992) takes Shortest Move as well as Fewest Steps as a fundamental Economy principle. As far as a single application of Move-α counts as one derivational step, the two principles in question impose two conflicting requirements. In order to reduce the number of movements in a derivation we must have longer movements. In order to reduce the distance of movements we must have more movements. Chomsky’s resolution of this conflict is to take chain-formations as derivational steps.

However, whether such a conflict is really undesirable or not depends on how the distance of a movement is calculated. If the distance of a movement is, say, the number of maximal projections it crosses, then a

---

16 We could make a distinction in terms of the number of Shallowness violations if Shallowness were taken to constrain chain-formations rather than movements. However, the present system provides an unified explanation for various ECP effects, as we will see.
conflict will usually arise. To reduce the length of every movement, it is necessary to adjoin to every maximal projection, as illustrated as follows:

(19) a.  

b.  

Every sentence has alternative derivations like these, if we assume that head movements and movements for Case-checking, for example, crosses at least one maximal projection. If it is chain-formations that counts as derivaional steps, then (19b) will be ruled out in favor of (19a) since Shallowness rates (19b) more economical than (19a) while Fewest Steps rates them equally economical, assuming that the successive movements form one chain in (19b). If derivational steps are movements, then Fewest Steps will rates(19a) more economical than (19b), yielding a conflict with Shallowness. This conflict could be interpreted in two ways. The derivations for a given sentence are all ruled in, or they are all ruled out. Under the former interpretation Fewest Steps and Shortest Move do not work at all to rule out derivations. Under the latter interpretation no grammatical sentence is derived. Thus the conflict in question is a serious problem that should be resolved before anything else.

On the other hand, if the distance of a movement is the number of possible landing sites that the movement crosses (a version of Relativized Minimality), the situation is less clear. Consider the following two derivations, which compete with (16):
In (20) what is raised to the matrix Spec CP by a single movement. In (21) what first moves to adjoin to who, which has been raised to the embedded Spec CP, and then it further moves to the matrix Spec CP. Suppose that who in Spec CP is “crossed” in (20) but not in (21) in the relevant sense. Then a conflict arises between Fewest Steps and Shortest Move. This will be a desirable situation if both derivations are taken to be illegitimate, since it is possible to assume that all derivations in conflict are ruled out. It depends in part on how the derived structure, namely (2a), is evaluated with regard to its acceptability, which has been controversial as we noted. We take it to be “good,” but we admit that it is also possible to take it to be “bad” and develop a theory on such a factual assumption.

In any case, suppose for the sake of discussion that the conflict of the kind under consideration must be resolved. Chomsky’s solution is straightforward but not the only possible one. Another solution is to “order” Fewest Steps and Shortest Move, maintaining the assumption that derivational steps are movements. Shortest Move precedes Fewest Steps in examining a set of competing derivations. That is, Fewest Steps only compares the derivations that have passed Shortest Move. Thus (20) is blocked by (21) under Shortest Move, which is the same result as is obtained under Chomsky’s solution.
This kind of ordering is employed in any theory. The requirement that derivations must converge can be regarded as a principle preceding Fewest Steps, for example. Thus, introducing an ordering itself is methodologically not a harmful move.

Under our present system, which does not incorporate Shortest Move, the assumption that derivational steps are movements does not give rise to a conflict of the kind seen above. But we need to impose orderings on Economy principles in a similar way. Thus Shallowness must apply prior to Fewest Steps.

One might think that the ordering just mentioned between Fewest Steps and Shallowness would automatically follow as intrinsic if Shallowness, unlike Fewest Steps, were not a transderivational constraint but an (intra)derivational one in the sense that a derivation is evaluated without referring to other derivations. This is a wrong idea, however. It presupposes that the derivational constraints precede the transderivational ones. There is no reason to think that this presupposition is a priori correct. Rather it is just a claim, which calls for an empirical justification.

Moreover, Shallowness is surely transderivational by nature. Comparing alternative operations is comparing alternative derivations involving those operations. It is just for convenience that we treated it as if it were a derivational constraint. No matter how it is described, its transderivationalness cannot be dismissed. Hereafter, we will try to be explicit in this regard.

To see our problem, let us now compare (20) and (21) with (16), reproduced here:
Note that these derivations cannot be distinguished in terms of Shallowness. This is because the domain of alternative Move-\( \alpha \) operations with respect to Shallowness is defined by (15), reproduced here:

(15) Two Move-\( \alpha \) operations are alternative to each other if and only if
    a. they have the same input structure,
    b. they are movements to the same position, and
    c. they are motivated by the feature-checking operations applied in the same position.

We want to rule out (20) in terms of Shallowness by saying that the raising of \textit{what} in (20) has a shallower alternative operation. But neither (16) or (21) involves such an operation. The raising of \textit{what} to the matrix Spec CP has a different input structure both in (16) and (21) not alternative to any operation in (16) and (21). The relevant input structure in each case is represented as follows, supposing for the sake of discussion that it includes a single phrase marker:
What is adjoined to IP in (16'), in situ in (20') and adjoined to the embedded Spec CP. Thus the operations having these input structures never count as alternative to each other, by virtue of (15a).

The following derivation blocks (20) under Shallowness:

\[(22)\]

First who is raised to the intermediate Spec CP, and then it is raised from there to the matrix Spec CP. In covert syntax what is raised to the intermediate Spec CP. This derivation has the following ill-formed overt structure:\(^{17}\)

\[(23)\]

* who did you decide [CP t' [IP to persuade t [CP to buy what ]]]

\(^{17}\) The corresponding surface form is acceptable under an irrelevant usage of decide, namely, if it is interpreted as taking a non-interrogative control clause.
The LF structure is the same as that of (17), assuming that the intermediate trace in the embedded Spec CP is ignored or deleted. Given (15), derivations (20) and (22) involve alternative operations: the movement of what [2] in (20) and the second movement of who [2] in (22). (20) and (22) compete with each other, and Shallowness chooses the latter over the former.

The problem that calls for the ordering of Shallowness and Fewest Step is the following. Fewest Steps would choose derivation (20) over derivations (16) and (21). Therefore, unless (20) has been ruled out when Fewest Steps comes into play, then (16) and (21) would be ruled out. Since (20) is independently ruled out by Shallowness, there would be no derivation yielding (2a), contrary to our expectation. If Shallowness applies prior to Fewest Steps, (20) is correctly ruled out by Shallowness before Fewest Steps.

Derivation (22) must be ruled out too, though it converges and blocks (20). Notice that who undergoes two wh-checking operations in this derivation. This counters the idea of Last Resort. We will restate Chomsky's (1992) principle of Greed, a self-serving Last Resort, as follows:

(24) Greed

A Move-α operation must be motivated by a nonredundant feature-checking operation.

A Move-α operation is motivated by a feature-checking operation if and only if the latter is the first feature-checking operation that the former is connected to. An operation is connected to another if and only if the former precedes the latter in the derivation and they affect (move or feature-check)
the same element. And no connection can be established if SPELL-OUT intervenes. Greed is regarded as a condition to decide the legitimacy of a convergent derivation rather than as a prerequisite for convergence.

Greed also rules out (21). Thus, among the four derivations under consideration, only (16) passes Shallowness and Greed. And derivation (17) also passes these two principles. (16) and (17) enter into a comparison by Fewest Steps, and (16) wins out.

We will see later how our Greed plays a role in explanation for ECP phenomena, given the assumption that checked features do not disappear. Later we will turn to the discussion of this matter. We will then argue that it may not be the case that Greed precedes Shallowness in application, and also that it must be the case that it precedes Fewest Steps.

Another Economy principle of Chomsky (1992) is Procrastinate: covert operations is cheaper than overt ones. Thus overt syntax must be minimal in terms of the number of devizational steps. We have a slightly different concept. The idea is that it costs more to produce directly observable effects. Let us restate this as follows:

(25) Procrastinate

The overt raising of an element must be minimal.

This principle applies to a set of convergent derivations yielding the same LF structure.

Consider the following, to see how Procrastinate works:

(26) a. John [VP bought the book ]
    b. John [AGRO the book [VP bought ti ]]
c. \[[CP \text{ what}_i \text{ did } \text{ John} \ [\text{VP buy } t_i]]\]

d. \[[CP \text{ what}_i \text{ did } \text{ John} \ [\text{AGR}_{oP} t'_i \ [\text{VP buy } t_i]]\]

In (26a) *that book* remains in situ in overt syntax and is covertly raised to Spec AGRoP for Case-checking. Assuming that no feature involved in Accusative Case-checking in English is strong in the sense of Chomsky (1992), this example has a convergent derivation. (26b) also has a convergent derivation, *that book* being overtly raised. These two examples will have the same LF structures, and therefore their derivations compete with each other. There is a sense in which *that book* is overtly raised farther in (26b) than in (26a). For preciseness, let us suppose that how far an element is raised is equivalent to how many maximal projections it is extracted from and that an element \( \alpha \) is extracted from a maximal projection \( \beta \) if and only if Move-\( \alpha \) moves \( \alpha \) from a position which is contained by \( \beta \) to a position which is not contained by \( \beta \). Thus, (26b) is ruled out in favor of (26a) by Procrastinate, since the derivation for (26b) involves a longer overt movement of *that book* than the derivation for (26a).

In (26c) *what* is immediately raised to Spec CP in overt syntax. The trace of *what* must be raised to Spec AGRoP for Case-checking in order for the derivation to converge.\(^{18}\) In (26d), on the other hand, it is raised through Spec AGRoP. In both cases a morphological requirement on C forces the overt raising of *what* That is, the failure of an overt raising leads to a crashing derivation. Procrastinate does not distinguish between the

\(^{18}\) It is not unreasonable to allow this movement under Condition (18) by assuming that no element outside the domain of the relevant checker counts as a higher copy.
convergent derivations for (26c) and (26d), since *what* overtly ends up in the same position. Thus both examples pass Procrastinate.

Chomsky’s version of Procrastinate rules out (26d) in favor of (26c), since *what* is subject to two operation in (26d) and one operation in (26c) in overt syntax, in terms of either Move-α or Form-Chain. We will later argue that a *wh*-phrase can and must be raised through its Case-checking position, supporting our version of Procrastinate.\(^{19}\)

One thing we have to be concerned about here is the VP-internal subject hypothesis. To make the subject and the object raise to Spec AgrsP and Spec AGRoP, respectively, under this hypothesis, Chomsky (1992) proposes a version of relativized minimality, which is supplemented with the notion of equidistance. However, his analysis does not allow the object to raise in overt syntax in languages like English, which does not raise V.\(^{20}\) We actually do not adopt Chomsky’s analysis, and therefore the problems doesn’t arise in the first place. But the fundamental question, which motivates Chomsky’s analysis, is still present: Why is it the case that the subject ends in the higher Spec AGRP and the object in the lower one? We simply drop the VP-internal subject hypothesis, and assume that the subject is generated in Spec of some head which is higher than AGRo but lower than TNS in the first place and then raised from there for Case-checking in the course of derivation. For arguments against the VP-internal subject hypothesis, see Koizumi (1993) among others.

\(^{19}\) Branigan (1992) also argues that *wh*-raising through SpecAGRoP is possible. Instead of revising Procrastinate, however, he proposes that a feature involved in Accusative Case-checking can be optionally strong.

\(^{20}\) Branigan (1992) gives a solution, but it does not fit into our theory.
Now we have the following picture to show how a comparison of competing derivations proceeds:

\[(27)\]

\[\text{Procrastinate} \rightarrow \text{Shallowness} \rightarrow \text{Greed} \rightarrow \text{Fewest Steps}\]

The derivations that compete with each other first undergo evaluation by Procrastinate. Second, the surviving derivations undergo evaluation by Shallowness and Greed. Finally the derivations that still survive undergo evaluation by Fewest Steps.\(^{21}\)

4.2. Comparison domains.

Now let us tum to another departure from the above mentioned recent works on Economy. With Shallowness and Fewest Steps we compare two derivations which yield different LF structures, while the above mentioned works assume that two derivations should be compared with each other only if they yield the identical LF structure.\(^{22}\) Our position is much closer to that

\(^{21}\) Greed is a derivational (or intraderivational) constraint that only examine a derivation without referring to other derivations, whereas the other principles are transderivational in the sense that they give a relative evaluation of a derivation by comparison with others. Thus, it will not make a difference if Greed is ordered between Shallowness and Fewest Steps. For a derivation violating Greed still can block a derivation under Shallowness.

The present system has the same number of Economy principles as Chomsky's (1992), who has Shortest Move instead of Shallowness, though we employ different definitions. We consider the requirement that derivations must converge to be ordered prior to all Economy principles.

\(^{22}\) Following Chomsky's (1992) generalized transformation approach, we assume that the structure at each point in the derivation is a set of phrase markers. The structure at a point after SPELL-OUT is a single phrase
of Epstein (1992), who takes a Move-α approach and does not restrict the comparison to the derivations yielding the identical LF structure, although we restrict the comparison domain in a different way.

By a comparison domain, we refer to an equivalence class of derivations, which compete with each other under the Economy principles. All the derivations in a comparison domain in a block enter into an examination by the Economy system in (27).

Given a convergent derivation D, its comparison domain is the set of all convergent derivations yielding LF structures that are homomorphic with the LF structure of D. The relevant notions are defined as follows:23

(28) A phrase marker K₁ is homomorphic with a phrase marker K₂ if and only if K₁ is identical to K₂ except that the interchange of a subphrase marker C₁ with a subphrase marker C₂ in R(K₁) yields R(K₂), where C₁ and C₂ are of the same type.

(29) R(K) is the reduced phrase marker of K, which is yielded by disjoining every adjunct to reduce every multiple-segment category to a single-segment category.

---

marker. Thus, by an LF structure we refer to the phrase marker at the final stage in the derivation. We will ignore intermediate traces. They can be regarded as invisible since they are internal to LF objects, which are chains. Or they can be regarded as being deleted under the Economy of representation. We will keep away from “reconstruction,” which we consider a process applied after the transformational derivation so that it is not subject to Economy principles.

23 I am indebted to Noam Chomsky (personal communication) for suggesting to me the notion of reduced phrase marker. Note that our usage of this term should not be confused with Lasnik and Kupin’s (1977).
Take the superiority case we discussed above. Derivation (11) yields (1a) and derivation (13) yields (1b):

(1) a. who did you persuade t [to buy what]
b. ??what did you persuade who [to buy t]

(11) \[ LF \]
\[
\text{CP} \quad \text{who} \quad \text{IP} \quad \text{t} \quad \text{CP} \quad \text{what} \\
\]

(13) \[ LF \]
\[
\text{CP} \quad \text{what} \quad \text{IP} \quad \text{who} \quad \text{CP} \quad \text{t} \\
\]

(11) and (13) have been assumed to compete with each other. They yield the following LF structures, which are represented graph theoretically, irrelevant details being omitted:

(30) a. PM yielded by (11) b. PM yielded by (13)

The reduced phrase markers of the phrase markers in (30) are as follows:
(31) a. RPM yielded by (11) 
RPM yielded by (13)

The interchange of the subphrase markers rooted with DP$_i$ and DP$_j$ in (31a), namely the phrases *what* and *whom*, yields (31b). Thus the phrase makers in (30) are homomorphic, so that the derivations yielding them enter an equivalence class for comparison.

The examples in (2) have derivations (16) and (17), which yield LF structures (32a) and (32b), respectively:

(2) a. *what did you decide [ who [ to persuade t [ to buy t ]] ]
   b. *who did you decide [ what [ to persuade t [ to buy t ]] ]

(16) \[
\text{[CP what [IP [CP who [IP t [CP t]]]]]}
\]

(17) \[
\text{[CP who [IP [CP what [IP t [CP t]]]]]}
\]
Since no adjunction is involved in the relevant part, these phrase markers are equivalent to their reduced phrase markers. Since the interchange of the phrases *what* and *who* in (32a) yields (32b), the derivations in question yield homomorphic LF structures so that they are subject to comparison.

Generally, two identical LF structures are homomorphic. Given the definition in (28), take the case where \( K_1 = K_2 \), so that \( R(K_1) = R(K_2) \). The interchange of \( C_1 \) and \( C_2 \) in \( R(K_1) \) yields \( R(K_2) \), where \( C_1 = C_2 \). That is, if one and the same subphrase marker is selected as \( C_1 \) and \( C_2 \), then the interchange is vacuous. For example, the interchange of the subphrase marker rooted by \( DP_i \), namely the phrase *whom*, with itself in the reduced phrase marker (31b) vacuously yields (31b). Therefore the phrase marker (30b) is homomorphic with itself. This phrase marker is yielded not only by derivation (13) but also by derivation (12). These two derivation yield identical and homomorphic LF structures, so that they (12) would block (13) under Fewest Steps. But Shallowness compares the two derivation before Fewest Steps and rules out (12). Thus, (12) never get the chance to block (13). (13) is blocked by (11) under Fewest Steps.
Now consider the following example:

(33)  \[ \text{who}_i \text{ t}_i \text{ decided [ who}_j \text{ [ to persuade t}_j \text{ [ to buy what}_k \text{ ]]]} \]

This example is not only acceptable but also ambiguous in interpretation. The wh-phrase in situ, \textit{what}, can be interpreted with matrix scope or with embedded scope. These two readings are yielded by the following derivations, which satisfy Shallowness:

(34) a. \[
\begin{array}{c}
\text{LF} -[4]\text{--}[3]\text{--}[1]
\end{array}
\]
\[
[ \text{who}_i \text{ [ t}_i \text{ [ who}_j \text{ [ t}_j \text{ [ what}_k \text{ ]]]} ]
\]

b. \[
\begin{array}{c}
\text{LF} -[3]\text{--}[1]
\end{array}
\]
\[
[ \text{who}_i \text{ [ t}_i \text{ [ who}_j \text{ [ t}_j \text{ [ what}_k \text{ ]]]} ]
\]

One might think that since at the point of the raising of the phrase \textit{who}_j to the embedded Spec CP the phrase \textit{who}_i has not been introduced in the targeted phrase marker, the generalized transformation that introduces \textit{who}_i into that Spec CP would be a shallower alternative operation. Note however that the phrase \textit{who}_i is not directly drawn from the lexicon. Since it is an independent phrase marker, it must be constructed outside the lexicon. Thus it can be the case that the movement of \textit{who}_j applies prior to the construction of \textit{who}_i, so that the former operation is not ruled out by the latter under Shallowness.\footnote{Similarly, we may be able to assume that X^0 categories are composed of features or feature complexes stocked in the lexicon and are constructed before they are introduced into a phrase marker. Otherwise, it would not be always possible to raise an X^0 category. For the raising of an X^0 category}{24}
The two derivations in (34) yield different LF structures, as follows:

(35) a. PM yielded by (34a)

(35) b. PM yielded by (34b)

The following reduced phrase markers are obtained:

necessarily crosses one maximal projection and a binary Move-\(\alpha\) operation that introduces a new \(X^0\) category has the depth \(\emptyset\). Though we cannot expect that there will be many empirical consequences are not provided, this line of considerations will allow us to treat the singular and binary Move-\(\alpha\) operations in an undistinguished way with respect to Shallowness. We will argue that Shallowness plays the central role in deriving the Strict Cycle, which does not distinguish between singular and binary operations.
A subphrase marker is generally characterized in terms of dominance and precedence relations with others. A disjoined subphrase marker has lost the information concerning dominance, but keeps the information concerning precedence by virtue of the terminal strings. Thus, the phrase what immediately precedes the phrase who in (36a) and the phrase whom in (36b), assuming for concreteness that adjunction is only to the left of the target.25

25 The discussion at hand will not be essentially affected if adjunciation can be to the right, since the two Spec CPs are separated by the matrix materials...
(36a) and (36b) are not equivalent since they have the disjoined adjunct in linearly different positions. There can be no interchange of two subphrase markers in (36a) to yield (36b), so that the phrase markers (35) are heteromorphic. Thus the derivations in (34) are not subject to comparison by Fewest Steps, leading to the ambiguity of (33).\textsuperscript{26,27}

Let us now turn to a restriction on the interchange of subphrase markers in a reduced phrase marker. Roughly speaking, two subphrase markers are interchangeable only if they are of the same type, as is stated in the definition (28). \(C_1\) and \(C_2\) are of the same type if only if there is a feature type \(F\) that is materialized on the root nodes of \(C_1\) and \(C_2\). Thus, as we have so far assumed, two \textit{wh}-phrases are interchangeable since they include \textit{wh}-features.

\textsuperscript{26} If two derivations having the identical “D-structure” must be compared with each other, as Epstein (1922) assumes, derivation (34a) would be incorrectly blocked by derivation (34b).

Epstein does not consider the present case, but it would not provide an argument against him since he assumes that at LF a \textit{wh}-phrase is raised over another by one step. It seems that all the cases that Epstein discusses still can be analyzed under his system, even if he adopts the assumption that derivations compete only if they yield the identical LF structure.\textsuperscript{27}

Lasnik and Saito (1992) considers examples like the following:

(i) \[ \text{wh}_i t_j \text{ wonders [ what}_j [ you told wh}_k [ \text{PRO to read } t_j ]] \]\n
They claim that the \textit{wh}-phrase in situ can not have the embedded scope but can have the matrix scope. This fact suggests that the interchange of the two subphrase markers is further constrained in terms of the structural relation between them, shrinking the comparison domain so as to permit more derivations. The desirable result will be readily obtained by a technical reformulation.
We assume that wh-features are carried not only by interrogative wh-phrases but also by relative pronouns, topicalized phrases and empty operators. All kinds of wh-phrases are of the same type for the purpose of homomorphism. Moreover, they share potential checkers so that they interact with each other with regard to Shallowness. They are distinguished from each other in terms of sub-features, [+Question], [+Relative], [+Topic] and [+Empty]. These features, unlike the super-feature [+wh], do not belong to the class of features that must be checked for convergence. Rather, they are more closely associated with the semantic contents of the phrases on which they are manifested, and play a role in determining whether semantic restrictions are observed.

The position that interrogative wh-phrases and other phrases of the kinds just mentioned form a class has been taken since Chomsky (1977), who showed that they behave alike with respect to island effects.

What is interesting in the present context is that, as Pesetsky (1982) observes, the raising of different kinds of wh-phrases reveals a nesting effect.

Consider the following, for example:\(^{28}\)

(37) a. ?the book was easy [Op_i [to decide [who_j [to persuade t_j [to buy t_i ]]]]]

   b *the boy was easy [Op_i [to decide [what_j [to persuade t_i [to buy t_j ]]]]]

Example (37a) involves a nesting of two wh-phrases, whereas example (37b) involves a crossing. However, the LF structures of (37b) is not

---

\(^{28}\) For concreteness, suppose that the empty operator is probably in Spec CP. As for topicalized phrases, \(w\) will argue later it is in the Spec of some functional head other than C.
homomorphic with that of (37a). Rather, it is homomorphic with the LF structure of (37c), which involving a nesting:

(37) c. *the boy was easy [what_i[to decide [Op_j[to persuade t_j [to buy t_i]]]]]}

Here both what and Op successfully undergo a wh-checking operation by C. (37a) is ruled out in favor of (37c), the former being derivationally less economical than the latter.

Example (37c) has a convergent derivation with what and Op being LF legitimate objects. But it yields a deviant interpretation, failing to satisfy selectional restrictions. The clause headed by what is interpreted as a "question," but easy does not semantically select a question. Similarly, decide semantically selects a question whereas the clause headed by Op is not interpreted as a question.

5. Impossible extractions.
5.1. Adjunct extraction.

Now let us turn to ECP effects. First consider the argument-adjunct asymmetry observed in the extraction from a wh-island:

(3) a. ?what do you wonder [ whether [ John fixed t yesterday ]]
    b. *when do you wonder [ whether [ John fixed it t ]]

We expect that (3) will have the following derivations:
Suppose for the sake of discussion that whether is a wh-phrase inserted in Spec CP by a binary transformation.\textsuperscript{29} Both derivations involve an intermediate IP-adjunction, assuming that when, like what, is raised from a lower position and that whether behaves exactly like other wh-phrase in the relevant respects. The unacceptability of example (3b) suggests that derivation (38b) should be somehow ruled out. We will see below that the Case-theoretic property of adjuncts like when plays a role to yield the desired result.

To devise the necessary mechanism, suppose first that a clause has the following articulated structure:\textsuperscript{30}

\textsuperscript{29} Alternatively, we may assume that whether is C and that its Spec is occupied by an empty operator. Moreover, if whether is raised to Spec CP from somewhere inside the clause, as is argued in Larson (1985), the claim we are making here will remain essentially unaffected, although the discussion would become somewhat complicated.

\textsuperscript{30} The functional element \( \Sigma \) has been proposed by Laka (1990). As for ASP, we will shortly discuss a motivation to posit it.
In overt syntax TNS and ASP are raised to AGRs and AGRo, respectively, to form a complex, regardless whether V is raised or remains in situ. The subject is also overtly raised to Spec AGRsP to be Case-checked by AGRs-TNS complex. Similarly, the object is raised to Spec AGRoP to undergo a Case-checking operation by AGRo-ASP complex. The raising of the subject is always forced in overt syntax, while the raising of the object is not.

Let us use the term "structural Case" to refer to those which are checked by a complex of AGR and TNS/ASP, including Nominative Case and Accusative Case. An instance of structural Case is a complex of morphological agreement features and a morphological case feature. AGR is

---

31 See Chomsky (1992) for the proposal that PRO has Null Case checked by AGRs-TNS.
responsible for the checking of the agreement features and TNS/ASP is responsible for the checking of the case feature. TNS and ASP have some of their features, including Case-checking features, that must be checked by C and V, respectively. From this it indirectly follows that structural Case-checking often depends on the properties of C and V.

We will use the term "nonstructural Case" to refer to other instances of abstract Case. We will further propose that nonstructural Case is not restricted to argumental DPs, broadening the notion of abstract Case, and that it is checked by a functional element. Thus, argumental PPs have nonstructural Case and are externalized out of VP at some point of derivation to undergo a Case-checking operation by ASP. And adjuncts also have nonstructural Case, and ASP and TNS are possible Case-checkers for them. When ASP is raised to AGRo, the checking domain of ASP is extended to include Spec AGRoP and adjuncts to AGRoP as well as Spec ASPP and adjuncts to ASPP.32 Similarly for TNS.

Now derivation (38b) should be illustrated with a more accurate structure, as follows:

(40) \[ \text{[CP when [ \text{[CP whether [AGRoP John [TNSP t} \]

32 Given Chomsky's (1992) definition of checking domain and our slight revision presented in the footnote [CC] above, ASPP-adjuncts as well as Spec ASPP will be no longer within the checking domain of ASP after ASP is raised and adjoined to AGRo, if ASPP is regarded as a complement of the raised ASP. We simply assume that the category XP is a complement of the category Y0 if and only if the first segment dominating XP dominates Y0 and the first segment dominating Y0 dominates XP, so that in the above case the ASPP-adoined positon is within the checking domian of ASP. This revision does not affect Chomsky's discussions.
But this derivation is blocked under Shallowness by the following derivation:

\[(41)\]
\[
\begin{array}{llll}
\text{CP when} & \text{CP whether} & \text{AGRsp John} & \text{TNSp t} \\
\end{array}
\]

First John is adjoined to AGRsP, and then when is adjoined to AGRsP and further raised to matrix Spec CP.

Assuming that adjuncts like when can be Case-checked by TNS as well as ASP, the intermediate AGRsP-adjointed position in (40) serves as a Case-checking position for when. Feature-checking operations, unlike Move-\(\alpha\) operations, automatically apply whenever the structural conditions are met. Thus the feature-checking operation that motivates the AGRsP-adjunction of when is the one applied in the AGRsP-adjointed position. Note that the AGRsP-adjointed position could also serve as a Case-checking position for the subject John, since it has Nominative Case, which can be checked by the AGR-TNS complex. Therefore, the AGRsP-adjunction of John in (41) would be motivated by a feature-checking operation in the AGRsP-adjointed position. It should count as an alternative operation to the AGRsP-adjunction of when in (40) with respect to Shallowness, given (15), reproduced here:

\[(15)\] Two Move-\(\alpha\) operations are alternative to each other if and only if

a. they have the same input structure,

b. they are movements to the same position, and
c. they are motivated by the feature-checking operations applied in the same position.

Shallowness is violated in (40). That is, (40) is blocked by (41) under Shallowness with respect to the AGRsP-adjunction, assuming that when is raised from a deeper position than the Spec AGRsP, which is occupied by John.

Derivation (41) is ruled out by Greed:

(24) Greed

A Move-\(\alpha\) operation must be motivated by a nonredundant feature-checking operation.

For the AGRsP-adjunction of John is motivated by a redundant feature-checking, namely the Nominative Case-checking applied in the AGRsP-adjoined position.\(^{33}\)

The following derivation satisfies Shallowness:

(42) \[ \text{[CP when [ [CP whether [\text{AGR}\_\text{SP} \_\text{John} [\text{TNSP} t \] ] ] ] ]} \]

\(^{33}\) Note that this derivation must be allowed by Procrastinate:

(25) Procrastinate

The overt raising of an element must be minimal.

Otherwise it would not be able to block (40) under Shallowness, since Procrastinate applies before Shallowness. Since the distance of the raising of an element is calculated in terms of maximal projections from which that element is extracted, the raising to the Spec of a head and the raising to the adjoined position to its maximal projection are equivalent in distance.
*When* is first adjoined to TNSP to reach the AGRsP-adjoined position. But the TNSP-adjoined position is also a Case-checking position for *when* by TNS. *When* undergoes two movements, both of which induce a Case-checking operation.

This derivation violates Greed, however. The second movement of *when* is connected to but not motivated by the *wh*-checking operation applied in Spec CP. It is motivated by the Case-checking operation applied in the AGRsP-adjoined position. But this feature-checking operation is redundant since Case has been once checked in the TNSP-adjoined position.

Thus, *when* cannot reach the AGRsP-adjoined position without violating Shallowness or Greed, so that it cannot cross over the *wh*-island.34

From the above discussion it is clear that Greed may not apply prior to Shallowness. If Greed must apply prior to Shallowness, then (41) would be ruled out before it has a chance to block (40) under Shallowness.

A question that has been put aside is whether Greed applies prior to Fewest Steps. Actually we have decided so without any argument, as shown in (27):

---

34 To be more precise, derivation (42) will yield the same LF structure as derivation (i), if the copy of *when* in the TNSP-adjoined position is deleted at LF. If the copy adjoined to AGRsP is deleted, it will yield the same LF structure as the following derivations:

(i) \[
\text{[CP when [CP whether [AGRsP John [TNSP t]]]]}
\]

This derivation violates Shallowness since the second movement of *when* crosses over a shallower *wh*-phrase, namely *whether.*
Now we are ready to provide some argument for this decision.

A prediction is that we can have a grammatical crossing example whose nesting counterpart involves the extraction of an adjunct from a *wh*-island. Any derivation with an adjunct extraction from a *wh*-island must violate Greed if it satisfies Shallowness. If Greed applies prior to Fewest Steps, no such derivation reaches the stage in which Fewest Steps applies. Thus, if such a derivation is a crossing one, it is not the case that it blocks the competing nesting derivations under Fewest Steps.

This prediction is borne out, though the facts are delicate. Consider the following nesting examples first:

(43) a. ?? what_i do you wonder [ to whom_j it seemed t_j [ Mary bought t_i ]]
    b. * when_i do you wonder [ to whom_j it seemed t_j [ Mary bought it t_i ]]

(43a) can be derived by a derivation which satisfies Shallowness and Greed. But derivations for (43) necessarily violates Shallowness or Greed. The crossing counterparts are as follows:

(44) a. * to whom_j do you wonder [ what_i it seemed t_j [ Mary bought t_i ]]
    b. ?? to whom_j do you wonder [ when_i it seemed t_j [ Mary bought it t_i ]]
When Fewest Steps applies, the remaining derivation for (44a) is blocked by the remaining derivation for (43a), revealing an ordinary nesting effect. On the other hand, the derivation for (44b) that has got past Shallowness and Greed goes all the way without competitors, because no derivation for (43b) has survived when Fewest Steps applies. Thus, (44b) is ruled in. This is an amazing result, but it is not beyond expectation for us.

Interestingly, crossing examples like (44b) are not always legitimate. Observe the following:

\[(45)\]
\[
b. \quad * \quad \text{when}_j \text{ do you wonder [ who}_i \text{ I proved } t_i \text{ to have said [that Mary bought } t_j \text{ ]}]
\]

\[(46)\]
\[
b. \quad * \quad \text{who}_i \text{ do you wonder [ when}_j \text{ I proved } t_i \text{ to have said [that Mary bought it } t_j \text{ ]}]
\]

The examples in (45) involve a nesting and the ones in (46) involve a crossing. (46) is not so good as it is expected to be. But this problem is readily solved.
Suppose that when a wh-phrase is raised to Spec CP it must be raised through a Case-checking position, and that arguments and adjuncts must have a Case-checking position in the clause where they are originally generated to receive the right interpretation. In (46b) when is raised through a Case-checking position (an adjoined position to TNSP, ASPP, or AGRoP) in the clause of bought. Who is raised through the Spec AGRoP in the clause of proved. Suppose that the raising of who to Spec AGRoP precedes the raising of when to Spec CP. Under Shallowness, when must get adjoined to the ASPP in the clause of proved to get as shallow as who, which has been raised to Spec AGRoP. When gets Case-checked there again, violating Greed.

If the upstairs ASP could not Case-check when, the derivation would be permitted without a violation of Greed. There is indeed such a case:

(47) a. ??what j do you wonder [ who i I believe t j to have said [that Mary bought t j ] ]
   b. * when j do you wonder [ who i I believe t j to have said [that Mary bought it t j ] ]

(48) a.* who i do you wonder [ what j I believe t j to have said [that Mary bought t j ] ]
   b. ??who i do you wonder [ when j I believe t j to have said [that Mary bought it t j ] ]

---

35 Later we will discuss these assumptions again.
The derivation for (48b) should proceed in the same way as the derivation for (46b). Nevertheless, (48) is more acceptable.\(^{36}\) This suggests that the ASP that is associated with *believe* is incapable of Case-checking *when* in the first place, so that *when* can be adjoined to the ASPP without violating Greed.

We may note the difference between the aspectual properties inherent in *prove* and *believe*: *prove* is eventive and *believe* is stative. It is not unreasonable to assume that these properties are reflected in morphological properties. ASP is also classified into two kinds: eventive ASP, which feature-checks the aspectual features on eventive verbs like *prove*, and stative ASP, which feature-checks the aspectual features on stative verbs like *believe*. And eventive ASP has a feature to Case-check temporal/locative phrases like *when*, but stative ASP does not.

An argument can be made from an observation by Branigan (1992). Lasnik and Saito (1991) have shown that the subject of an ECM construction can be higher than an adjunct phrase in the matrix in the stage where dependencies like anaphor binding are checked, drawing the conclusion that the embedded subject is raised to the matrix Spec AGRoP for Case-checking at LF:

\[(49)\] The DA proved the defendants to be guilty during each other's trials

However, Branigan has observed that it is true only if the ECM verb is eventive:\(^{37}\)

\(^{36}\) A degrading factor involved here is "Surprising Asymmetry." For this matter, see Pesetsky (1987) and Browning (1987) among others.

\(^{37}\) Branigan actually classifies ECM verbs in terms of "causative" (or "performative").
(50) *The audience believed Bill and Mary to have committed the crime during each other's speeches.

Suppose that in both cases the embedded subject is raised to the matrix Spec AGRoP. The *during*-phrase can in principle get Case-checked by either TNS or ASP. Thus, in (49) the *during*-phrase can get adjoined to ASPP for Case-checking so as to be lower than the embedded subject. In (50), on the other hand, the *during*-phrase must get adjoined to TNSP in order to get Case-checked, so that it cannot be lower than the embedded subject.

We have been assuming that PP complements are Case-checked by ASP. Thus, in (44b) *to whom* must get adjoined to ASPP or AGRoP before it is raised to Spec CP. If the adjunction to ASPP/AGRoP can be applied after the raising of *when* to Spec CP, then *when* crosses over *to whom* in situ through an intermediate adjunction to somewhere inside VP. However, we will argue later that the movement to a feature-checking position obeys the Strict Cycle regardless of whether it is a substitution or an adjunction. Thus, in (44) *to whom* must be raised to a Case-checking position before *when* is raised to Spec CP. In order to satisfy Shallowness *when* must get adjoined to the maximal projection to which *to whom* is adjoined, namely ASPP or AGRoP. Since *seem* is stative, its ASP does not Case-check *when*, so that the derivation does not violate Greed, a desired result.
The argument-adjunct asymmetry concerning the neutralization of the
nesting effect, which is found in examples like (44) and (48), has never been
noticed and does not seem to be easy to explain in the previous frameworks.
We consider it to be a good support for our framework, where the nesting
effect and the unextractability of an adjunct from a *wh*-island are captured on
the same Economy basis.

5.2. Other impossible extractions.

Our approach will have a number of consequences. An interesting one
is that it is expected that there can be a case where an argument is unable to
be extracted from an *wh*-island if it has Case to be checked by TNS.

Belletti (1988) has proposed that the indefinite argument in a *there-
construction has Partitive Case. Suppose that this Case can be checked by
TNS. Then it should be unable to be extracted from an *wh*-island. This
prediction is borne out, as is observed by Frampton (1991, 1992) and Cresti
(1992). The following examples are due to Frampton (1992):³⁸

(43) a. how many ducks do you think there are t in the pond
   b. * how many ducks do you wonder whether there are t in the pond

Shallowness does not allow *how many ducks* in (43b) to reach the AGRsP-
adjointed position in a single step, since it can be Case-checked in that
position by TNS, which also checks *there* in Spec AGRsP. *How many ducks*
must first get as shallow as *there*. Thus it must get adjoined to TNSP or to
*there*. The latter Move-α operation will be necessary for convergence
anyway, if Chomsky's (1989) analysis of the *there*-construction is correct.

³⁸ Compare these examples with the examples in (52).
In any case, *how many ducks* cannot be further adjoined to AGRsP under Greed.

An argument to support the idea that Partitive Case can be checked by TNS will be provided by languages like Icelandic and German.

Jonas and Bobaljik (in preparation) observe that in Icelandic indefinite subjects prefer to follow a sentential adverb whereas definite subjects must precede such an adverb:

(44) a. í gær hefur sennilega einhver lokið verkefninu
    yesterday has probably someone finished the assignment
    "Someone has probably finished the assignment yesterday."
    b. ??í gær hefur einhver sennilega lokið verkefninu

(45) a. *í gær hefa sennilega stúdentarnir lokið verkefninu
    yesterday has probably the students finished the assignment
    "The students have probably finished the assignment yesterday."
    b. í gær hefa stúdentarnir sennilega lokið verkefninu

They also show that it cannot be the case that the (transitive) subject after a sentential adverb is inside VP:

(46) a. í gær lasi sennilega einhver þessar bækur alveg allar
    yesterday read probably someone these books completely all
    "Someone probably read all of these books completely yesterday."
    b. í gær lasi sennilega þessar bækur einhver alveg allar
    c. í gær lasi sennilega þessar bækur alveg einhver allar
The subject must precede a shifted object, which is separated by a VP adverb from a floated quantifier associated with it. Thus they claim that sentential adverbs are adjoined to TNSP and that subjects are in Spec AGRsP or Spec TNSP.

Chris Collins (personal communication) has pointed out that Jonas and Bobaljik's assumption about the two subject positions can be supported in German too. The following examples are due to him:

(47) Es sind 4 Studenten hier.
"There are 4 students here."
  a. Ich bin glücklich, weil zwei Studenten ja doch mein Auto schnell reparieren.
  b. Ich bin glücklich, weil ja doch zwei Studenten mein Auto schnell reparieren.
"I am happy, because indeed two students quickly fix my car."

The subject in (47a) has the presuppositional reading in the sense that it presupposes the existence of a set of entities, i.e. the set of 4 students which has been established in the discourse in this case (cf. Diesing (1992), and also Milsark (1974)). On the other hand the subject in (47b) must have the nonpresuppositional reading, unless a contrastive focus is on zwei.39 Definite subjects must precede a sentential adverb. Note that the subject in (47b) still precedes the shifted object, which in turn precedes a VP adverb. German, like Icelandic, has two subject positions.

39 We will discuss the effect of contrastive focalization later.
Suppose that whereas definite phrases may not have Partitive Case, indefinite phrases can have it optionally.\(^{40}\) Whether an indefinite phrase has Partitive Case or not may result in a difference in interpretation. In German an indefinite phrase is interpreted presuppositionally under a normal intonation if and only if it does not have Partitive Case.\(^{41}\) It does not seem that this effect is observed in Icelandic. According to Jonas and Bobaljik, some notion of heaviness appears to play a role so that \textit{margir stúdentar} 'many students' will be much better than \textit{einhver} 'someone' in the pre-adverbial position in (44b), while it can apparently be non-specific here too. We take light indefinites like \textit{einhver} 'someone' to be more inclined to have Partitive Case.

Jonas and Bobaljik's generalization is now restated as follows: a subject ends up in Spec AGRsP if it does not have Partitive Case, and in Spec TNSP if it has Partitive Case. Our approach provides a way to explain this generalization.

A subject without Partitive Case, which has Nominative Case, is raised to Spec AGRsP to undergo a Nominative Case-checking operation in overt syntax, assuming that a strong feature is involved. Assuming that a sentential adverb has a strong Case to be checked by TNS, it is adjoined to TNSP along the line we discussed above, yielding the desired word order.

\(^{40}\) Here we are departing from Belletti's (1988) assumption that Partitive Case is \(\theta\)-related. See Lasnik (1990) for arguments against this assumption.

\(^{41}\) A clarification about the terminology is in order. A indefinite phrase \textit{three books}, for example, can be paraphrased by a so-called "partitive" phrase \textit{three of the books} under the presuppositional interpretation. Accordingly, the presuppositional interpretation is sometimes dubbed the "partitive" interpretation. We refrain from using the latter term to avoid a possible confusion. For a presuppositional phrase does not have Partitive Case under our assumptions.
The question we have to answer now is: why is it impossible to raise a subject with Partitive Case to Spec AGRsP. Suppose that whenever subjects raise to Spec AGRsP, they go through Spec TNSP. Greed does not allow a subject with Partitive Case to raise to Spec AGRsP, which would induce another Partitive Case-checking operation on the subject. Thus it gets stuck in Spec TNSP. A sentential adverb can be raised over the subject in Spec TNSP to get adjoined to TNSP, since it can go through an intermediate adjunction to the complement of TNS, namely ΣP. This derivation yields the desired word order.

---

42 This is a kind of successive cyclicity, which we will later argue can be reduced to a general consideration based on Shallowness.

43 The adverb could be adjoined to AGRsP instead of TNSP, if nothing is in Spec AGRsP. However, we follow Jonas and Bobaljik in assuming that a phonologically null expletive element occupies Spec AGRsP. There are reasons for making this assumption. For one thing, a phonologically non-null expletive will appear in the initial position of a matrix clause in German (and also of an embedded clause in Icelandic):

(i) a. Es haben ja doch zwei Studenten mein auto schnell gerepariert.
    "Two students indeed fixed my car quickly."
    b. það hefur sennilega einhver lokið verkefninu
    "Someone has probably finished the assignment."

The non-null expletives in the initial position can appear only if the subjects are nonpresuppositional. Thus it is natural to assume that they are raised from Spec AGRsP, and that Spec AGRsP must be always occupied by something regardless of whether it is phonologically non-null or not.

Another reason is more theory-internal. As noted above, as far as the extraction of an adjunct from a wh-island is invariably impossible, the present theory forces us to conclude that something that is feature-checked by TNS must occupy Spec AGRsP.

One thing we will leave open here is why the (transitive) subject must be external to VP even if an expletive occupies Spec AGRsP, which is not a trivial question under the VP-internal subject hypothesis. See Bures (1992) for an explanation, which is adopted by Jonas and Bobaljik.
We assume that Partitive Case can be checked not only by TNS but also by ASP. This makes us expect that objects behave exactly like subject in the relevant respects except that objects, unlike (transitive subject) do not always have to be external to VP in overt syntax. This prediction is borne out in German, as has been pointed out by Diesing (1992), among others. The following examples are due to Chris Collins (personal communication):

(48)   Es sind 4 Autos in der Garage.
       "There are 4 cars in the garage."
   a.   Ich bin glücklich, weil der Hans zwei Autos schnell repariert.
   b.   Ich bin glücklich, weil der Hans schnell zwei Autos repariert.
       "I am happy, because Hans quickly fixes two cars."

The indefinite object precedes a manner adverb in (48a) and follows it in (48b). The object in (48a) can only have the presuppositional reading. On the other hand, the object in (48b) can have both presuppositional and nonpresuppositional readings, although the latter reading is preferred unless a contrastive focus is on zwei. A definite object behaves exactly like a presuppositional indefinite object. It can appear on either side of a manner adverb, with some preference.44

The raising to Spec AGRoP must be through Spec ASPP. An object without Partitive Case can reach Spec AGRoP, but an object with Partitive Case gets stuck in Spec TNSP, since ASP is overtly raised to AGRo just as

44 How strong the preference seems to depend on the speaker (cf. Moltmann (1991)). For the sake of discussion, we simply take this preference to be due to some additional factor irrelevant here.

As for Icelandic, it is not so clear again whether the specificity of an indefinite makes a difference.
TNS is overtly raised to AGRs. Suppose that whatever may turn out to derive the extended portion of the EPP allows Spec AGRoP, unlike Spec AGRsP, to be introduced in covert syntax. Then objects can also stay in situ inside VP in overt syntax. Manner adverbs have Case to be checked by ASP in overt syntax. If an object is in Spec AGRoP, a manner adverb is adjoined to ASPP. If not, it can be adjoined to AGRoP or ASPP. Thus, an object with Partitive Case follows a manner adverb, and an object without Partitive Case precedes it, as is expected.

A piece of evidence for the availability of Spec ASPP for an object can be found by looking at the VP-remnant topicalization. Consider the following example:

(49) gelesen hat Hans das Buch immer
read has Hans the book always
"Hans always read the book."

Thiersch (1985) and Den Besten and Webelhuth (1990) argue that the topicalized constituent in examples like (49) is VP, from which the object is extracted prior to the topicalization. We assume that in (49) the object is in Spec AGRoP and the adverb is adjoined to ASPP. The stranded object would be unable to follow the adverb, as is shown in the following:

(50) * gelesen hat Hans immer das Buch

Since a definite object, which cannot have Partitive Case, cannot stay in Spec ASPP, it must stay inside VP when it follows an ASPP-adjointed

45 The sentence improves with a contrastive focus on the object.
adverb. On the other hand, an indefinite object with Partitive Case can stay in Spec ASPP when it is forced to be externalized from VP. We should expect that an indefinite object can be stranded and follow an adverb. This expectation is borne out:

(51) gelesen hat Hans immer Bücher

read has Hans always books

"Hans always read books."

An indefinite object can be overtly raised to Spec ASPP just as an definite object can be shifted, suggesting that some feature involved can be optionally strong or weak.

It does not seem that English reveals the definite-indefinite or presuppositional-nonpresuppositional distinction of the sort which we just saw concerning Icelandic and German. However, this does not necessarily mean that in English Partitive Case is restricted to the context of there construction. Rather, we assume that Partitive Case is freely distributed in principle in this language too. And we assume, basically following Jonas and Bobaljik, that in languages like English TNS and ASP cannot have Spec to begin with, so that an indefinite phrase with Partitive Case can be raised to Spec AGRsP.

Since indefinites can optionally have Partitive Case, it is not apparently easy to find a signal of whether they have Partitive Case. However, we expect that the presence/absence of Partitive Case may be reflected in an interpretive difference, which we just saw was the case in German.
The referential-nonreferential distinction discussed by Cinque (1990)\textsuperscript{46} can be recast in terms of Partitive Case. His discussion is based on Longobardi's observation concerning scope interaction. Consider the following:

(52) a.  [how many books]j do you think that everyone read tj
b.  [how many books]j do you wonder whether everyone read tj

*How many books* can be interpreted as taking narrow scope with respect to *everyone* in (52a) but not in (52b). Let us assume for the moment with Cinque that a *wh*-phrase can take narrow scope with respect to another quantifier phrase only if it is interpreted nonreferentially.\textsuperscript{47} Under this assumption, the contrast found in (52) will straightforwardly follow if nonreferential *wh*-phrases cannot be extracted from a *wh*-island in the first place, as Cinque argues.

Cinque proposes that the notion of "referentiality" is the ability to refer to specific members of a set in the mind of the speaker or preestablished in the discourse and that it subsumes Pesetsky's (1987) notion of D-linking.\textsuperscript{48} We identify Cinque's referential-nonreferential distinction with our presuppositional-nonpresuppositional distinction. If the *wh*-phrase in (52) is used nonreferentially, then it will have to obtain Partitive Case and will

\textsuperscript{46} For discussions of the matter under different assumptions, see Longobardi (1987), Rizzi (1990), Lasnik and Saito (1990), Kroch (1989), Frampton (1991, 1992), Szabolcsi and Zwarts (1993), Cresti (1992) and Murasugi and Saito (1992), among others.

\textsuperscript{47} We will see below that this assumption is not precisely correct, though.

\textsuperscript{48} Enç (1991) claims that the notion of D-linking is the same as her notion of specificity: specificity involves a ("weak") link to previously established discourse referents.
therefore be unable to be extracted from a wh-island, yielding the contrast between (52a) and (52b) with respect to scope interaction.

Rizzi (1990) argues that the referential-nonreferential distinction should be defined so as to treat quasi-argumental phrases, including measure phrases, idiom chunks, etc., as nonreferential. Consider the following:

(53) a. [how much]i do you think that the computer weighed/cost ti
   b. * [how much]i do you wonder whether the computer weighed/cost ti

(54) a. [how much advantage]i do you think that John took ti of Mary
   b. * [how much advantage]i do you wonder whether John took ti of Mary

The wh-phrases in (53b) and (54b) are quasi-argumental and resist the extraction from a wh-island. Under our system this fact simply suggests that quasi-arguments have a feature which can be checked by TNS, which we identify with Partitive Case.

A serious question is why nonreferential phrases must have Partitive Case. Since it is natural that morphological properties are reflected in semantic properties and vice versa, it is not surprising that the absence/presence of Partitive Case is correlated with the referential-nonreferential distinction. But how should the correlation be as it is, not in the opposite way? Below we will consider a possible direction to answer this question.

Suppose first that noun phrases have the following clause-like structure:

(55) ([DP ... D] [IP ... I [NP .... N ... ]]](])
D and I are functional elements heading their own projections. D is comparable with C. Some noun phrases lack D, just as some clauses lack C. I is a complex of features including those which check inflectional or declensional features on N, and may be actually split to distinct heads, as is the case in clauses. N is raised overtly or covertly and adjoined to I to undergo a checking operation for its inflectional features. Arguments and adjuncts generated in NP are also Case-checked by I. I itself has features to be checked. Some features will be checked by D, when I is raised to D.

D, like C, plays a crucial role in determining the meaning of the entire noun phrase. Suppose that D is always presuppositional in the sense of Milsark (1974), i.e., it presupposes a set of referents, and that only D can convey such a presupposition. Then the capability of being referential only emerges from D. Thus an indefinite is referential if and only if it includes D. Suppose further that Partitive Case is a morphological feature (or feature complex) that is necessary associated to I.

TNS, ASP and D constitute a class of potential checkers of Partitive Case. If a noun phrase includes D, then Partitive Case on I is checked when I is raised to D. If D is missing, then Partitive Case is checked when the noun phrase (=IP) is raised to enter the checking domain of TNS or ASP, since Partitive Case is automatically percolated up to IP.

Put differently, Partitive Case is embedded in every noun phrase. But whether it surfaces depends on the structure of the noun phrase. If D is

49 We abstract away from the morpho-phonological realization of D. Also we do not assume that determiners such as the, a, every, some, etc. necessarily belong to D.
50 I-to-D raising in a noun phrase can be considered to be analogous to I-to-C raising in a clause, which has been proposed in the literature.
51 Ken Hale (personal communication) has informed me that languages like Miskitu have the form [[[N]indef]def] to refer to referential indefinites.
included, Partitive Case does not surface, since it is prevented from percolating up to DP. Consequently, only nonreferential phrases have Partitive Case on the surface.\textsuperscript{52}

Now remember that in the discussion of the German examples above we noted that a presuppositional (or referential) indefinite and a definite can behave like a nonpresuppositional (or nonreferential) one if it is assigned a contrastive focus. We assume that in this case the noun phrases include an additional feature, which we may or may not call Case. This feature affects both phonological and semantic interpretations to derive the characteristics of contrastive focalization. If the feature in question must be checked by TNS or ASP, it cannot be checked inside the noun phrase and must be percolated up to the surface. Now the focused phrase behave exactly like nonreferential indefinite phrases with respect to the checking by TNS or ASP.

We expect that a similar phenomenon can be found in English too. Consider the following:

(56) a. [which/how many of those books]$_i$ do you think that everyone read $t_i$

b. [which/how many of those books]$_i$ do you wonder whether everyone read $t_i$

\textsuperscript{52} We may be able to assume equivalently for the purpose of the present discussion that $D$ can be just defective in the relevant respects rather than entirely missing.
The wh-phrases in (56) are clearly referential. But here we find the same contrast as we found in (52). The wh-phrases can take the narrow scope with respect to everyone in (56a) but not in (56b).

We assume that referential phrases, unlike nonreferential phrases, are not necessarily made quantificational. Thus the wh-phrases in (56) must acquire a quantificational force to scopally interact with everyone. Under our approach it is natural to assume further that those phrases include an additional feature, which is associated with quantificational force. If this feature must be checked by TNS (and ASP), no phrase including it can be extracted from a wh-island.

We have so far proposed various features to be checked by TNS. We do not need to further pursue their identities. The point we would like to make here is that if a phrase bears a feature which can be checked by TNS, then it cannot be extracted from a wh-island, regardless of the exact nature of that feature.

Furthermore, it does not have to be the case that TNS is the only checking feature which blocks the extraction from a wh-island. In principle, any feature or feature complex will have the potential of blocking a movement over the Spec of a head $\alpha$, if its checking domain includes the adjoined position to the complement of $\alpha$.

With AGRsP being the complement of C, the AGRs-TNS complex can block the subject extraction from a wh-island, yielding a wh- $t$ effect. Consider the following:

---

53 It is unclear whether we can identify this additional feature with that which is associated with a contrastive focus in the case of German.
(57) a. * which student do you wonder [ whether [ t_i read the book ] ]
b. which book do you wonder [ whether [ John read t_i ] ]

Here the contrast is sharp: (57a) is totally unacceptable while (57b) is marginal at worst, as has been pointed out in the literature.\(^5\)

The only difference between the subject and object wh-phrases is that the former, unlike the latter, has Nominative Case which can be checked by an AGRs-TNS complex. If the subject is necessarily raised to Spec AGRsP, where its Nominative Case is checked, an adjunction to AGRsP from here in order to get extracted from the wh-island induces a redundant Case-checking. It is this redundant Case-checking that is regarded as motivating the AGRsP-adjunction, violating Greed. On the other hand, the AGRsP-adjunction of the object does not induce Case-checking, so that it does not violate Greed since it will be motivated by the wh-checking applied in the matrix Spec CP.

The assumption that the subject must get raised to Spec AGRP can be derived by restating the idea that structural Case-checking is essentially a form of Spec-head agreement. We take “agreement” to be just a label for reciprocal checking. A phrase is Case-checked by an AGR-TNS/ASP complex if and only if the former has checked the latter. A head can be

\(^{54}\) We do not assume that (57a) is deviant for the totally same reason as that-\(t\) examples such as the following:

(i) * which student do you think [ that [ t_i read the book ] ]

For, as has been pointed out by Sobin (1987), some speakers from the midwest accept that-\(t\) but not wh-\(t\). One of my informants is such a speaker.

We do not assume either that (57) is derived by raising whether over the subject wh-phrase to yield a crossing. For we have been assuming that whether can be introduced in Spec CP by a binary transformation.
checked by a maximal projection only in an X-bar configuration, that is, only if the latter is a complement or Spec of the former. Thus a Nominative/Accusative phrase must get raised to Spec AGRs/o so that it can check AGRs/o, which in turn can check the phrase.

This line of considerations leads us to an interesting prediction. A subject should be able to get extracted from a wh-island if it is raised from a further embedded clause. This prediction is borne out, as has been noted in the literature (see Lasnik and Saito (1984, 1992), among others). The following examples are from Browning (1987: 296):

(58) a. ?? which studenti did John wonder [ whether [ to believe [ t_i understood the problem ]]]
   
   b. ? which problemi did John wonder [ whether [ to believe Bill understood t_i ]]]

(58a) is clearly more acceptable than (57a), though it is still worse than (58b). To derive (58a), which student must get adjoined to the AGRsP of

---

55 The contrast between (58a) and (58b) has been dubbed "Surprising Asymmetry" since Pesetsky (1987) (See also Browning (1987)). It is one of a number of yet unexplained subject-object asymmetries with regard to wh-movement. I will leave this matter open here.

Notice also that (58b) is worse than (57b). This is probably due to an irrelevant factor that has been noted in the literature (see Rizzi (1982), Chomsky (1986) and Frampton (1990), among others), namely, a further embedding. The following data are from Frampton (1990: 69-70):

(i) a. Sam, who I know when to try to see t,...
   b. ??Sam, who I know when to say I saw t,...
   c. ? Sam, who I know why you tried to see t,...
   d. * Sam, who I know why you said you saw t,...
the *whether*-clause. However, no position adjoined to the AGRsP in question is a Case-checking position for *which student*, since it is not raised to the Spec of the AGRsP to feature-check the AGRs-TNS complex.

From the above stated property of structural Case-checking it follows that a subject extraction will improve even when the *wh*-island is a tensed clause, which also has been pointed out in the literature. The following examples are from Rizzi (1992: 95):

(59) 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 ]]]

Again we find the “surprising” asymmetry and the Tensed-S effect here. But the subject extraction is better than the adjunct extraction, which shows the difference between structural Case and nonstructural Case, the latter being able to be checked by any instance of TNS.

Interestingly, in some languages other than English the object behaves exactly like the subject. Browning (1987: 312) cites the following Dutch examples from Koopman and Sportiche (1986)\(^5\)

(60) a. *wie hij zich afvroeg of jij t aardig vond
   who he wondered whether you liked
   “who did he wonder whether you liked”

---

Embedding a tensed clause in a *wh*-island reduces the extractability, though we do not take the degraded examples in (i) to be as bad as indicated. It worth noting that in spite of this effect (58a) is more acceptable than (57a).\(^5\) Their work was orally presented, and I do not have a copy of the handout.
b. * wiei hij zich afvroeg of ti jouw aardig vond
   who you wondered whether you liked
   "who do you wonder whether liked you"

c. * waaromij hij zich afvroeg of Jan ti ontslagen was
   why he wondered whether Jan fired was
   "why did he wonder whether Jan had been fired"

d. met wiei hij zich afvroeg of hij ti zou kunnen praten
   with whom he wondered whether he would be able to talk
   "with whom did he wonder whether he would be able to talk"

Comorovski (1990) shows a similar fact in French: 57

(61) a. [Subject]
   * Jean, quiij j’ignore oùj ti se trouve maintenant tj, nous avait promis
     de nous rendre visite ce soir
     “John, who j I do not know wherej ti is now tj, had promised us to
     pay us a visit tonight”

b. [Object]
   * il fredonnait des mélodies quei je ne savais pas quij tj avait
     composées ti
     “he was humming melodies whichj I did not know whoj tj had
     composed ti”

57 (61) also involves a superiority violation.
c. [Dative object]
Jean, à qui je ne sais pas quel colis ils ont envoyé, s’attend à un cadeau imposant
“John, to whom I do not know which parcel they have sent, expects a large gift”

d. [PP complement]
cet appartement, dans lequel j’ignore combien d’étudiants ont habité, me semble tout à fait délabré
“this apartment, in which I do not know how many students have lived, seems to me to be very run down”

e. [PP adjunct]
le maintien du niveau bas des salaires - voilà une raison pour laquelle je sais qui démissionnerait
“the continued low salaries - this is a reason for which I know who would resign”

The impossibility of the object extraction implies two things: the object is raised through Spec AGRo, and AGRo-ASP complex is raised to AGRs-TNS complex so as to include the AGRsP-adjoined position in its checking domain.

If the object wh-phrase were raised without going through Spec AGRo so that it failed to check AGRo-ASP complex, then the AGRsP-adjoined position would not function as a checking position for Accusative Case, even if it is structurally included in the checking domain of AGRo-ASP complex. Accusative Case-checking would be successfully made possible by raising the original trace of the object in covert syntax.
In the next section we will argue that Shallowness derives a version of successive cyclicity: a Spec position must be filled in overt syntax if possible. The object can be raised through Spec AGRo, since there should be no feature conflict in the first place and Procrastinate is satisfied in this case. Thus it must be raised through there.

As for the second implication, it is justified by the correlation of the object extractability and V-raising, which is observed by Comorovski (1990):

(62)  cet événement, que je ne sais comment

a.  complètement oublier t₁ t₂
b.  oublier complètement tᵥ t₁ t₂

m’a changé la vie

“this event, which I don’t know how to forget completely, has changed my life”

(63)  c’est le courage de répondre que je ne vois pas comment

a.  ? ne pas avoir t₁
b.  *n’avoir pas tᵥ t₁

dans cette situation t₂

“it is the courage to answer that I do not see how not to have in this situation”

Basically following Pollock (1989), suppose that a tensed verb must be raised to AGRs-TNS and that an infinitival verb is optionally raised to AGRo-ASP if it is a thematic one and optionally raised to AGRs-TNS if it is a non-thematic one. Suppose further that V can be raised to AGRs-TNS only by first raising to AGRo-ASP and then pied-piping it. Thus the data in
(61)-(63) shows that the object extraction is impossible if and only if AGRo-ASP is raised to AGRs-TNS, supporting our analysis.\(^{58}\)

Comorovski (1990) reports the observation, which she attributes to Hirschbühl and Valois (1989),\(^{59}\) that in French objects will become extractable from a \textit{wh}-clause if it is further embedded in a declarative clause inside the \textit{wh}-island, just as subjects in English. This is certainly what should be expected under our analysis.

A question is how Dative objects and complement and adjunct PPs can be extracted if AGRo-ASP is raised to AGRs-TNS. Dative Case and other lexical Cases for PPs are checked exclusively by ASP. Thus they will fail to get extracted through AGRsP-adjunction, if (i) they need to get raised to a Case-checking position by ASP before getting adjoined to AGRsP or (ii) the AGRsP-adjointed position enters the checking domain of ASP. Neither condition is met.

As for the first condition, remember that the reason that the object must be raised through Spec AGRoP is given by the assumption that a Spec position must be filled in overt syntax if possible. If no Spec is available for a phrase with Dative or some other lexical Case, then there will be no reason that such a phrase needs to be raised through a Case-checking position.

As for the second condition, let us follow Chomsky’s (1992) assumption that the domain of a head is determined derivationally. Suppose that the raising of AGRo-ASP is done in the following way, ignoring possibly intervening heads:

\(^{58}\) German does not apparently show the correlation because of head-finalness. Also it may be possible in principle that AGRo-ASP raises to AGRs-TNS by itself, and that V is raised to AGRs-TNS by skipping or excorporating from AGRo-ASP. In such cases the wanted correlation will never be found.

\(^{59}\) This paper has not been available to me.
ASP is adjoined to AGRo to form AGRo-ASP complex so that its checking domain is extended. When [[[AGRo]ASP]V] is adjoined to [[AGRs]TNS], ASP is an adjunct to AGRo and therefore is not regarded as being "moved," so that the checking domain is not redefined. On the other hand, AGRo-ASP as well as AGRo and AGRo-ASP-V is considered to be moved, and its checking domain is extended. Thus the AGRsP-adjoined position does not enter the checking domain of ASP, though it enters the checking domains of AGRo-ASP, AGRs-TNS and TNS. Therefore, only elements with Accusative Case, Nominative Case and features checked by TNS are not extractable from a wh-island.

This line of considerations leads us to a possible explanation for the that-t effect, which is exemplified in the following:

(65) a. who do you think [ (*that) [ t_i hates Mary ]]
   b. who do you think [ (that) [ John hates t_i ]]
   c. why do you think [ (that) [ John hates Mary t_i ]]

---

60 As noted above, there are speakers who accept examples like (65a) with that. See Sobin (1987) for this matter. There is also a variation concerning for-t, which we believe can be accounted for in the same way as that-t.
What is interesting here is that not only the object but also the adjunct can be extracted from a *that*-clause. Thus, the question to answer is: what singles out the subject?

It has recently been suggested in the literature that I or AGRs is raised to C. To make this idea precise, suppose that TNS is generally feature-checked by C, deriving the co-occurrence restriction between TNS and C. If the relevant feature of C is strong, TNS must be raised to C for feature-checking in overt syntax. If the feature in question is weak, TNS must be covertly raised, given Procrastinate.

Suppose that *that*, unlike empty complementizers, is strong with respect to feature-checking of TNS. Then, TNS must be overtly raised and adjoined to C in overt syntax. However, TNS cannot raise to C by itself, because it must be first adjoined to AGRs to form a complex. Thus it is the AGRs-TNS complex that is raised to C.61

After raising of AGRs-TNS, Spec CP is included in the checking domain of AGRs-TNS, but not of TNS for the same reason as discussed above. Suppose further that a *wh*-phrase must be raised through the Spec of *that*, a successive cyclicity effect. Under these assumptions the subject *wh*-phrase in (65a) is doubly Case-checked in Spec CP and in Spec AGRsP.

---

61 A question is, what position is occupied by an auxiliary or modal verb is in a *that*-clause? Consider the following, for example:

(i) I think that John can solve the problem

A possibility is that *can* is adjoined to AGRs-TNS, and that AGRs-TNS is "excorporated" in the sense of Roberts (1991) so that it raises to C leaving *can* behind, though we are assuming that TNS cannot be excorprated from AGRs-TNS. Another possibility is that *can* is substituted for or adjoined to a head that is lower than AGRs or TNS.
violating Greed. On the other hand, the *wh*-phrases in (65b, c) are not Case-checked in Spec CP, so that they are successfully extracted.

It has been noted that the effect in question is clause-bounded:

(66) who do you think [ (that) [ John said [ (*that) [ t; hates Mary ]]]]

This is one of the predictions we make along the same line as in the case of the *wh-t* effect discussed above.

We will later argue that the successive cyclicity of the kind in question is derived from Shallowness. A consequence of this approach is that it is expected that there should be no such effect in covert syntax. That is, a *wh*-phrase does not go through Spec CP when it is extracted from a *that*-clause. Thus we should expect that there is no effect such as *that-t*. Though there seems a variation among speakers with regard to whether a subject *wh*-phrase in situ is allowed, as has been noted in the literature, Howard Lasnik (personal communication) has noted that few speakers are sensitive to the presence/absence of *that* with respect to the acceptability of the subject *wh*-phrase in situ. We will later discuss the *wh*-in-situ in general.

Culicover (1991) has pointed out that the *that-t* effect is suppressed by intervention of certain sentence-level adjuncts, such as *for all intents and purposes, just yesterday, in NP’s opinion, and under normal circumstances.* The following is one of his examples:

(67) Robin met the man [who [Leslie said [that [*for all intents and purposes] t; was the mayor of the city ]]]]
A question here is how the adjunct can appear in front of the subject in the first place, assuming that it is not topicalized. Remember that we have argued that an adjunct which can be Case-checked by TNS cannot be raised over the subject and adjoined to AGRsP without violating Greed.

Suppose that the adjunct in question is generated in the Spec of a functional head. Call this head $\Psi$ for convenience' sake. $\Psi$, which functions as inserting a parenthetical phrase, can rather freely appear in a clause structure. In (67) $\Psi$ is generated between C and AGRs. That is, $\Psi$ takes AGRsP as its complement, and $\Psi P$ is taken by C as its complement. Suppose further that $\Psi$, just like C, feature-checks TNS, and that the relevant feature of $\Psi$ is optionally strong. In the case in which the feature in question is strong the overt raising of TNS to C is accomplished under Procrastinate and Greed by first adjoining AGRs-TNS to $\Psi$ and then adjoining the whole complex to C. As a result the feature-checking domain of AGRs-TNS is not extended to include Spec CP. The subject wh-phrase is not Case-checked in Spec CP, so that Greed is not violated. Thus the that-t effect disappears.

Noam Chomsky (personal communication) has pointed out that the wh-t effect is also suppressed by an intervening parenthetical adjunct:

(68) who did you ask [whether [* (under normal circumstances) t would win the election]]

This fact simply suggest that $\Psi$ can optionally carry a strong or weak feature to feature-check TNS. When $\Psi$ is weak, AGRs-TNS stays in situ. And the subject wh-phrase can be extracted from the wh-island through an adjunction to $\Psi P$ without violating Greed, since the $\Psi P$-adjoined position is external to the feature-checking domain of AGRs-TNS.
Interestingly, an adjunct is still unextractable from a *wh*-island even if a parenthetical adjunct is inserted:

\[(69) \quad * \text{why}_i \text{ did you ask [whether [(under normal circumstances) John would win the election \(t_i\)]]}\]

If \(\Psi\) is capable of feature-checking of sentence-level adjuncts in general, adjuncts like *why* cannot be extracted from a *wh*-island through an adjunction to \(\Psi\). For the \(\Psi\)-adjoined position will be a Case-checking position for such an adjunct. For *why* to get adjoined to \(\Psi\), it must be first adjoined to AGRsP since the parenthetical phrase, which is feature-checked by \(\Psi\), is interfering with regard to Shallowness. The AGRsP-adjoined position is also a Case-checking position for *why*. Thus there is no way for *why* to get extracted from the *wh*-island without violating Greed.

Under the present system, an explanation will be provided along the same line for Lasnik and Saito’s (1992) observation that the subject fails to reveal the capability of undergoing a (clause-bounded) topicalization.

They present two diagnoses for topicalization. The first is concerning the extraction domain. Consider the following examples they raise:

\[(70)\]
\[
a. \quad ?* \text{who}_i \text{ do you think that [pictures of \(t_i\) ] are on sale}
   b. \quad ?? \text{who}_i \text{ do you think that [pictures of \(t_i\) ] John wanted}
\]
\[
(71) \]
\[
a. \quad ?? \text{who}_i \text{ do you wonder [which pictures of \(t_i\) ] Mary bought}
   b. \quad ?? \text{who}_i \text{ do you wonder [which pictures of \(t_i\) ] is on sale}
   c. \quad ?? \text{who}_i \text{ do you think that [pictures of \(t_i\) ] Mary believes are on sale}
\]
If (70a) can be yielded by a topicalization of the subject, then there should be no reason that it must be worse than (70b) whereas (71a) is as good as (71b). The relative acceptability of (71c), which clearly involves a (long-distance) topicalization of subject, also forces us to conclude that a topicalization cannot be involved in (70a).

The second diagnosis is concerning anaphor binding. Consider the following examples from them:

(72) a. * John thinks that Mary likes himself
   b. John thinks that himself Mary likes
   c. * John thinks that himself likes Mary

(73) a. John thinks that himself Mary thinks Susan likes
   b. ?? John thinks that himself Mary said won the race

The contrast between (72a) and (72b) shows that a topicalization of the anaphor object extends its binding domain. Given this, it should be expected that (72c) will be improved with a topicalization of the subject if it can be analyzed as involving a topicalization of the subject. The acceptability of (73b) suggests that in the case of a long distance topicalization of the subject does extend the binding domain, as is expected, although the degradation of (73b) in comparison with (73a) shows that there should be an additional factor. Whatever the additional factor is, (72c) should be expected to be as good as (73b) as far as a topicalization of the subject is possible.

These facts undoubtedly supports Lasnik and Saito’s claim that the subject cannot undergo a clause-bounded topicalization.

Suppose now that a topicalized phrase is in the Spec position of a functional head. Call this head TOP. TOP is generated between C and
AGRs. TOP feature-checks its Spec for a wh-feature. Suppose also that TOP has a strong feature that checks some feature that is additionally carried by TNS. Then AGRs-TNS is overly raised and adjoined to TOP, so that Spec TOPP is included in the checking domain of AGRs-TNS. If the subject is raised to Spec TOPP from Spec AGRsP in the same clause, it will be doubly Case-checked and Greed will be violated.

A prediction is that if Ψ is generated between TOP and AGRs, then the subject should become able to be topicalized. Consider the following:

(74) a. who do you think that [pictures of t₁] * (?under normal circumstances) would be on sale
b. John thinks that himself * (?under normal circumstances) would win the race

Though the fact is not so clear, it does not seem to be unreasonable to think that our prediction is borne out.

The most significant advantage of our system in dealing with extraction phenomena is that relevant variations across languages and within a language are provided with an explanation based on the morphological properties of the elements involved. The essential idea of the recent “principle-and-parameters” approach to language is that UG gives a computational system consisting of invariant principles and a finite set of finitely-valued parameters. The parameters belong to the Lexicon, and furthermore they are only associated with morphological properties of lexical items. Variations should be derived from an interaction of the
computational system with the morphological properties of employed items. Our analysis of extraction phenomena readily fits in this approach.

So far we have assumed that derivations proceed in a cyclic manner. This is crucial for almost all discussions above. In this section we will argue that Shallowness will derive a version of Strict Cycle.\textsuperscript{62}

Our version is extensionally comparable to Chomsky's (1992) "extension" version of Strict Cycle: transformations must extend the phrase marker.\textsuperscript{63} This requirement is stated as follows:

\[(75)\] Extension Condition

A phrase-marker $K$ must be properly included by $K'$, where $K$ is transformed to $K'$.

Here $K$ is properly included by $K'$ if and only if $K$ is identical to a proper subphrase-marker of $K'$, whose root category is irreflexively dominated by that of $K'$.

Consider the following schematic structure, for example:

\[(76)\] $[YP \beta [Y' Y [XP \alpha [X' X ZP ]]]]]$

Suppose that we are dealing with substitution of four elements, $X$, $Y$, $\alpha$, $\beta$. The question is, in what order are the four substitutions applied? The X-bar

\textsuperscript{62} Strict Cycle is originally defined by Chomsky (1973) as follows:

\[(i)\] No rule can apply to a domain dominated by a cyclic node $A$ in such a way as to affect solely a proper subdomain of $A$ dominated by node $B$ which is also a cyclic node.

\textsuperscript{63} Chomsky, unlike us, takes Form-Chain and GT rather than (singular and binary) Move-$\alpha$ to be characterized in terms of the Strict Cycle.
theory reduces possibilities. Thus, the substitution of Y must follow that of X since XP is Y's complement. The substitution of α must follow that of X, since α is the Spec of XP. Similarly the substitution of β must follow that of Y.

Three possibilities still remain:

(77) \[
\left[ \begin{array}{c}
Y' \ Y' \ X' \ X' \ ZP
\end{array} \right]
\]

(78) \[
\left[ \begin{array}{c}
Y' \ Y' \ X' \ X' \ ZP
\end{array} \right]
\]

(79) \[
\left[ \begin{array}{c}
Y' \ Y' \ X' \ X' \ ZP
\end{array} \right]
\]

The Extension Condition rules out (78) and (79), where the substitution of α fails to extend the phrase marker.

Two provisos are necessary for the Extension Condition. First, adjunctions must be exempted from this condition. (An adjunction can apply noncyclically.) Second, the condition must be inapplicable in covert syntax. (Covert syntax is noncyclic.)

The first proviso is necessary, because an adjunction never extends the phrase marker, so that it would necessarily violate the condition. Take the following operation, for example:
The adjunction of $Y''$ in (80) does not extend the input phrase-marker in such a way that it is properly included in the output phrase marker.

The notion of "extension" could be defined in such a way that $K$ is extended to $K'$ if and only if the root category of $K'$ contains that of $K$. The adjunction of a maximal projection to another would thus "extend" the phrase marker.

However, head-to-head adjunctions would not be saved in this way. Consider the following, for example:

Here $Y^0$ is adjoined, and the input phrase marker is not extended with either definition of "extension." Thus, the Extension Condition may not apply to adjunctions.\(^{64}\)

A similar reasoning leads us to the second proviso: the Extension Condition does not hold in covert syntax. In covert syntax the phrase-

---

\(^{64}\) Technically, the Extension Condition is easily defined as a constraint on substitutions. Thus, Chomsky (1992) suggests to take it as a requirement on $\Delta$-insertion, which is not employed by an adjunction.
marker is never extended, since the Lexicon is inaccessible. Consider sentences like John loves Mary, for example. The object Mary must be raised to Spec AGRoP in covert syntax, without extending the phrase marker. This kind of considerations will immediately require that the Extension condition should not apply in covert syntax.

Under our present system, the Strict Cycle follows directly from Shallowness, together with the just mentioned two properties (exemption of adjunction and noncyclicness of covert syntax).

Let us see how derivation (77) is ruled in and derivations (78) and (79) are ruled out.

(77) \[ [YP \beta [Y' Y [XP \alpha [X' X ZP ]]]] \]

(78) \[ [YP \beta [Y' Y [XP \alpha [X' X ZP ]]]] \]

(79) \[ [YP \beta [Y' Y [XP \alpha [X' X ZP ]]]] \]

65 Possibly, except in limited cases. For example, we might be able to think that in wh-in-situ languages like Japanese a sentence with a matrix wh-question involves a covert wh-raising to extend C' to C''.

66 Chris Collins (personal communication), however, has suggested the possibility that given the first proviso, transformations in covert syntax are all adjunctions.

Incidentally, Branigan (1992) claims under a Form-Chain approach that a Form-Chain operation with an apparent Strict Cycle violation actually ends up with an adjunction. Thus, in the case of overt object shift in languages like Icelandic the object is raised through Spec AGRoP and adjoined to some position. We might be able to think that the covert raising of an object proceeds in the same manner.
The substitution of $\alpha$ in (77) satisfies the Strict Cycle, but those in (78) and (79) violate it. However, they are not alternative to each other with regard to Shallowness, given the conditions in (15).

(4) Shallowness

An operation must be the shallowest.

(15) Two Move-$\alpha$ operations are alternative to each other if and only if

a. they have the same input structure,

b. they are movements to the same position, and

c. they are motivated by the feature-checking operations applied in the same position.

For their input phrase markers are different:

(77') $[X' X \ ZP ]$

(78') $[YP \beta \ [Y' Y [XP \ [X' X \ ZP ]]]]$ 

(79') $[Y' Y [XP \ [X' X \ ZP ]]]$

The operation that should be considered here is the substitution of $X$, which has the same input structure in all cases. A question is how (15c) is satisfied. Note that the substitution of $X$ were not motivated by a feature-checking operation in the first place, it would be ruled out by Greed. We assume that a head is feature-checked by its complement (and its Spec), expanding the concept of feature-checking.
One might think that Greed does not have to apply to binary Move-α operations (GT), and therefore the substitution of a head does not have to be motivated by a feature-checking operation. Under our approach there is no reason to distinguish between singular and binary Move-α operations. Furthermore, as far as we adopt the Larsonian VP structure, it is anyhow necessary to assume that a head is feature-checked by its complement, since in a Larsonian recursion structure necessarily involves a singular Move-α operation that substitute a head.67

If the substitution of X in (77) is shallower than those in (78) and (79), then the latter derivations are ruled out. Noam Chomsky (1992 class lectures) has proposed that in the process of substituting a head Δ is introduced and projected up in accordance with the X-bar theory, and it is arbitrary in principle whether it is projected up to the single-bar level or up to the double-bar level. Thus we have the following two possibilities:

\[(\text{82})\]

\[
\begin{array}{c}
\begin{tikzpicture}
  \node (Z) at (0,0) {$Z''$};
  \node (Delta) at (1,1) {$\Delta$};
  \node (Z) at (2,0) {$Z''$};
  \draw (Delta) -- (Z);
  \node (X) at (4,1) {$X'$};
  \node (Z) at (5,0) {$Z''$};
  \draw (X) -- (Z);
\end{tikzpicture}
\end{array}
\]

\[(\text{83})\]

\[
\begin{array}{c}
\begin{tikzpicture}
  \node (Z) at (0,0) {$Z''$};
  \node (Delta) at (1,1) {$\Delta$};
  \node (Z) at (2,0) {$Z''$};
  \draw (Delta) -- (Z);
  \node (X) at (4,1) {$X''$};
  \node (Z) at (5,0) {$Z''$};
  \draw (X) -- (Z);
\end{tikzpicture}
\end{array}
\]

67 Given the assumption that binary Move-α operations are subject to Greed and that heads can be feature-checked by their complements but cannot feature-check them, head-complement structures must always be constructed by an operation that affects the head and targets the phrase to be complemented, since heads do not feature-check their complements.
The operation shown in (82) has the depth $\emptyset$, and the operation shown in (83) has the depth $\{X''\}$.

In derivations (78) and (79) the second option for the substitution of $X$ cannot be taken. If the first option is taken so that $X$ is only projected up to the single-bar level, the substitution of $Y$ will fail since the $X$-bar theory requires that a complement must be a maximal projection. In derivation (77), on the other hand, the first option can be taken, since the substitution of $\alpha$ will project $X'$ to $X''$, which is ready for the substitution of $Y$. Now noncyclic derivations (78) and (79) are blocked by cyclic derivation (77) under Shallowness.

Let us now turn to the above mentioned two properties of the Strict Cycle: it does not apply to adjunctions, and it does not hold in covert syntax.

A head can refrain from projecting up to its maximal projection when it is substituted, only if its Spec can be immediately supplied. That is, if an element is raised to the Spec of a head in the course of derivation, it has to be raised there as soon as possible in order to make the substitution of the head the shallowest. Raising to Spec is a substitution. And no adjunction helps the substitution of a head with this regard. As a result, adjunctions, unlike substitutions, can apply noncyclically.

Suppose that no maximal projection is newly created inside the phrase marker without access to the Lexicon. The phrase marker has been fully extended in covert syntax. Thus any covert operation will have the same depth regardless of the application order with respect to others. Covert operations can apply in any order as far as Shallowness is concerned. That
is, covert syntax is noncyclic. Now we have derived the two properties in question without stipulation.

Besides deriving a version of the Strict Cycle, Chomsky's Extension Condition has one more significant consequence. It opens a way to derive the binary branching requirement on X-bar structures. This requirement gives rise to a conceptually desirable result that makes the X-bar theory more restrictive, and seems to be empirically justified.

Consider the following substitution operations:

(84)  
\[
\begin{array}{c}
X' \\
X \\
Y''
\end{array} \hspace{1cm} \rightarrow \hspace{1cm} \begin{array}{c}
X' \\
X \\
Y'' \\
Z''
\end{array}
\]

(85)  
\[
\begin{array}{c}
X'' \\
Y'' \\
X'
\end{array} \hspace{1cm} \rightarrow \hspace{1cm} \begin{array}{c}
X'' \\
Z'' \\
Y'' \\
X'
\end{array}
\]

Z'' is affected and inserted in a second complement position in (84), and in a second Spec position in (85), yielding a ternary branching structure in each case. No operation like these is allowed by the Extension Condition.

Shallowness does not help us rule out multiple branching structures.\(^{68}\) But it is undesirable to adopt the Extension Condition in the present system.

---

\(^{68}\) In (84), unlike in (85), Greed is violated, given the assumption that heads do not include their complements in their checking domains. However, consider the following:

(i)  
\[
\begin{array}{c}
X' \\
X \\
Y''
\end{array} \hspace{1cm} \rightarrow \hspace{1cm} \begin{array}{c}
X' \\
X \\
Y'' \\
Z''
\end{array}
\]

\[
\begin{array}{c}
Y \\
X
\end{array} \hspace{1cm} \cdots tY \cdots
\]

\[
\begin{array}{c}
Y \\
X
\end{array} \hspace{1cm} \cdots tY \cdots
\]
for the purpose of deriving the binary branching requirement. For the Strict Cycle, another motivation for this condition, is redundantly derived by Shallowness, a principle which is independently motivated in the present system. Thus we should look for a different possibility to derive the binary branching requirement.

A natural direction under the Minimalist program is to think that the binary branching is part of the minimalism of the phrase structure. The relations allowed by the X-bar theory are head-complement (sister relation) and Spec-head (non-sister relation). Each relation must be minimally defined on the elements. A head may or may not have a relation with other elements. If a head has a relation, then it has only one for each kind. That is, a head can have at most one complement and at most one Spec.

This line of speculation may lead to deriving the requirement that a X-bar projection can only reach the double-bar level. Consider the following structure:

(86)

\[
\begin{array}{c}
\text{X}''' \\
\quad \beta \quad \text{X}'' \\
\quad \alpha \quad \text{X}' \\
\quad \text{X} \\
\ldots
\end{array}
\]

Here the head Y is adjoined to the head X before Z'' is inserted. Remember that we have defined the notion of complement domain in such a way that Z'' is included in the complement domain of Y in (i). If Y has a feature that can check a feature of Z'', Greed will be satisfied. Thus, we are not totally free from the problem of multiple branching at the single-bar level.
Here the head X has two Spec-head relations, one with α and another with β, referring as Spec to an element that is allowed by the X-bar projection but is not a sister of the head. The structure (86) is ruled out under the minimalism of the phrase structure. If α or β, or both are missing in (86), the triple-bar and double-bar levels are conflated since they are not significantly differentiated. (The single bar-level is significantly differentiated from the double-bar level since it determines the sisterhood with respect to the head. Thus the double-bar and single-bar levels are not conflated even if there is no Spec.)

Leaving the binary branching requirement, let us now consider further consequences of our system.

First, in a similar way it derives the Strict Cycle, Shallowness derives the obligatory intermediate substitution, by which we mean the effect that an element must be raised through an intermediate Spec. We saw three such cases above. When we discussed Icelandic/German indefinite subjects/objects, we assumed that they cannot skip Spec TNSP/ASFP. When we discussed object extraction from a wh-island in French and Dutch, we assumed that Spec AGRoP cannot be skipped. When we discussed the that-t effect, we assumed that a wh-phrase cannot skip the Spec of an intermediate that.

Take the last case for illustration, and consider the following:

(87) a. \[
\text{[CP what do you think [CP t' that [ John bought t ]]]}
\]

b. \[
\text{[CP what do you think [CP that [ John bought t ]]]}
\]
When *that* is introduced in (87a), the phrase-marker is extended to C'. The next Move-α operation is the raising of *what* to the Spec of *that*, which extends the phrase marker to CP. In (87b), on the other hand, the substitution of *that* must extend the phrase marker to CP. These two derivations compete with each other since the LF structures yielded by them are homomorphic, indeed identical, assuming that the copy of *what* in the intermediate Spec is deleted or ignored. Though (87b) has fewer Move-α operations, it is blocked by (87a) under Shallowness before Fewest Steps applies, since the substitution of *that* is shallower in (87a). (Neither of the two movements of *what* in (87a) are alternative to the movement of *what* in (87b).)

The first movement of *what* in (87a) observes Strict Cycle. For the second movement to immediately follow the first, the substitution of *that* must extend the phrase marker to CP to enable the substitution of *think*. Thus, the "successive movement" of *what* is not really successive.

Under our approach it follows that the obligatoriness of intermediate substitution, just like Strict Cycle, does not hold at LF. That is, the Spec of *that* can be and therefore must be skipped at LF. As a result it is predicted that no *that*-t effect will be observed in covert syntax, which is probably correct though there is a complication as noted before.

Since Procrastinate precedes Shallowness, we expect that a non-*wh* cannot be raised to Spec AGRoP to make shallower the operation introducing AGRo, and that Spec of *that* cannot be occupied by any element at surface, whether it is a *wh*-phrase or not.
Shallowness does not require adjunction to obey the Strict Cycle. This yields a good result in some cases. Consider the following example, for example:

(88)  ? what; do you wonder [ who; [ t j bought t i ]]

Here what must get AGRsP-adjoined before it is raised to the matrix Spec CP. If this intermediate adjunction obeyed the Strict Cycle, it must apply prior to the raising of who to the embedded Spec CP. In order to cross over what under Shallowness, who would have to get adjoined to AGRsP, violating Greed. Thus, our analysis of the impossible wh-island extraction depends on the fact that adjunction does not necessarily obeys the Strict Cycle.

However, there is a case where we would like adjunction to be subject to the Strict Cycle. First consider a fact noted by Fiengo (1980):

(89)  a. * what; did you tell who about t i
   b. what; did you talk to whom about t i

(89a) is ruled out as a crossing example, since who is shallower than what in its original position. In (89b), on the other hand whom is not as shallower than what in its original position, since there is a maximal projection dominating who but not what, namely the PP headed by to.

But Pesetsky (1982) observes that the nesting effect is still found in an analogous construction:
This fact is in a conflict with the above fact under our analysis, because it is not the case that one of the two wh-phrases involved is shallower than the other in their original positions in either case.

Suppose that the complement of P is Case-checked by adjoining to the PP. Then who in (90) must get adjoined to the PP to who before it is raised to Spec CP. Suppose further that the PP-adjunction of who must be applied prior to the raising of what subject. Now who, in the PP-adjointed position, is shallower than what subject in its original position. Under Shallowness what subject must get adjoined to the PP when it crosses over who, taking an extra Step to yield a nesting effect.69

Now the question is: what makes the PP-adjunction apply prior to the raising to Spec CP in this case? If the adjunction in question obeys the Strict Cycle, we will get the right result. We just saw a case in which an adjunction does not obey the Strict Cycle. Note however that there is a difference between the two cases: in one case it is an intermediate adjunction that is involved and in the other case it is an adjunction that induces a Case-checking operation. We may consider feature-checking operations to obey the Strict Cycle. Because the Strict Cycle for Move-α

69 Our analysis will naturally carry over to the prominence of the complement of P. (See Anderson (1979), Pesetsky (1991) and Branigan (1992), among others.) The following examples are from Branigan:

(i)  a. Jake likes to talk to his daughter about herself
    b. *Sarah wrote to him about Cliff’s smoking habit
    c. Sarah wrote to his wife about Cliff’s smoking habit
operations is derived from Shallowness, we may think that this principle should be extended to cover feature-checking operations. Actually, when we presented the definition of Shallowness, we did not restrict it to Move-α operations:

(4) Shallowness

An operation must be the shallowest.

The question to ask is how alternative feature-checking operations are restricted with regard to this principle. Let us assume that approximately, feature-checking operations are alternative to each other if they are applied in the same position. Since feature-checking operations are automatically applied when checking environments are created by movements, Shallowness amounts to stating that roughly, movements for feature-checking must apply as soon as possible. (Remember that Procrastinate precedes Shallowness.) Thus, an overt adjunction resulting in a feature-checking operation must target a separate phrase marker so as to make its depth ∅.

7. A-over-A.

The multiple wh-constructions we discussed above involve two wh-phrases that are originally in a superiority (or "c-command") relation. Our system can also deal with "A-over-A" cases, where two wh-phrases are in a dominance/inclusion relation. Shallowness subsumes the superiority condition and the A-over-A condition, thanks to our formulation of the relative depth of two elements, where it is irrelevant whether they are in dominance or not.
Consider first the A-over-A version of nesting effect, which Saito (1989) discusses, such as follows:

(88) a. ??which pictures of who [John saw t_i] do you wonder 
    b. *[which pictures of who [John saw t_i] do you wonder]

The two wh-phrases involved here are which pictures of who and who, the former including the latter. That is, the former is originally shallower than the latter.

The following derivation yields (88a):

(89)

First the wh-phrase \( \alpha \) (=which pictures of who) is raised to the embedded Spec CP in a single step. The wh-phrase \( \beta \) (=who) cannot be raised from inside \( \alpha \) to the matrix Spec CP in a single step under Shallowness. The adjunction to IP (=AGRsP) is an option for \( \beta \) to get as shallow as \( \alpha \). \( \beta \) is raised from the intermediate IP-adjoined position in a single step. The derivation involves three (relevant) steps. (Another option to get as shallow as \( \alpha \) is to get adjoined to \( \alpha \), which will lead to a Moderateness violation.)

The following derivation yields (88b):
First $\beta$ gets adjoined to $\alpha$ to get as shallow as $\alpha$, and then gets raised to the embedded Spec CP. ($\beta$ could get as shallow as $\alpha$ by adjoining to a maximal projection that is distinct from $\alpha$ but is as shallow as $\alpha$. If $\alpha$ is raised to Spec AGRoP for Case-checking before the extraction of $\beta$, then ASPP will be available for an intermediate adjunction site. The discussion here will not be essentially affected under this option.) $\alpha$ gets first adjoined to the embedded IP and then raised to the matrix Spec CP. The derivation involves four (relevant) steps. Assuming that the two derivations compete with each other, (90) is blocked by (2) under Fewest Steps.\footnote{\noindent (88a) is not perfect. We take this to be a subjacency effect, which is an additional factor we do not really understand.}

We should be able to have a similar explanation for the A-over-A counterpart of the superiority effect, which is exemplified in the following:

\begin{enumerate}
\item [(91) a.] [which pictures of who] did you see t
\item [(91) b.] ??who did you see [which pictures of t] \end{enumerate}

The following derivation yields (91a):
In overt syntax $\alpha$ is raised to Spec CP, and in covert syntax $\beta$ is adjoined to $\alpha$.

The following derivation yields (91b).

(93) \[
\begin{array}{c}
\text{[CP [\beta who] [IP [\alpha which pictures of t\beta ]]} \\
\end{array}
\]

$\beta$ is overtly raised to Spec CP through an intermediate adjunction to $\alpha$, taking two steps. $\alpha$ is covertly adjoined to $\beta$ so that the derivation involves three steps. Derivation (93) is blocked by derivation (92) under Fewest Steps.

A problem arises here. There is a reason to cast a doubt on the possibility of LF raising the whole wh-phrase (namely, $\alpha$). Consider the following:

(94) a. John$_i$ wondered [which picture of himself$_i$ [Mary saw t ]]

b. *John$_i$ wondered [when [Mary saw which picture of himself$_i$ ]]

In (94a) the raising of the wh-phrase containing himself allows the binding domain for himself to extend so as to include John. If the wh-raising were carried out in the same way in covert syntax, then himself should be successfully bound by John in (94b).

Similarly, we have the following contrast:

(95) a. ?? which picture of John$_i$ did he$_i$ see t

b. * when did he$_i$ see which picture of John$_i$
(95a) is far from perfect, but better than (95b), which is hopeless. If the entire complex wh-phrase can be covertly raised in (95b), the contrast will be surprising. These facts suggest that it cannot be the case that a complex wh-phrase is raised in covert syntax.

Hornstein and Weinberg (1990) argue that a part of a complex wh-phrase can be raised in covert syntax. Observe that complex wh-phrases do not reveal superiority effects (the following examples are due to Hornstein and Weinberg):

(96) a. which book did which friend review  
       b. which book did you expect which boy to buy  
       c. which book did you persuade which man to buy

They claim that these examples should be treated in the same way as the following example:

(97) whose book did whose friend review

Suppose that it is not whose friend but whose inside it that is covertly raised and adjoined to whose book, which is overtly raised. Whose in situ is originally not shallower or deeper than whose book, so that no superiority effect should be expected. If which acts like whose, it is raised in covert syntax without a superiority effect.

Suppose that the wh-feature on whose/which can be optionally transferred to a larger phrase which contains them. If a phrase has inherited
a *wh*-feature by transfer, it will function as a *wh*-phrase and the original *wh*-phrase will not function as such.

The following derivations will yield (91b), if the *wh*-feature of *which* is not transferred:

(98) a.  \[ \text{[CP who \text{[IP \text{[which \text{[pictures of t]}}\text{]}]}\text{[1]}]} \]

b.  \[ \text{[CP who \text{[IP \text{[which \text{[pictures of t]}}\text{]}]}\text{[2]}\text{[1]}]} \]

Both derivations are ruled out in favor of derivation (99), which yields (100):

(99) \[ \text{[CP which \text{[IP \text{[t \text{[pictures of who]}}\text{]}]}\text{[1]}]} \]

(100) which did you see [ t pictures of who ]

(99) blocks (98a) under Shallowness, and (98b) under Fewest Step.\textsuperscript{71}

Saito (1989) argues that the unacceptability of example (88b) is due to a violation of the Proper Binding Condition, which imposes an “S-structure” requirement that traces must be properly bound (cf. Fiengo (1977), May

\textsuperscript{71} We have to leave open two questions here: Why is it impossible to raise a complex *wh*-phrase in covert syntax, and why is it impossible to raise a part of a complex *wh*-phrase in overt syntax.

-107-
(1977). However, there is a reason to think that the account in terms of the Proper binding Condition is untenable.

Consider the following:

(101) a. [hit Mary] John did
   b. [proud of Mary] John is
   c. [hit by Mary] John was

Here we have interactions between “NP-movement” (the movement to the subject position for Case-checking) and “Predicate-fronting.” An “NP-trace” is left in the fronted predicate in (101a) and (101b), given the predicate-internal subject hypothesis, and in (101c) regardless of the validity of this hypothesis. These examples are acceptable even though the Proper Binding Condition is violated.

Lasnik and Saito (1992), however, argue that NP-traces do obey the Proper Binding Condition. They discuss the following examples (cf. Kroch and Joshi (1985), Barss (1986)):

(102) a. [how likely to win] is John
   b. * [how likely to be a riot] is there
   c. * [how likely to be taken of John] is advantage

They consider the unacceptability of (102b) and (102c) to be due to a violation of the Proper Binding Condition, a trace of the subject John being left in the fronted predicate. The reason that (102a) is acceptable is, they argue, that it does not involve a raising construction but a control construction, so that it is PRO that is in the fronted predicate. We are
dealing with a lexical ambiguity here. *Likely* is subject to a double entry: Raising and Control. With (102a) *likely* can be analyzed as a control verb since the matrix subject is an expression that has a capability of control. On the other hand, the control option is not available in (102b) and (102c), where the matrix subject is an expletive or an idiom chunk. (The control analysis should hold in such examples as in (101), which they do not discuss.)

There is evidence, however, to show that Raising and Predicate-fronting should be able to co-occur, as far as the standard assumption that Control, unlike Raising, obstructs a “backward” anaphor-binding is correct:

(103) a. the picture of himselfi shocked Johni
       b. each other'si pictures shocked themi

(104) a. the picture of himselfi is likely to shock Johni
       b. each other'si pictures are likely to shock themi

(105) a. [ shock Johni ] the picture of himselfi did
       b. [ shock themi ] each other'si pictures did

(106) a. [ how likely to shock Johni ] is the picture of himselfi
       b. [ how likely to shock themi ] are each other'si pictures

the Proper Binding Condition incorrectly rules out the examples in (105) and (106). This is enough to show that NP-traces do not obey the Proper Binding Condition, regardless of whether Control can be involved in the case of Predicate-fronting.

The fact that NP-traces are insensitive to the Proper Binding Condition is correctly predicted under our approach. Predicate-fronting is motivated by a feature-checking operation with respect to a *wh*-feature. (We are
assuming that such a non-interrogative Predicate-fronting as in (101) is a case of topicalization, and that the fronted predicate in that case has a wh-feature just as other kinds of topic phrases.) On the other hand NP-movement is motivated by a Case-checking operation on the subject. There is no interaction between these two kinds of movements with regard to Shallowness, so that no nesting effect can be expected.

Now the Proper Binding Condition is descriptively correct only in dealing with multiple wh-constructions as in (88), and its effect is deduced from more general considerations under our Economy system. (Actually, Saito (1989) argues that the condition properly deals with cases that involve scrambling, which we will not discuss here.)

We have to leave it open here what is additional factors to rule out (102b) and (102c), which are “overgenerated” by our system.72

As is expected, the trace left by Object shift does not observe the Proper Binding Condition. Consider the following German examples from den Besten and Webelhuth (1990):

(107) a. [gelesen] hat Hans das Buch nicht
       b. weil Hans das Buch nicht gelesen hat

72 The following contrast was pointed out in a class of Noam Chomsky in 1989:

(i) a. how likely to be a stone is there believed to be
     b. *how likely to be taken of John is advantage believed to be

This suggests that (102b) and (102c) may be degraded for different reasons. Note also that the Proper Binding Condition incorrectly rules out (ia).
They argue that the fronted (or topicalized) element in (107a) is not V but VP. The object is externalized from VP prior to VP-fronting. The possibility of the externalization of the object is independently attested by examples like (107b).

However, they also show that the externalization from VP is not sufficient for the fronting of the “remnant” VP to be successful, raising the following examples:

   b. weil Hans da nicht mit gerechnet hat

The complement of a preposition can be externalized from VP, as is shown by (108b). But the remnant VP-fronting is impossible, as is shown by (108a).

A difference between the object of a verb and the complement of a preposition is that the former is Case-checked outside VP, namely in Spec AGRoP, whereas the latter is Case-checked inside PP, therefore inside VP as far as PP is inside VP. Thus, in the case of the complement of a preposition, the externalization from VP is performed by scrambling. The externalization of the object of a verb from VP, on the other hand, is able to be (and has to be under Shallowness) done by raising to Spec AGRoP (Object shift). (The relative ordering of the object and the negative element indicates that the shifted object is further scrambled from Spec AGRoP.) Given this, the acceptability of (107a) is just one of our predictions. The unacceptability of (108a) suggests that traces left by scrambling obey the Proper Binding Condition. We are not in a position to ask why scrambling
behaves in that way, because we do not have sufficient understanding of it yet.

In English the object cannot be left behind by VP-fronting:

    b. * [ read ] John did the book not

This is because English does not allow Object shift under Procrastinate, and does not have scrambling, either.

Interestingly, there is a way to force Object shift even in English, however. Compare the following German example from den Besten and Webelhuth (1990) and English example from Takano (1992):

(110) (?*)[ t$_i$ gelesen ]j weiβ ich nicht [ wa$_a$i [ er t$_j$ hat ]]
(111) ?? [ fix t with the hammer ], I wonder which car John did

Den Besten and Webelhuth observe that although (2) is unacceptable in a dialect with a strong wh-island effect, it is fairly good in a dialect with a weak wh-island effect. (They also note that the Dutch counterpart is good.) This is not a surprise at all. The object is raised to Spec AGROPO before it is raised to the embedded Spec CP. The raising to Spec CP (Object shift) is motivated by an Accusative checking operation rather than a wh-checking operation, the former being applied prior to the latter. A movement that affects VP, which counts as a wh-phrase, will never be an alternative operation to Object shift under Shallowness. The derivation after Object shift will be a nesting with regard to the two wh-phrases, was and VP. With (111), which only reveals a weak wh-island effect, the derivation proceeds in
the same way. As argued before, Object shift before wh-raising satisfies Procrastinate, and therefore it is forced under Shallowness.

Compare further the following examples, (112) from den Besten and Webelhuth (1990), and (113) from Takano (1992):

(112)  * [ gesagt [daß er t\_i lesen will] ]j weiß ich nicht [ was\_i [ er t\_j hat ]]
(113)  * [ claim that John fixed t with the hammer ], I wonder which car Bill did

Here Object shift is applied inside the most deeply embedded clause, and therefore inside the fronted VP. Thus, the derivations yield an A-over-A version of crossing.

Takano (1992) also observes that a preposition stranding is impossible in English, just as in German:

(114)  * [ speak to t in that manner ], I wonder who John did

Again, this is because the complement of a preposition never has its Case-checking position outside VP.73

An entire PP complement is Case-checked by ASP, however. Thus we expect that it behaves like an object in the relevant respect. This expectation is borne out, as is observed by Takano (1992):

73 Howard Lasnik (personal communication) has pointed out the possibility of "reanalysis" of speak and to as a single verb so that the complement of the preposition is Case-checked in Spec AGRsP just as the ordinary Accusative object of verb, which will be problematic for our analysis. He also has pointed out that evidence for reanalysis is provided not only by pseudo-passivization, as has been often argued in the literature, but also by antecedent-contained deletion.
Here to whom is adjoined to ASPP (or AGRoP) to get Case-checked before it is raised to the embedded Spec CP.

Takano (1992) argues that the wh-phrase inside the fronted VP undergoes an “extraposition” from it to yield acceptable examples (111) and (115). The unacceptability of (113) and (114) is a reflex of the fact that extraposition cannot be from CP or PP. A problem with this approach is that VP-fronting cannot leave behind an extraposed element:

(116) a. *[fix t with the hammer], I wonder who did [the car which he bought from Mary]
b. *[speak t in that manner], I wonder who did [to the woman from Tokyo]

Probably an extraposed element is still inside VP (cf. Culicover and Rochemont (1990), Larson (19xx), Pesetsky (19xx)). Note also that the examples in (116), just like those in (109), cannot be derived through overt raising of a stranded element to a Case-checking position, since it would violate Procrastinate. Under our approach the unacceptability of (113) and (114) is regarded as a reflex of the locality of a raising for Case-checking.

A raising Case-checking is usually clause-bounded. But the following examples show that it is not always so:

(117) a. *[want [to buy]], I wonder what you did
b. * [ want [ to put the computer on ] ], I wonder which desk you did
c. ??[ want [ to put the computer ] ], I wonder on which desk you did
d. ?? [ want [ to sing a song ] ], I wonder where/at which party you did

With (117a), the object must be raised through the most internal Spec AGRoP under Shallowness. Even if there is an available Spec AGRoP in the next clause, which the fronted VP is originally from, it would be a moderateness violation if the object is further raised there. With (117b), there is no appropriate Case-checker for the complement of a preposition in a higher clause. Thus there is no way to avoid an A-over-A crossing in these cases.

With (117c), the PP complement is raised out of the VP which want is the head of to get adjoined to ASPP or AGRoP for Case-checking, and it is further raised to Spec CP. And then the VP is fronted, yielding a nesting. Similarly, with (117d), the adjunct is raised out of the VP to get adjoined to ASPP or TNSP to avoid an A-over-A configuration.

The contrast found concerning the remnant VP-fronting is duplicated concerning the ordinary superiority effect:
(118) a. * [ promise Mary [ to buy ] ], I wonder what you did
    b. * [ promise Mary [ to put the computer on ] ], I wonder which desk
        you did
    c.?? [ promise Mary [ to put the computer ] ], I wonder on which desk
        you did
    d.?? [ promise Mary [ to sing a song ] ], I wonder where/at which party
        you did

(119) a.?? what did you promise whom [ to buy ]
    b.?? which desk did you promise whom [ to put the computer on ]
    c. on which desk did you promise whom [ to put the computer ]
    d. where/at which party did you promise whom [ to sing a song ]

We assume that an arguments or adjunct is generated in a Spec of a
lexical head, $\alpha$, and is interpreted in its Case-checking position. To receive
the interpretation with respect to $\alpha$, it must be in the domain of $\alpha$, where the
relevant domain of $a$ is the set of nodes dominating $\alpha$ at LF. Usually lexical
heads only raise for feature-checking in its clause. However, in Control (, and ECM and Raising) constructions embedded heads raise to the matrix in the course of derivation, extending their domains beyond the clause boundary. Thus, the Case-checking position can be outside the original clause in that case.


Thiersch, Craig. 1985. VP and Scrambling in the German Mittelfeld. Ms., Tilburg University.