

# LANGUAGE-PARTICULAR PROCESSES AND THE EARLINESS PRINCIPLE\*

David Pesetsky/MIT  
June 1989

## **1 Earliness vs. Economy**

### *1.1 Economy of Derivation*

In a recent paper, Chomsky (1989) has proposed two principles which choose among competing transformational derivations. He calls them principles of "Economy of Derivation". These are the Least Effort principle and the Last Resort principle, seen in (1a-b). (The nomenclature is partially my own: Chomsky uses the term "Principle of Least Effort" for (1a-b) together.)

#### (1) Principles of Economy

- a. Principle of Least Effort: If two derivations from a given D-structure each yield legitimate outputs, and one contains more transformational steps than the other, only the shorter derivation is grammatical.
- b. Last Resort Principle: "UG principles are applied wherever possible, with language-particular rules used only to "save" a D-structure yielding no output."

This paper proposes an alternative to these principles. Chomsky's arguments for his approach involve a reanalysis of Pollock's (1989) discoveries concerning verb movement and INFL. I will begin by briefly summarizing how Chomsky's system works.

Chomsky uses a filter due to Lasnik (1981) as a prime mover for the various transformations and insertions found in the English and French verbal auxiliary system. Lasnik's filter requires morphemes designated as affixes to be "supported" by lexical material at PF. It is informally stated in (2):

#### (2) Lasnik's Filter: An affix must be lexically supported at PF.

I will assume that lexically supported is defined as "sister to a non-empty category marked [-affix]", though other cases of lexical support are imaginable (e.g. PF cliticization involving mere linear adjacency). The configurations of lexical support relevant to this paper are those created by adjunction of the non-affix to the affix or by adjunction of the affix to the non-affix.<sup>1</sup>

Now consider the problem of satisfying Lasnik's filter for inflectional affixes. UG and English grammar provide three possibilities: V-to-I Raising, I-to-V Lowering, and do-support.<sup>2</sup> Lasnik's filter requires that one of these possibilities be chosen, but does not explain a rigid pecking order that exists among these possibilities. Thus, Chomsky notes that V-to-I Raising is required whenever it is possible, as it is in French finite clauses:

---

\*This is a very rough draft, minimally expanded from a talk read at GLOW 1989. Here and there I indicate what I intend to elaborate on later. In particular, the citations and footnotes are quite incomplete, and dates for citations have been supplied from memory.

(3) **V-to-I Raising**

- a. Marie ne parle<sub>i</sub> pas e<sub>i</sub> français.
- b. Marie parle<sub>i</sub> souvent français.

**I-to-V Lowering**

- c. \*Marie ne e<sub>i</sub> pas parl-e<sub>i</sub> français.
- d. \*Marie e<sub>i</sub> souvent parl-e<sub>i</sub> français.

On the other hand, we know that Lowering of I-to-V does exist, but only where V-to-I is impossible, as it is with English main verbs:

(4) **V-to-I Raising**

- a. \*Bill speaks<sub>i</sub> not e<sub>i</sub> French.
- b. \*Bill speaks<sub>i</sub> often e<sub>i</sub> French.

**I-to-V Lowering**

- c. Bill [<sub>INFL</sub> e<sub>i</sub>] speak-s<sub>i</sub> French.
- d. Bill often [<sub>INFL</sub> e<sub>i</sub>] speak-s<sub>i</sub> French.

Why does V-to-I Raising take precedence over I-to-V Lowering? Chomsky suggests that length of derivation makes the crucial distinction. Given the ECP, it is somewhat surprising that Affix Lowering should be allowed at all. Chomsky proposed that the ECP is satisfied for the trace of finite affix Lowering because at LF the trace position is actually filled by Raising of V back to INFL. Crucially, this sort of “round trip” derivation involves two steps, Lowering and Raising, while a derivation that simply raises V to INFL involves only one, Raising. The “Least Effort” principle requires that the shorter derivation be picked: hence the round trip induced by Lowering is picked only when the one-way trip involving S-structure Raising is unavailable.

The round trip derivation found with I-to-V Lowering is empirically supported by the Head Movement Constraint (HMC) effects (Travis 1984) induced by negation. Chomsky follows the analysis assumed in the first part of Pollock (1989) in treating negation as a head of a NegP (but cf. section ??? for an alternative suggestion, also due to Pollock). If the HMC derives from the ECP, then an intervening negation should not affect the S-structure Lowering portion of the round trip, but should block the LF Raising portion. Indeed, it is exactly in negative clauses that simple I-to-V Lowering is impossible. Instead, an inflected do is inserted above Neg to support INFL — Do support:

- (5) a. \*Bill e<sub>i</sub> not speak-s<sub>i</sub> French.
- b. Bill does not speak French.

But this possibility raises a new question. Insertion of do in INFL involves only one step, and should produce derivations no longer than those involving Raising. Why isn't do-insertion the method of choice for satisfying Lasnik's filter where Raising is blocked? If do insertion could apply in these cases, it would yield ungrammatical examples like (6), with do non-contrastive:

- (6) \*Bill does speak French.

To explain the fact that do is inserted only when neither V-to-I nor I-to-V is possible, Chomsky invokes an additional condition, over and above the Least Effort principle: an absolute ban on the use of language-particular insertion rules like do-support when there is an alternative legal derivation involving UG movement — be it Raising or Lowering. This is the Last Resort Principle, seen in (1b).

Summarizing, Chomsky proposes (i) an absolute preference for UG principles over Language Particular rules, and (ii) a “Least Effort” principle that favors the shortest derivation from a given D-structure.

Let us assume that the movements and insertions that are assumed on Chomsky’s analysis are correct, including the round trip derivation for cases of finite-I-to-V Lowering. In the beginning, however, I will assume a simplified version of his analysis in order to keep our rather complicated discussion manageable. In particular, two questions arise with respect to Chomsky’s analysis that are quite important, but avoidable for the moment. The first is seen in (7a):

- (7) a. Question: Why can main verbs in English raise to INFL at LF, but not at S-structure?
- b. Provisional Answer: English INFL is “ $\theta$ -opaque” at S-structure, but the effects of this type of opacity are turned off at LF.

Chomsky’s actual answer to this question follows Pollock in part by relying on the structural distinction between an I adjoined to a V (which is a V) and a V adjoined to an I (which is an I).<sup>3</sup>

Another question can be seen in (8a):

- (8) a. Question: Why can have and be verbs in English, and all verbs in French finite clauses, raise over negation at S-structure, while negation blocks raising at LF?
- b. Provisional Answer: Not, like Pollock’s pas, is a modifier of NegP, not its head. The head of NegP is in fact empty at S-structure, allowing movement through it, but is filled at LF, blocking movement through it.

The phenomenon of LF lexical insertion of functional heads has been argued to exist for the English complementizer for in Pesetsky (in prep.). Chomsky’s actual answer to (8a) is quite different, and has to do with issues to which I return in section ???.

The simplifications in (7) and (8) will introduce certain distortions into the argument, which will be cleared up in section ???. Leaving the questions in (7) and (8) aside, and assuming the basic facts of Chomsky’s analysis, I will propose an alternative characterization of the reasons for the various steps.

## 1.2 The Earliness Principle

Chomsky’s treatment of the auxiliary system hinges indirectly on the notion “satisfaction of a filter”, in the sense that Economy singles out the derivation that satisfies all grammatical filters in the fewest number of steps. My alternative relies in a different way on a somewhat extended notion of “satisfaction at a level”. The notion can be described informally as follows: a filter is satisfied at the level at which its actual requirements have been met and the chains of all elements affected by the filter have been made legal. By “affected by a filter” I mean “mentioned in the SD of the filter”. This notion has the consequence for Lasnik’s filter described in (9):

- (9) Lasnik’s Filter is satisfied at level  $L_j$  for Affix  $\alpha$  iff
- (1)  $\alpha$  is supported at  $L_j$ , and
  - (2) no members of the chain containing  $\alpha$  at  $L_j$  violate any grammatical principles at  $L_j$  or any later level.

The notion “satisfaction” can be described more formally as in (10):

- (10) **A filter F whose SD involves elements  $A=(\alpha_1, \dots, \alpha_n)$  is satisfied at level  $L_j$  iff for any  $\alpha_i$  in A, no chain-mate of  $\alpha_i$  (including  $\alpha_i$  itself) violates any grammatical principles (including those of F itself) at  $L_{(k)j}$ .**

Now let us ask a new question — how early in the derivation is Lasnik’s filter satisfied in each of the cases described by Chomsky (where I understand satisfied as in (10))? There are three cases to consider, which are schematized in (11) below.

When the preferred option, Raising, can take place, Lasnik’s Filter is satisfied at S-structure. This is (11a).

When the next most preferred option, Lowering, takes place, the filter is satisfied at LF — the level at which the head of the chain containing INFL satisfies the ECP. This is (11b).

When do insertion has applied, however, something more needs to be said. Suppose do-insertion, as a Language-Particular process, is not merely a special type of rule, but is a rule that applies at a special level of representation set aside for Language-Particular insertion rules. Let us call this level LP-structure and position the level after S-structure, perhaps in the position of Chomsky and Lasnik’s (1977) “Phonetic Form”. Then, when do insertion has applied, the earliest level at which Lasnik’s Filter is satisfied is LP-structure. This is (11c):

- (11) **At what level is Lasnik’s Filter satisfied?**

a. <b>V-to-I Raising</b>		<b>S-structure.</b>
b. <b>(Finite) I-to-V Lowering + V-to-I Raising</b>		<b>LF</b>
c. <b>Do-support</b>		<b>LP-structure.</b>

Now recall that the preferences for satisfaction of Lasnik’s filter are as in (12):

- (12) **V-to-I > I-to-V > do**

Putting (11) and (12) together, it is clear that the combined effects of Chomsky’s Least Effort and Last Resort principles can be captured by the Earliness Principle in (13) below:

- (13) **Earliness Principle: Satisfy filters as early as possible on the hierarchy of levels: (DS >) SS >LF >LP.<sup>4</sup>**

Crucially, the ordering required by the Earliness Principle involves a linearization of the forked structure of the T-model, so that LF precedes LP-structure. I have no great light to shed on why this particular linearization is chosen by the theory, as opposed to a linearization that would place LF last. Speculation on this question might proceed as follows: one possibility is that the linearization might differ, depending on whether a morphological filter like Lasnik’s is at stake, or a filter more closely related to LF. Another possibility is that D-structure, S-structure and LF form a unit as the levels whose content is given by UG, while LP-structure stands apart.<sup>5</sup> These suggestions remain entirely speculative for now, however.

### 1.3 Differences

The Earliness Principle differs in three salient respects from Chomsky’s proposals.

First, it does not require examination of derivations as a whole, but only the structure of chains affected by the Structural Description of Filters. This difference may have some computational interest, but since I have not attempted to examine these questions, I will leave them for now.

Second, Earliness is a homogeneous condition, unlike the two distinct conditions of Economy given in (1). In the next part of this paper, I will try to establish that a “homogeneous” Earliness Principle is feasible in one crucial respect. The Earliness Principle relies upon the idea that Language-Particular insertion rules like do-support are not merely marked as Language-Particular, but actually apply at a level of representation set aside for rules of this sort. This idea must be defended. I will attempt to do just this, by developing two arguments that do-support does not feed Move  $\alpha$ . The arguments will involve inversion constructions and adverb placement.

Third, the Earliness Principle is both stronger and weaker than Economy, in ways that I will argue are advantageous:

**The way in which it is Earliness is correctly weaker than Economy:** since Earliness does not keep track of derivation length, Earliness, but not Economy, should allow “spontaneous” movement — movement that takes place for no reason at all. In Part Three of this paper, I will argue against the Economy approach by showing that spontaneous movement over an adverb can be found in the English verbal and auxiliary systems.

**The way in which Earliness is correctly stronger than Economy:** consider two derivations of equal length, where one satisfies a filter earlier than the other. Earliness predicts that only the derivation that satisfies the filter earlier should be permitted, but Economy predicts equal status for the two derivations. In the last section, I shall provide a case of this sort, where Earliness seems to be supported over Economy.

Let us now begin with the argument for LP-structure.

## 2 LP Structure: The Inertness of do

The grammar we are arguing for has the articulations in (14):



This grammar predicts that LP insertion rules should not feed any S-structure processes. Suppose Move  $\alpha$  is part of the mapping from D-structure to S-structure and LF, but not part of the mapping from S-structure to LP-structure. We predict that inserted do should not undergo the sort of head-movement that other verbs undergo. I will provide two arguments that this is correct. The first argument involves the inability of do to undergo I-to-C Raising. The second argument involves the inability of do to raise over adverbs.

## 2.1 Argument #1 for the Inertness of do: Conditional Inversion (C-Inv)

English, like a number of Germanic and Romance languages, shows a type of subject-verb inversion in conditionals. The complementary distribution of Conditional Inversion (C-Inv) and lexical if, seen in (15), suggests, as noted by den Besten (1983) and Holmberg (1986), that the Conditional Inversion crucially involves COMP. Let us therefore assume, with these authors, that C-Inv is a rule that moves the contents of INFL to C.

- (15)
- a. If John had read the book, he would have known the answer.
  - b. Had John read the book, he would have known the answer.
  - c. \*If had John read the book, he would have known the answer.
  - d. \*Had if John read the book, he would have known the answer.

We will show that do does not undergo this rule, and use this fact to argue for LP-structure. To accord any significance to this fact, however, it is first necessary to look at what elements of the verbal and auxiliary systems may undergo this rule.<sup>6</sup>

In English, C-Inv applies only to counterfactuals. Thus, whenever C-Inv is found, the protasis is always presupposed to be false (or unlikely). I will use the ability to take an apodasis with the modal would as a sign that a conditional is counterfactual, which seems correct.

### 2.1.1 Non-inverted Counterfactuals

Before asking what elements can occur in counterfactuals with inversion, we must ask what elements can occur in the INFL of the protasis of counterfactuals without inversion. A complete listing can be found in (16), where (16a-g) contain auxiliaries and modals that do occur in non-inverted counterfactuals, and (16h-o) contains those modals that do not occur in counterfactuals.

- (16)
- a. If John had solved the problem, he would have shown up.
  - b. If Mary should meet him, she would certainly come and tell us. ["non-obligational should"]
  - c. If John were to solve the problem, we would be happy.
  - d. If Mary were dying, she would look worse.
  - e. If Mary could speak French, she would have shown up.
  - f. ?If we were to take out the garbage every day, they would have left us a note.  
(on the reading "if we were to" ≈ "If we were expected to")
  - g. If John would drive a little faster, he would get there a little sooner. ["agentive would"]
  - h. \*If Mary can speak French, she would have translated for us.
  - i. \*If Bill may leave, we would have been told.
  - j. \*If Bill might leave, Sue would have informed us.
  - k. \*If Sue must take out the garbage each morning, she would have asked for higher wages. [cf. (93f)]
  - l. \*If Bill ought to take out the garbage each morning, Sue would have informed us.
  - m. \*If Mary shall speak French, she would have started already.
  - n. \*If Bill should take out the garbage, we would have known about it. ["obligational should"]
  - o. \*If Sue will speak French, she would have told us.

(17) shows that paraphrases of many of the bad modals in (16) are acceptable. We therefore can't look to their meaning for a simple answer:

- (17) a. If Bill were permitted to leave, Sue would have informed us.  
(cf. (16g))  
b. If Bill were supposed to take out the garbage each morning,  
Sue would have informed us. (cf. (16l), (16n))

Now let us try to make sense out of the data in (16).

### 2.1.2 *Counterfactuals and Past Tense*

First, we need to make a few remarks about the “non-obligational should” seen in (16b). Except for an interesting usage in factives, non-obligational should is not found outside the protasis of a conditional. (I am ignoring the normative rule that substitutes should for would in the 1st person.) We thus cannot directly test whether a past-tense usage is available in normal assertions. I suggest that should is simply a form of would that occurs in the protasis of a counterfactual, as seen in (19):

- (18) a. If there should be a riot, it would be bad for the cause.  
b. \*If there would be a riot, it would be bad for the cause.  
[cf. (16g) with agentive would]  
c. \*If there should be a riot, it should be bad for the cause.  
d. \*If there would be a riot, it should be bad for the cause.

- (19) would<sub>non-agentive</sub> → should / protasis of a counterfactual

With should treated in this fashion, the generalization in (20) appears to be true:

- (20) What can occur in the protasis of a counterfactual?  
An auxiliary or modal  $\alpha$  is acceptable in INFL of the protasis of a counterfactual iff  $\alpha$  is acceptable as a past-tense form.

The relevance of the discussion of should is that there does exist a past tense usage of would, seen in (21b).<sup>7</sup> The letters on the examples of (21) correspond to the letters in (16):

- (21) a. Bill had finished the book the previous Friday.  
b. Mary would often repeat your remarks during those years.  
c. Bill was to die in the Great War.  
d. They were dying by the thousands.  
e. That's strange – Sue could speak French yesterday!  
f. Mary was to take out garbage every day, in return for which she received a small salary.  
g. When we pulled into the service station, the mechanic would not look at our car, on the grounds that it was Sunday.  
h. \*That's strange – Sue can speak French yesterday.  
i. \*Bill may leave yesterday.  
j. \*Bill might leave yesterday.  
k. \*Sue must take out the garbage each morning in those days.  
l. \*Bill ought (not (to)) take out the garbage each morning in those days.  
m. \*Mary shall speak French yesterday.  
n. \*Bill should take out the garbage yesterday.  
o. \*Sue will speak French yesterday.

### 2.1.3 *Do in counterfactual conditionals*

Now let us consider the do of do support. Do behaves as predicted so far. It is acceptable as a past-tense form, as seen in (22):

(22) It didn't rain yesterday.

— and thus is acceptable in the protasis of a counterfactual, as seen in (23):

- (23) a. If it didn't rain, we wouldn't have crops.  
b. If it did rain in Spain, it would fall mainly on the plain, right?

These observations will be important in the next section.

#### 2.1.4 Inversion

Finally, we return to inversion. We are not surprised to see in (24h-o) that modals that do not participate in simple counterfactuals also do not participate in inverted counterfactuals. On the other hand, the examples in (24a-g) are surprising: every modals and auxiliary in (21a-g) occurs in simple counterfactuals, but only had, should, were and non-obligational were-to occur in inverted counterfactuals. Could, obligational were-to, and agentive would all fail to occur in the inverted forms. The chart in (25) summarizes the data seen so far:

#### (24) Auxiliaries that independently occur in counterfactuals with 'if' (surprising facts):

- a. Had John solved the problem, he would have shown up.
- b. Should Mary meet him, she would certainly come and tell us.
- c. Were John to solve the problem, we would be happy.
- d. Were Mary dying, she would look worse.
- e. \*Could Mary speak French, she would have shown up.
- f. \*Were we to take out the garbage every day, they would have left us a note.  
(on the reading "were we to" ≈ "we were expected to...")
- g. \*Would John drive a little faster, he would get there a little sooner.

#### Auxiliaries that do not independently occur in counterfactuals with 'if' (unsurprising facts):

- h. \*Can Mary speak French, she would have translated for us.
- i. \*May Bill leave, we would have been told.
- j. \*Might Bill leave, Sue would have informed us.
- k. \*Must Sue take out the garbage each morning, she would have asked for higher wages. [cf. (93f)]
- l. \*Ought Bill to take out the garbage each morning, Sue would have informed us.
- m. \*Shall Mary speak French, she would have started already.
- n. \*Should Bill take out the garbage, we would have known about it.
- o. \*Will Sue speak French, she would have told us.



(25)

	ok in counter- factual protasis	ok as true Past tense	undergoes I-to-C	
a. had	+	+	+	<b>non-<math>\theta</math>-assigners</b>
b. should/would	+	+	+	
c. were-to (NON-OBL)	+	+	+	
d. were	+	+	+	
e. could	+	+	-	<b><math>\theta</math>-assigners</b>
f. were-to (OBL)	?	+	-	
g. would (AGENT)	+	+	-	
h. can	-	-	-	
i. may	-	-	-	
j. might	-	-	-	
k. must	-	-	-	
l. ought	-	-	-	
m. shall (AGENT)	-	-	-	
n. should (OBLIG)	-	-	-	
o. will	-	-	-	

The contrast between (25a-d) and (25e-g) are immediately and strikingly reminiscent of Pollock’s description of verb movement to INFL in English and in French infinitives.

Recall that Pollock suggested that certain instances of INFL are designated “[+ $\theta$ -opaque]” (Pollock called them simply “opaque”). A  $\theta$ -assigner that moves to such an INFL cannot assign its  $\theta$ -role through its trace; this creates a  $\theta$ -criterion violation. It is worth noting that the arguments for Pollock’s Description were somewhat weak in that they rested on a particular theory (due to Guéron and Kayne) under which possessional have is treated as a licenser of a possessional Small Clause, rather than as a  $\theta$ -assigning predicate in its own right. This assumption was necessary to account for V-to-I by possessional have in some registers of English usage. If have were itself a  $\theta$ -assigner, such movement should be impossible:

- (26)a. Sue hasn’t any conception of the truth.
- b. Have you a proper appreciation of art?

Pollock’s treatment of have may be correct, but surely has at least the status of a point of debate.

The evidence mounts for the correctness of this linkage between  $\theta$ -assignment and movement when we observe the same generalization governing I-to-C in (25). I-to-C affects a wider range of verbal elements than are affected by V-to-I, because I-to-C also affects elements that are base-generated in INFL, e.g. the modals.

I thus argue that conditional COMP is  $\theta$ -opaque just like English INFL — and that the cut between (25a-d) and (25a-g) is simply the cut between  $\theta$ -assigners and non- $\theta$ -assigners. Whether I-to-C in these Cases is an adjunction, as in Pollock’s analysis of movement to T and to AGR, or is a substitution, as the complementary distribution with if suggests, I will leave open.

One point should be noted here, which might cause confusion. The same  $\theta$ -assigning modals that cannot move to C may, of course, occur in INFL. Why is this? An answer can be given if the effect of  $\theta$ -opacity is crucially tied to movement (whether specifically adjunction, as in Pollock’s work, I leave open). There is evidence, familiar since Emonds, that English modals are

base-generated in INFL. Note the facts in infinitives, where aspectual have and be merely fail to raise, but modals cannot occur at all:<sup>8</sup>

- (27) a. I believe Bill to not have read the assignment.  
b. I consider someone to not be doing the work when they...  
c. \*I consider someone to not be able to speak French when they...  
d. \*I consider someone to not can speak French when they...

Finally a speculation. We have accounted for which of the verbs that occur in counterfactuals can also occur in inverted counterfactuals. We have not accounted for the restriction of I-to-C to counterfactuals in the first place. I suggest that the restriction of I-to-C to counterfactuals is also an example of Pollock's Description. Assume that the availability of a past tense form for a modal really means the availability of an unmarked tense form to which a past-tense interpretation can be assigned in certain environments. Imagine that the relation borne by tense to its clause is in some fashion analogous to the relation of a  $\theta$ -assigner to its arguments. The impossibility of moving to C a form with full tense interpretation can be likewise analogized to the impossibility of moving a  $\theta$ -assigner to a  $\theta$ -opaque category. This would prevent all but verbs in an unmarked tense form from moving to C in English. Among conditionals, this in effect restricts movement to counterfactuals.

### 2.1.5 Non-inversion with do

Leaving these speculations aside, now consider do. We saw in (22)-(23) that do participates in counterfactuals. This is not surprising, since it has a past-tense form in accordance with (20).<sup>9</sup>

Now let us consider do's expected behavior with respect to Pollock's observations. Do is a non- $\theta$ -assigner par excellence. This can be easily demonstrated by (28):

- (28) **Do is a non- $\theta$ -assigner**  
There didn't seem to be any need for this example.

In this light, it is very surprising that it does not undergo I-to-C, as can be seen in (29):

- (29) a. If it rained, the game would be cancelled.  
b. \*Did it rain, the game would be cancelled.  
c. If it didn't rain, we wouldn't have crops.  
d. \*Did it not rain, we wouldn't have crops.

Let us call the inability of do to move "do-inertness". It is important to be clear about why (29) is interesting: we have developed a coherent and possibly correct characterization of the verbs that undergo C-Inv. By all rights, do should undergo this process. It is in this light that do-inertness cries out for an explanation, which I will now provide.

I begin by specifying that do is inserted in under (INFL, IP), as in (30):<sup>10</sup>

- (30)  $\emptyset \rightarrow \text{do} / (\text{INFL}, \text{IP})$

Given (30), LP-structure explains do-inertness: if rule (30) applies at LP-structure, it applies after Move  $\alpha$ . The failure of do to undergo I-to-C (in the light of our understanding of what other elements undergo the rule) is directly explained by the assumption that LP-structure exists and is a level past S-structure and Move  $\alpha$ . In turn, LP-structure is a necessary part of the Earliness account of the English and French auxiliaries.

2.1.6 A Necessary Aside: Question Inversion

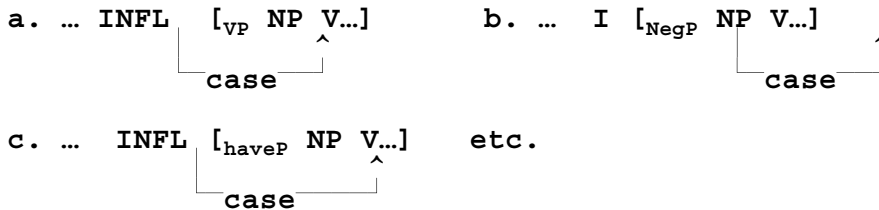
A brief aside is in order: If the data above are explained using “deep” considerations like  $\theta$ -assignment and the order of components, then we expect all I-to-C movement to behave alike. We might therefore expect Question Inversion or Negative Inversion to distinguish among  $\theta$ -assigning modals, non-assigning modals, and do. This is a patently false prediction, as (31)-(34) show:

- (31) a. What can Bill accomplish?  
       b. What did Bill accomplish?
- (32) a. At no time did she leave.  
       b. At no time could she leave.

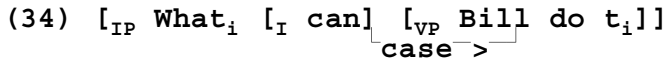
I would like to briefly propose a solution to this problem. With Hale (English), Diesing (Yiddish), and Rögnvaldsson/Thráinsson (Icelandic), I explain the lack of restrictions on modals and do in these constructions by showing that they are not instances of I-to-C; in fact, they are not instances of I-to-anywhere. Instead, in keeping with an old suggestion of McCawley’s, made in “English as a VSO Language”, let us assume that the “inversion” seen in questions involves not INFL raising to C, but incomplete subject raising within IP.

In agreement with much recent work by Kitagawa, Kuroda, Sportiche/Koopman and Fukui/Speas, among others, I assume that [SPEC,IP] is a non- $\theta$  position in simple sentences. Main verb subjects receive their  $\theta$ -roles within VP, and typically raise to [SPEC,IP] to receive nominative Case by (SPEC, Head) agreement. But the theory leaves open another possibility (explored first by Kitagawa and by Koopman/Sportiche), that the subject may raise only part-way, and INFL may assign nominative Case via ECM to the SPEC of its complement, as seen in (33):<sup>11</sup>

(33) Exceptional Case Marking by INFL



Let us assume that INFL may indeed assign case via ECM as seen in (33). I suggest that this is exactly what happens in English Question Inversion constructions. The reason the subject does not move all the way to the (SPEC,IP) is that the (SPEC,IP) is the landing site for matrix WH-movement — in other words an A-bar position. This analysis is sketched in (34):



This analysis follows if matrix clauses are IP, rather than CP. This can be argued on independent grounds. The argument goes as follows: consider the familiar observation that matrix clauses in English do not contain a visible COMP ((35)); yet there is no apparent head-governor to license an empty COMP. The need for such a governor can be seen in the familiar contrast between (35b) and (35c). We have a way out if we claim that matrix clauses are bare IPs.

- (35) a. (\*That) Mary left.  
       b. \*[[<sub>C</sub> e] Mary left] comes as a surprise.  
       c. I believe (that) Mary left.

Let us also propose that the converse of is true: embedded clauses are always CPs. This is clearly true for tensed clauses, as seen again by (35b).<sup>12</sup> In other work (Pesetsky (in preparation)), I defend this assumption for infinitives as well (arguing against bare IPs or CP-deletion analyses).

I thus assume (36)

- (36) a. (Non-exclamative) matrix clauses are IP.  
 b. Embedded clauses are CP.

Consider some consequences of these assumptions. First, in matrix questions, the feature Q that triggers WH-movement cannot be in C (since matrix clauses are IPs), but must be in INFL. In embedded clauses, by contrast, Q must be in C to be accessible to selection by higher predicates. Next, in matrix questions, WH-movement cannot be movement to the specifier of a non-existent CP, but must be adjunction to IP or substitution for [SPEC,IP]. Assume that [SPEC,IP] is the target for matrix WH-movement, and can function as an A-bar position.<sup>13</sup>

As a consequence, in matrix questions, unless the subject itself is a WH-phrase, it does not move to [SPEC,IP], but remains in the next highest SPEC, as seen in (37a-b). There it does receive Case from INFL by ECM. Evidence for this is provided by the adjacency effect seen in (37c):

- (37) a. [<sub>IP</sub> What<sub>i</sub> [<sub>I</sub> can] [<sub>VP</sub> Bill do t<sub>i</sub>]]  
 b. I wonder what<sub>i</sub> Bill can now do t<sub>i</sub>.  
 c. \*What can now Bill do t<sub>i</sub>?

The idea that [SPEC,IP] can be an A-bar position stems from work by Diesing on Yiddish, and by Rögnvaldsson & Thráinsson on Icelandic. Icelandic and Yiddish differ from English in allowing [SPEC,IP] to function as an A-bar position not only for question words and negatives, but also for topics— yielding the well-known phenomenon of “embedded V2” in these languages: the verb moves to INFL, and the topic moves to [SPEC,IP].<sup>14</sup> The proposal for English is summarized in (38):

- (38) a. Matrix clauses: WH moves to SPEC,IP  
 INFL Case-marks the subject of VP by ECM  
 b. Embedded clauses: WH moves to SPEC,CP  
 INFL Case-marks its SPEC by SPEC-Head agreement

This analysis is supported by data from topicalization and from the distribution of do support in questions.

**Topicalization evidence:** Suppose matrix WH-movement lands in (SPEC,IP), but embedded WH-movement lands in (SPEC,CP), as I have suggested. Consider the consequences of Baltin’s observation that English topics are IP-adjoined. Topics should land to the left of a moved WH-phrase in a matrix question, and to the right of the moved WH-phrase in an embedded question. While all topicalization in a question is somewhat odd, these predictions seem correct, as the data in (39)-(42) show.<sup>15</sup>

**Matrix questions**

- (39) a. ?<sub>[IP</sub>A book like this, <sub>[IP</sub> why should I buy?]]  
 b. ?This book, to whom should we give? (Watanabe (1988))  
 c. ?These prices, what can anyone do about?  
     (Langendoen (1979), via Watanabe)  
 d. ?To Bill, what will you give for Christmas?  
 e. ?And a book like this, to whom would you give?  
 f. ?And to Cynthia, what do you think you will send?  
     (Delahunty (1983), via Watanabe)  
 g. ?With Bill, what did you discuss?
- (40) a. \*<sub>[IP</sub> Why [<sub>?</sub> a book like this [<sub>?</sub> should I buy?]]]  
 b. \*To whom this book should Bill give?  
 c. \*\*To whom should this book Bill give?

**Embedded questions**

- (41) a. ?I wonder <sub>[CP</sub> why <sub>[IP</sub> A book like this, I should buy]].  
 b. ?I wonder to whom this book we should give?  
 c. ?Tell me what to Bill you're going to give for Christmas?  
 d. ?Ask him what book to John he would give.  
 e. ?I need to know what with Bill he's going to discuss.
- (42) a. \*I wonder <sub>[CP</sub>A book like this, <sub>[CP</sub> why I should buy]]?  
 b. \*I wonder this book to whom we should give?  
 c. \*I need to know to John what he would give.  
 d. \*Tell me this book to whom you will give.

**Evidence from do support:** The presence of do in non-subject matrix questions vs. its absence in subject matrix questions and in embedded questions also follows. The facts seen in (43)-(44), are clear enough. Note that it is not enough to rule out do when a matrix subject is questioned, as in the ECP account of Koopman (1984) — given that it is obligatory elsewhere —, nor to require do by a general condition on a Q-morpheme in C, as in Chomsky (1989) — given its absence with subject questions.

I suggest that the presence of do is required by the need to Case-mark the subject. Do is inserted only where necessary to satisfy this requirement — in keeping with its Last Resort character seen elsewhere in the system.

Let us consider the relevant cases. Suppose a non-subject is WH-moved in a matrix clause. The result we want is the obligatory insertion of do. Assume the structure is as in (43). The problem is to assign Case to the subject John. If INFL contains a modal, have or be, it will assign Case by ECM. Suppose INFL lowers to V due to the absence of a M, have, or be: it will not be able to assign Case by ECM. There are two possible reasons for this: possibly the next XP down is not L-marked by the trace of INFL, or else perhaps the trace of INFL (unlike INFL itself) cannot assign structural Case. Either way, it follows that INFL cannot lower to V. To prevent a Case-filter violation, do is inserted. Do, being lexical, is allowed to Case-mark by ECM, like modal and aspectual verbs:

- (43) a. \*<sub>[IP</sub> What [<sub>I</sub> e] <sub>[VP</sub> John accomplished]]?  
 b. <sub>[IP</sub> What [<sub>I</sub> did] <sub>[VP</sub> John accomplish]]?  
 c. <sub>[IP</sub> What [<sub>I</sub> can] <sub>[VP</sub> John accomplish]]?

Suppose now that the subject is WH-moved, as in (44)a-b. Here we have an interesting situation: (SPEC,IP) is an A-bar position as in (43), yet it is filled by the subject, as if it were an A-position. Nominative Case is quite straightforwardly assignable by INFL to (SPEC,IP)

via the SPEC-Head relation. The A-bar chain containing who receives Case; hence no do-support is needed.<sup>16</sup> The Last Resort character of do-support ensures that do-support applies only here.

- (44) a. [<sub>IP</sub> Who [<sub>I</sub> e] [<sub>VP</sub> t<sub>i</sub> left]]?  
 b. \* [<sub>IP</sub> Who [<sub>I</sub> did] [<sub>VP</sub> t<sub>i</sub> leave]]?

Finally, in embedded questions, which are CPs, the subject can Raise to (SPEC,IP), so that there is no problem with nominative Case assignment in the normal fashion, and do is not needed.

Actually, something more needs to be said, given that we are deriving the “Last Resort” character of do Insertion from the Earliness Principle, rather than from any “Last Resort Principle” as such. In fact, Earliness as developed so far require INFL Lowering in examples like (44)b, thereby preventing Do Insertion and incorrectly allowing no grammatical output. We must thus revise the notion of “satisfies” in (9)-(10) so that Lasnik’s Filter is not satisfied until the Subject which is to be Case-marked by INFL is also legal with respect to all grammatical principles. I will not spell out this revision here [i.e. in this draft], but it will involve a generalization of the notion of “chain” presented there such that if INFL is to assign nominative Case to the subject, the two elements must form a chain.<sup>17</sup>

There is thus a plausible account of Question Inversion that does not interfere with our account of C-Inv, and brings with it certain advantages of its own.<sup>18</sup> The crucial point is that if matrix Question Inversion is not an instance of I-to-C, then the availability of do in questions is not a problem for do-inertness. Recall in turn that our whole discussion of do-inertness is an argument for LP-structure, which in turn is a crucial part of our Earliness hypothesis.

## 2.2 Argument #2 for the Inertness of do: Adverbs

I assume that adverbs are generated as right or left sisters of XP or X’, depending on the nature of the adverb. Emonds has argued that V-movement can move a verb leftward to INFL over an adverb, yielding V Adv NP order, as in (45):

- (45) Pierre parle<sub>i</sub> à peine t<sub>i</sub> l’italien.

Pollock added the important observation that movement over an adverb does not necessarily diagnose V-to-I, but could diagnose V movement to a position between V and INFL which he called “AGR”. I shall call this position, not AGR, but  $\mu$ . I will be arguing later that  $\mu$  is not an agreement element at all, but is, in fact, contentless (or, perhaps, an empty verb). Pollock’s evidence for the existence of the  $\mu$ -position comes from French infinitives. French infinitives do not allow  $\theta$ -assigners to move all the way to INFL, but do allow them to move “half-way” to  $\mu$  over an adverbial, as seen in (46a-b):

- (46) a. [<sub>IP</sub> PRO I [ <sub>$\mu$</sub>  parler<sub>i</sub>] [<sub>VP</sub> a peine [<sub>VP</sub> t<sub>i</sub> l’italien]]]...  
 b. \* [<sub>IP</sub> PRO [<sub>I</sub> (ne) parler<sub>i</sub>] [<sub>NegP</sub> pas [<sub>VP</sub> t<sub>i</sub> l’italien]]]...

Now let us add a new observation: leftward movement over an adverb in English is possible for almost every verbal element in IP — not merely for the main verb:

**Either side of modals, base-generated in INFL.**

- (47) a. Bill absolutely must be shovelling his walkway by 6:00.  
 b. Bill must absolutely be shovelling his walkway by 6:00.  
 c. \*Bill must be absolutely shovelling his walkway by 6:00.
- (48) a. John soon will have won the Nobel prize.  
 b. John will soon have won the Nobel prize.  
 c. \*John will have soon won the Nobel prize.

**Either side of aspectuals in INFL.**

- (49) a. Mary lately has read War and Peace. [cf. above]  
 b. Mary has lately read War and Peace.  
 c. Mary has lately been reading War and Peace.  
 d. ??Mary has been lately reading War and Peace.  
 e. ??Mary was lately reading War and Peace.  
 f. ??Mary lately said that the world is round.

**Either side of aspectuals not in INFL.**

- (50) a. ?Mary will lately have read War and Peace. [cf. above]  
 b. Mary will have lately read War and Peace.  
 c. Mary will have lately been reading War and Peace.  
 d. ??Mary will have been lately reading War and Peace.  
 e. ??Mary will be lately reading War and Peace.
- (51) a. Mary soon will recently have won the Nobel prize.  
 b. Mary soon will have recently won the Nobel prize.  
 c. Mary will soon recently have won the Nobel prize.  
 d. Mary will soon have recently won the Nobel prize.  
 e. \*Mary will have soon recently won the Nobel prize.

These data suggest that  $\mu$  is not limited to the area immediately above the main verb. Each member of the auxiliary system, including INFL itself, may have a  $\mu$ -position above it. (52a) demonstrates such a position above have, and (52b) demonstrates such a position for modals in INFL. Note that I am suggesting that  $\mu$  projects a maximal phrase containing a specifier into which Bill moves in (52b).

- (52) a. [<sub>IP</sub> Bill<sub>j</sub> might<sub>i</sub> [ <sub>$\mu$ P</sub> t<sub>j</sub> have<sub>i</sub> [<sub>PerfectP</sub> long t<sub>i</sub> [<sub>VP</sub> said ...]]]]  
           b. [ <sub>$\mu$ P</sub> Bill<sub>j</sub> will<sub>i</sub> [<sub>IP</sub> soon t<sub>j</sub> t<sub>i</sub>] [<sub>VP</sub> ...]]]

But now note something important: once again, do is exceptional. Since every other element in the auxiliary system can have its own  $\mu$ P above it, into which it moves, we might expect the same possibilities for auxiliary do. Yet the facts are otherwise: do does not move to  $\mu$ . Relevant data are seen in (53)-(54).

**The inertness of 'do' for movement over an adverb**

- (53) a. Sue cleverly has not opened her present.  
 b. Sue has cleverly not opened her present.  
 c. Sue cleverly did not open her present.  
 d. \*Sue did cleverly not open her present.

- (54) a. Mary certainly has not returned from the conference yet.  
b. Mary has certainly not returned from the conference yet.  
c. Mary certainly did not understand the question.  
d. \*Mary did certainly not understand the question.

These data are expected if verb Adv order derives from verb-movement to  $\mu$ . Such movement, like the I-to-C seen above, is impossible for do, since do is inserted under (INFL,IP) too late for Move  $\alpha$  to apply. And this is exactly what is predicted if do Support takes place at LP-structure.

Finally, to pick up the main thread once more: if do-insertion is not merely a “Last Resort”, but applies in a special late component of the grammar, then the Earliness proposal becomes feasible. We have argued that this is empirically correct: if do is always inserted in INFL of IP after Move  $\alpha$  has its last chance to apply, we can explain its inertness. We can now proceed to arguments directly in support of Earliness.

### **3 Economy is Too Strong: $\mu$ or AGR?**

Recall my observation earlier that Economy, but not Earliness, prohibits “spontaneous movement” unmotivated by any filter such as Lasnik’s Filter or the ECP. I now turn to an argument that this restriction imposed by Economy is too strong. My argument is designed to show a case where a longer derivation is taken even when the shorter derivation is available.

My arguments concern movement over an adverb to  $\mu$  in English — some of the facts we have just been discussing in another context. I will argue first that, contrary to Pollock and Chomsky analyses,  $\mu$  is not an AGR and is not any other type of contentful affix.

Assume that movement to  $\mu$  is motivated by Lasnik’s morphological filter. What kind of morphology is  $\mu$ ? We can see that though it is phonologically null, it is not the familiar “zero morpheme”, since it is semantically null as well: it makes no detectable contribution to meaning. Thus either  $\mu$  is not an affix, and hence cannot trigger Lasnik’s filter, or it has some syntactic role to play in the sentence. This is just what Pollock and Chomsky claim: they claim that  $\mu$  is some sort of agreement element. Pollock’s main evidence for identifying  $\mu$  with AGR is the claim that movement to  $\mu$  filters movement to INFL: if you can’t move to  $\mu$ , you can’t move to INFL. This is predicted by the HMC if  $\mu$  is an obligatory member of IP, and hence is an obligatory half-way house on the journey to INFL.

I will argue that Pollock was wrong in his description of what can move to  $\mu$ . Things can move to  $\mu$  that cannot move to INFL, and things can move to INFL that cannot move to  $\mu$ . Given the HMC, we conclude that  $\mu$  is not an obligatory member of IP.<sup>19</sup> Things can move to INFL that do not move to  $\mu$  first because  $\mu$  is not always present in the tree. The one restriction on movement to  $\mu$  will involve, not  $\theta$ -marking, but Case assignment. On the other hand, the one restriction on movement to INFL will involve, not Case-assignment, but  $\theta$ -marking.

I conclude from this that if  $\mu$  is an affix of some sort, it is not merely phonologically null and semantically null, but is syntactically null as well: it licenses nothing and makes no contribution to its structure. If this is a syntactic affix, it is like no other known affix. Notice as well that if such affixes exist that are semantically, syntactically and phonologically null, we expect to find affixes that are semantically and syntactically null, but not phonologically null. This would be an utterly optional and meaningless phonological string inserted as part of the morphology of certain words. No such affix exists, which strongly suggests that  $\mu$  is not an affix either. Therefore, I conclude that Lasnik’s filter does not force movement to  $\mu$ .



I will suggest finally that not even the ECP requires filling of  $\mu$  — though here I will have to cut short my explanation in the present draft. From all this I will conclude that movement to  $\mu$  is as unnecessary as  $\mu$  itself. Given that this movement does occur (as we've just seen), we have an argument for spontaneous movement.

### 3.1 Movement to $\mu$

#### 3.1.1 Prolegomena to main verb movement to $\mu$ in English

I will argue that the examples in (55a-c) have a structural description in which they involve leftward main-verb movement to  $\mu$ :

- (55) a. Bill INFL [ $\mu$  knocked<sub>i</sub>] [<sub>VP</sub> recently t<sub>i</sub> on it].  
 b. Sue looked carefully at him.  
 c. Harry relies frequently on it.  
 d. \*Bill pushed recently the door.  
 e. \*Sue saw frequently the movie.  
 f. \*Harry trusted frequently Mary.

The ordering seen with PP objects in (55a-c) is, of course, impossible with the NP objects seen in (55d-f). In fact, contrasts of this sort are the stock-in-trade of demonstrations of Case adjacency. I will suggest that this reflects a basic property of movement to  $\mu$  given in (56):

- (56) Elements in  $\mu$  may not assign Case through their trace.

We need to be careful to distinguish putative leftward verb movement over a left-peripheral adverb from rightward movement of a heavy PP over a right-peripheral adverb, as in (57), which would yield the same linear string:

- (57) Bill INFL [<sub>VP</sub> knocked t<sub>i</sub> recently] [on it]<sub>i</sub> .  
 >

There are a number of ways to make this distinction. Two are seen in (58). (58a) shows, following Rochemont, that rightward-shifted phrases are generally focused, hence incompatible with being old information. (58b) shows the familiar difficulty in extracting from a rightward-shifted phrase.

- (58) a. disambiguate HNPS by controlling the focus  
 #As for War and Peace, I gave to Bill that book  
 b. disambiguate HNPS by extracting from the object  
 ??War and Peace, which I gave to Bill a copy of..

Construction of examples involving as for phrases or extraction help structurally disambiguate these constructions from those involving rightward shift.<sup>20</sup>

As a modest demonstration that the extraction test is on the right track, we can examine examples that involve the order V Adv object, where (i) the adverb modifies some INFL-element  $\alpha$  (for example, Tense), and (ii) the verb cannot move to  $\alpha$ . Under these circumstances, V Adv object order can only be due to rightward movement of the object. Indeed, extraction is quite bad in such cases. An example is found in (59)-(60), involving the present perfect, interpreted as in traditional grammars (and cf. McCawley (1988)) as a present tense sentence with a past tense auxiliary. A present tense adverb like now in such a construction could only be taken to modify the present tense morpheme in INFL. Main verbs cannot move to (much lest to the left of) this present tense

morpheme. By contrast, a past tense adverb like recently in the present perfect could modify the perfect participle, allowing movement of the perfect participle leftward over recently to  $\mu$ . The data show that the main verb may move leftward over recently, but not over now:

- (59) a. \*As for this book, Bill has looked now at it.
- b. ?As for this book, Bill has looked recently at it.
- (60) a. This is what Bill has recently looked at -.
- b. This is what Bill has looked recently at -.
- c. This is what Bill has now looked at -.
- d. \*This is what Bill has looked now at -.

We must also exclude the possibility that the adverb in a V Adv object sequence modifies the Direct Object directly: looking at nominalizations is relevant here; the data is given in (61), and I will not discuss the issue here.

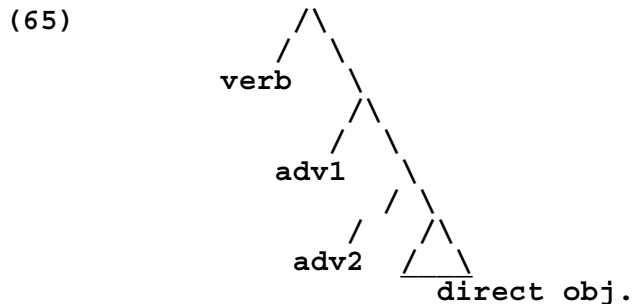
- (61) a. Bill relies merely on luck  
      (structurally ambiguous, given (61b))
- b. Bill's reliance merely on luck
- c. Bill relies heavily on luck  
      (not ambiguous, given (61d))
- d. \*Bill's reliance heavily on luck

Examples (62)-(64) make similar points:

- (62) a. Bill participated a bit in the proceedings.
- b. \*Bill's participation a bit in the proceedings.
- (63) a. Sue assisted partially with the preparations.
- b. \*Sue's assistance partially with the preparations
- (64) a. Bill thought carefully about the problem.
- b. \*Bill's thought carefully about the problem

### 3.1.2 Evidence for main verb movement to $\mu$ in English

The best evidence that (55a-c) may involve leftward main verb movement comes from constructions with stacked adverbs.<sup>21</sup> The form of the argument goes as follows: assume that the sisterhood requirement on verb and direct object is inviolate. Now suppose we find evidence in multiple adverb constructions for the hierarchy seen in (65):



This evidence will argue in favor of verb movement from a position between adv2 and the direct object. The following sections present evidence of this sort.

#### Scope argument:

Andrews (1983) noted that when adverbs are stacked on the left or right periphery of the VP, the relative scope of the adverbs is as predicted if the structures are “articulated” rather than “flat”, as indicated in (66)-(67):

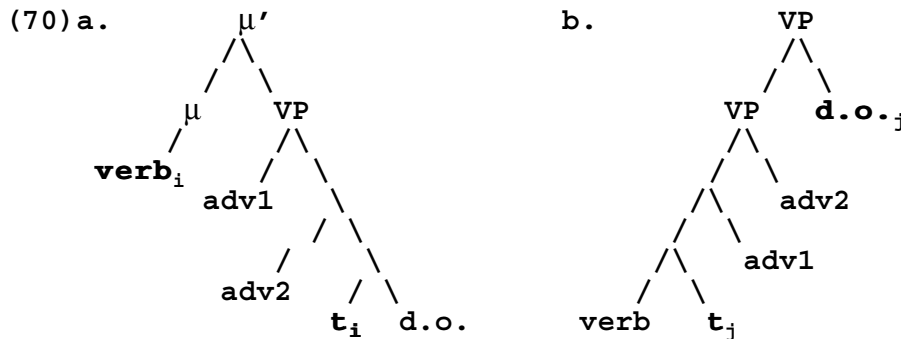
- (66) a. John [[[knocked on the door] intentionally] twice].  
 b. (?)John [twice [intentionally [knocked on the door]]].
- (67) a. John [[[knocked on the door] twice] intentionally].  
 b. (??)John [intentionally [twice [knocked on the door]]].  
 (Andrews (1983))

In (66a), twice unambiguously has scope over intentionally: the sentence can only refer to two events of intentional knocking. (66b) is unambiguous in the same way: it too can only refer to two events of intentional knocking. The examples in (67) have only the opposite scope interpretation: there was one intention, which was to knock twice.

Scope judgments of this sort give us a probe into the constituency of cases in which adverbs intervene between a main verb and its object. If we construct examples in which two adverbs come between a main verb and its object — examples of the form V Adv1 Adv2 PP, and the PP is modestly “heavy” — we observe that scope is suddenly ambiguous, as in (68a) or (69a).

I suggest that this ambiguity is structural: (68a) and (69a) either show adverbs stacked on the left of VP plus leftward verb movement, yielding the hierarchy in (70a), or else they show adverbs stacked on the right of VP plus rightward PP movement, yielding the hierarchy in (70b). Indeed, the heavier the PP gets, the more the latter interpretation is available.

- (68) a. John knocked intentionally twice on the door.  
 b. John knocked intentionally twice on the heavy oak door.
- (69) a. Bill relied stupidly twice on Mary.  
 b. Bill relied stupidly twice on the person you told me about.



Indeed, if we alter the examples in ways which tend to rule out a rightward Heavy Shift analysis of (70b), the relative scope of the two adverbs becomes unambiguously that predicted by (70a), as can be seen in (71).

- (71) **Disambiguating for rightward Heavy Shift**  
 a. As for Mary, Bill relied stupidly twice on her. (focus)  
 b. Mary's the one who Bill relied stupidly twice on \_\_. (extraction)  
 [unambiguously (stupidly (twice...))]

**Stacking restriction argument:**

A very similar argument can be constructed based on hierarchical restrictions on the stacking of adverbs. As is known from the work of Jackendoff and others, there are severe restrictions on the occurrence of adverbs of completion. For example, adverbs of completion can only attach to the X' that they modify, as can be seen from the examples in (72)-(73):

That it is the non-maximal X-bar and not XP can be seen in (73), where the idiomatic usage of the negated modal would not ('refuse') allows adverbs of completion only between the subject and the modal. If the modal is in INFL and heads the sentence, and if adverbs of completion could attach to XP, we might expect such an adverb to occur IP-initially, which is impossible:

- (72) a. Bill has completely finished his meal.  
b. \*Bill completely has finished his meal.  
c. \*Completely, Bill must have finished his meal.
- (73) a. Bill utterly would not leave the car.  
b. \*Utterly, Bill would not leave the car.

Subject-oriented adverbs show no such restrictions, as can be seen in (74):

- (74) a. Bill has cleverly finished his meal.  
b. Bill cleverly has finished his meal.  
c. Cleverly, Bill has finished his meal.

This difference in attachment site between adverbs of completion and subject-oriented adverbs leads to predictable restrictions on their cooccurrence. When both occur together on the left side of the VP, the subject-oriented adverb must precede the adverb of completion, as seen in (75a-b). Crucially, the same ordering restriction shows up when the two adverbs are niched between the verb and its direct object, as seen in (75c-d). Examples (76)-(78) provide more data of the same type:

- (75) a. Sue has been very cleverly completely staying in bed.  
b. \*Sue has been completely very cleverly staying in bed.  
c. Sue has been staying very cleverly completely in bed.  
d. \*Sue has been staying completely very cleverly in bed.
- (76) a. Mary has carelessly partially dealt with the problem.  
b. \*Mary has partially carelessly dealt with the problem.  
c. ?Mary has dealt carelessly partially with the problem.  
d. \*Mary has dealt partially carelessly with the problem.
- (77) a. Sue recently completely agreed with my comments.  
b. \*Sue completely recently agreed with my comments.  
c. ?Sue agreed recently completely with my comments.  
d. \*Sue agreed completely recently with my comments.
- (78) a. The French have at last completely given up on the Dutch.  
b. \*The French have completely at last given up on the Dutch.  
c. ?The French have given up at last completely on the Dutch.  
d. \*The French have given up completely at last on the Dutch.

We thus conclude — contra Pollock and Emonds — that main verbs do move leftward over adverbs in English, to the position that we are calling  $\mu$  — the position that Pollock called AGR-S, and Chomsky AGR-O.

### 3.2 Consequences of English Main Verb Movement to $\mu$

If my arguments have been correct, then Pollock was wrong about the properties of English  $\mu$ . Pollock (who identified  $\mu$  with AGR-S), claimed that English  $\mu$  is  $\theta$ -opaque. If I am right, English  $\mu$  is not  $\theta$ -opaque, but is “Case-opaque”— where “Case-opaque” means that a verb in  $\mu$  cannot assign Case via its trace.<sup>22</sup>

(79) English  
 $\mu$ : [- $\theta$ -opaque, +Case-opaque]

French  
 $\mu$ : [- $\theta$ -opaque, -Case-opaque]

Let us now ask about English finite INFL. English INFL must be [+ $\theta$ -opaque] — unlike  $\mu$  and unlike French finite INFL. The evidence for this is what it always was: English main verbs may not raise past negation to INFL or invert in questions.

English INFL is [+ $\theta$ -opaque], but is it [+Case-opaque]? I think the answer (unlike the answer for  $\mu$ ) must be “no”. The relevant evidence comes constructions in which it can be argued that a Case-marking verb has moved to INFL in English. Such examples are discussed in a recent manuscript by Lasnik.

Lasnik’s paper argues for a version of Belletti’s idea that existential be assigns Case to its object. In support of this claim, Lasnik notes that when existential be is not the main tensed verb, it cannot be separated from its object by an adverb. On the other hand, when existential be is the main tensed verb, it can be separated by an adverb. (80) shows this; (81)-(82) present similar data for have in INFL:

- (80) **Existential be moves to INFL**  
 a. There are never any cops when you need them.  
**Existential be moves to  $\mu$**   
 b. \*My whole life, there have been never any cops when I’ve needed them.  
**Non-existential be moves to  $\mu$**   
 c. My whole life, cops have been never where I’ve needed them.
- (81) **Have moves to INFL**  
 a. ?John has never time to do anything good.  
**Have moves to  $\mu$**   
 b. \*John has had never time to do anything good.  
**Movement to  $\mu$  over never is possible**  
 c. John relies never on anyone important.
- (82) a. John has always something on his mind.  
 b. \*John must have always something on his mind.  
 c. John knocks always on my door by mistake.

Lasnik’s data show that a verb moved to INFL can assign Case via its trace, while a verb moved to  $\mu$  cannot. Thus, although English  $\mu$  is [- $\theta$ -opaque, +Case-opaque], English INFL is [+ $\theta$ -opaque, -Case-opaque]. This is summarized in (83).<sup>23</sup>

(83) English

$\mu$ : [- $\theta$ -opaque, +Case-opaque]  
INFL: [+ $\theta$ -opaque, -Case-opaque]

French

$\mu$ : [- $\theta$ -opaque, -Case-opaque]  
INFL: [- $\theta$ -opaque, -Case-opaque]  
(finite)

But this result is of great importance: both Pollock and Chomsky assume tacitly that  $\mu$  is to be identified with some syntactically contentful position. Thus, Pollock suggests AGR-S; Chomsky, AGR-O. Neither of them give any real argument for these identifications — just, presumably, a background assumption that a head position must have some name and must fulfill some function other than merely acting as a landing site. The data summarized in (83) argue against this background assumption. If the HMC (or the ECP) holds of verb movement, then V-to-I should necessarily involve movement to  $\mu$  as an intermediate step. We should therefore be quite surprised to learn about verbs that can move to INFL but not move to  $\mu$ . Movement to  $\mu$  should filter movement to INFL, given the HMC.<sup>24</sup>

Yet we have just examined this type of “surprising” data. Verbs that do not assign a  $\theta$ -role but do assign Case — existential be and have — may move to INFL but may not move to  $\mu$ .

The simplest explanation of this phenomenon goes as follows: when a non- $\theta$ -marking verb moves to INFL, it does not have to pass through  $\mu$  because  $\mu$  does not have to be generated. In other words, I am claiming that  $\mu$  is not a syntactically contentful affix of any kind at all — neither AGR-S nor AGR-O nor any sort of contentful affix. Since it is also not a phonologically or semantically contentful affix, it would be an affix like none other we’ve seen, motivated only to save the Economy principles. I conclude that  $\mu$  is not an affix, and hence not subject to Lasnik’s filter. But remember that the argument is weak in one respect: if one is willing to investigate an affix like none we’ve seen to save the Economy principles, this possibility is, of course, open.

Let us now step back and consider how the argument has progressed. The argument is presented in (84).

- (84)i. A  $\mu$  position may exist above each auxiliary and main verb.
- ii. This position is not always present. Therefore it plays no crucial licensing role in the sentence.
- iii. It is therefore not an affix that attracts movement to it, nor does it provide something that V needs.
- iv. Yet verbs move to  $\mu$ .

We almost have what we need to conclude that a longer derivation may be chosen when a shorter is available — that is, an argument against Economy.

\*\*\*THIS SECTION TO BE EXPANDED (or spun off into a separate squib\*\*\*\*\*

There is one piece missing from the argument above, however. It may be the case that  $\mu$  is strictly an optional node (hence not AGR-anything) but we might still claim that when  $\mu$  is by chance generated in a tree, V movement to  $\mu$  is forced by the ECP. This would amount to claiming that empty  $\mu$  nodes must be head-governed or lexically filled — a not unreasonable claim. It looks like this claim is false for  $\mu$ , however, though a full development of the argument will await another draft or paper.

One rough-and-ready potential argument against an ECP motivation for movement to  $\mu$  comes if we allow or require semantically empty nodes to be deleted at LF, as in Chomsky's analysis of the auxiliary system. Deletion of  $\mu$  at LF should eliminate any ECP violation, making movement to  $\mu$  unnecessary.

Direct evidence for the legality of empty  $\mu$  comes from multiple quantifier floating. In a recent article, which I will assume is convincing, Sportiche has argued that the "quantifier floating" seen in (85) actually shows "quantifier stranding" in subject position internal to VP:

(85) **The kids<sub>i</sub> must have been [cleverly [VP [all t<sub>i</sub>] pretending to sleep]] (Sportiche (1988))**

If this idea is correct, then Floating phenomena give us a probe for intermediate subject positions. In a longer version of this paper, I demonstrate first that "floated" emphatic reflexives and subject-oriented even are also instances of the general Q-float phenomenon (for even it is first necessary to distinguish its floated use from its use in "association with focus" constructions in the sense of Jackendoff and Rooth). Now let us ask: if floated quantifiers are actually stranded in subject positions, what subject positions are relevant in multiple floating constructions like (86) (similar to some examples from Dowtie and Brodie (1984))?

- (86) a. **The kids have all each been awarded a prize.**  
 b. **The kids must have all themSELVES each been awarded a prize.**  
 c. **The KIDS must have even all themSELVES each been awarded a prize.**

It can be shown, based on ordering restrictions among the quantifiers that each of the floated elements in (86c) has been stranded separately, but I cannot pursue the matter here. The point of these examples for our purposes is to ask what position the middle floated elements are occupying. Even if one blames the order of the leftmost floater and have on leftward V-movement, we must still assume at least two phrases with subject positions sitting between have and been, whose heads are empty at S-structure. This is demonstrated in (87):

(87)  
**have<sub>i</sub> [<sub>μP</sub> even e<sub>i</sub> [<sub>μP</sub> all e [<sub>μP</sub> themSELVES e [<sub>VP</sub> each been awarded ...**

I thus conclude that empty-headed  $\mu$  projections do not need to be lexically filled, though the reasons for this are obscure for now.<sup>25</sup> Hence movement to  $\mu$  really is optional movement, and we have an argument against Economy principles.<sup>26</sup>

\*\*\*\*\*

### ***3.3 Alternatives to the argument that $\mu$ is not an affixal position.***

At this point, we must consider a number of alternatives to our conclusions concerning movement to  $\mu$  and to INFL. Up to now, I have contrasted my analysis with a drastically simplified version of Chomsky's hypotheses. If one considers the actual mechanisms that are needed to handle the cases Chomsky considered in his paper, a variety of important alternatives immediately suggest themselves.

For example, I have tacitly assumed that "Case-opacity" is a filter on movement, rather than a condition on representations. It followed from this assumption that when V-to-I fails to show Case-opacity effects, V does not stop in  $\mu$  on its way to I. I have also assumed the HMC of Travis. From all these assumptions plus our empirical observations, the conclusion followed that  $\mu$  is not an obligatory constituent of IP. Any of the various assumptions just discussed could be false, however. Indeed, the assumption about the HMC is quite crucially false in Chomsky's actual system: HMC

effects are derived from the ECP, and the HMC as a descriptive generalization is, in certain cases, argued to be wrong. The other assumptions are open to similar challenges.

Consider first Case-opacity. Case-opacity might not be a property of movement, but rather a property of chains created by movement. This alternative view might be summarized as in (88):

(88) **A chain that contains a Case-opaque position may not contain a position from which Case is assigned.**

Suppose that  $\mu$ , contrary to my arguments above, is an obligatorily present position. Consider the hypothesis that a trace in  $\mu$  can be deleted. If a trace in  $\mu$  could delete before condition (88) applies, we would still be able to maintain the hypothesis that movement to INFL stops in  $\mu$  in the face of the absence of Case-opacity effects on movement to INFL. Example (89) shows the derivation:

(89)

a.	INFL	[ $\mu$ e]	[ $_{VP}$ V...	$\xrightarrow{\text{V-to-}\mu}$
b.	INFL	[ $\mu$ V <sub>i</sub> ]	[ $_{VP}$ t <sub>i</sub> ...	$\xrightarrow{\text{V-to-I}}$
c.	V <sub>i</sub> -INFL	[ $\mu$ t <sub>i</sub>	[ $_{VP}$ t <sub>i</sub> ...	$\xrightarrow{\mu\text{-deletion}}$
d.	V <sub>i</sub> -INFL	[ $\mu$ $\emptyset$	[ $_{VP}$ t <sub>i</sub> ...	<b>satisfies (88)</b>

It is worth noting, however, that  $\mu$  now accomplishes nothing as a filter on movement to I, since the effects of Case-opacity are nullified by  $\mu$ -deletion before (88). A flowchart of questions and possible conclusions can be produced: We need to ask (A) whether there is any independent evidence for  $\mu$  as an obligatory constituent of IP. If the answer to is no, then we need to ask (B) whether there is at least any independent evidence against or in favor of  $\mu$ -deletion. If there is evidence in favor of  $\mu$ -deletion, then (C) the alternative view of Case-opacity effects shown above becomes plausible. If there is evidence neither against nor for  $\mu$ -deletion, we need to ask (D) whether  $\mu$ -deletion at least constitutes some sort of null hypothesis. If there is evidence against  $\mu$ -deletion (and no independent evidence in favor of the obligatory status of  $\mu$ ), then we may conclude (E) that the alternative sketched in (89) is implausible.

The answer to question (A) seems to be no, as we have seen above, but the answer to question (B), according to Chomsky (1989), is yes. Chomsky argues that deletion of  $\mu$  plays a crucial role in the derivation of grammatical structures which satisfy the ECP but violate Travis's HMC. Recall some of the basic properties of Chomsky's system, shared by the system I have been outlining as well:

1. Lasnik's filter requires I to participate in an adjoined structure containing V.
2. The filter is satisfied by V-to-I Raising whenever this is possible; V-to-I Raising is only possible for verbs of the have or be class, due to  $\theta$ -opacity.
3. Where V-to-I Raising is impossible, lower I-to-V. Except in the case where T may be deleted, the ECP motivates subsequent LF Raising to T.
4. Where Raising at LF is blocked, do may be inserted in I in the syntax.
5. What can block Raising at LF: a NegP with a filled head: [<sub>T</sub> e] Neg [<sub>V</sub> V AGR-S T]?

Now consider examples like (90a-b):



- (90) a. Bill has<sub>i</sub> not t<sub>i</sub> been reading his assignments.  
 b. Sue is<sub>i</sub> not t<sub>i</sub> happy.

Granted that Raising at LF past Negation is impossible (as noted in point 5 above), why is Raising at S-structure possible? This is where the replacement of Travis's HMC with the ECP is relevant, and where Chomsky makes crucial use of  $\mu$ -deletion.

Chomsky's suggests that the impossibility of movement over Negation at LF is due to the ECP. Chomsky here adopts Pollock's idea that Neg heads its own maximal projection.

The possibility of movement over Neg at S-structure is explained as follows: have or be moves from its underlying position over Neg to INFL in two steps. First, have or be moves to  $\mu$ , and then it moves from  $\mu$  to INFL:

- (91) Bill [<sub>INFL</sub> have<sub>i</sub>+INFL] not [ <sub>$\mu$</sub>  t'<sub>i</sub>] t<sub>i</sub> left.

The trace in  $\mu$ , t', antecedent-governs t<sub>i</sub>. As a consequence, t' assigns [+ $\gamma$ ] to t. This allows t to satisfy the ECP at LF. Of course, t' itself is not antecedent-governed by have, and should produce an ECP violation, since not intervenes. This is where deletion is important: t' may delete in the mapping from S-structure to LF, thus sidestepping the ECP.

Returning to the impossibility of movement over Neg at LF, Chomsky provides reasons why t' should be undeletable in these cases, which I will not discuss here. In any case, if Chomsky's discussion is correct, then S-structure movement, but not LF movement, can sidestep the ECP. In turn, if  $\mu$ -deletion is possible, then we also have a means of sidestepping Case-opacity formulated as in (8). Derivations that satisfy both the ECP and (8) may be schematized in (92):

- (92)
- |    |                      |  |                                    |                                       |                        |
|----|----------------------|--|------------------------------------|---------------------------------------|------------------------|
| a. | INFL                 | [ <sub><math>\mu</math></sub> e]               | [ <sub>VP</sub> V...               | $\xrightarrow{\text{V-to-}\mu}$       | >                      |
| b. | INFL                 | [ <sub><math>\mu</math></sub> V <sub>i</sub> ] | [ <sub>VP</sub> t <sub>i</sub> ... | $\xrightarrow{\text{V-to-I}}$         | >                      |
| c. | V <sub>i</sub> -INFL | [ <sub><math>\mu</math></sub> t <sub>i</sub>   | [ <sub>VP</sub> t <sub>i</sub> ... | $\xrightarrow{\text{\gamma-marking}}$ | >                      |
| c. | V <sub>i</sub> -INFL | [ <sub><math>\mu</math></sub> t <sub>i</sub>   | [ <sub>VP</sub> t <sub>i</sub> ... | $\xrightarrow{\text{\mu-deletion}}$   | >                      |
|    |                      |  | $\xrightarrow{\text{+\gamma}}$     |                                       | >                      |
| d. | V <sub>i</sub> -INFL | [ <sub><math>\mu</math></sub> $\emptyset$ ]    | [ <sub>VP</sub> t <sub>i</sub> ... |                                       | satisfies ECP and (88) |
|    |                      |  | $\xrightarrow{\text{+\gamma}}$     |                                       |                        |

Some of the data I have presented in this paper allow a rather strong argument against Chomsky's use of  $\mu$ -deletion as a method of sidestepping the ECP. This argument, if correct, eliminates any presently available independent motivation for  $\mu$ -deletion, but does require an alternative account of the S-structure/LF asymmetry in movement over Neg.

In section 2.2, we saw that  $\mu$  positions, while they may not be omnipresent, are ubiquitous: a  $\mu$  position can be found above each auxiliary or main verb. The relevant data was presented in (48)-(52), of which a sampling is repeated here:

- (93) a. Mary soon will recently have won the Nobel prize.  
 b. Mary soon will have recently won the Nobel prize.  
 c. Mary will soon recently have won the Nobel prize.  
 d. Mary will soon have recently won the Nobel prize.  
 e. \*Mary will have soon recently won the Nobel prize.

Now let us return to the way Chomsky's system allows S-structure movement over Neg. Verbs that can move over Neg like auxiliaries have and be move first to an intermediate position –  $\mu$ . The trace of this movement assigns  $+\gamma$  to the trace internal to VP, then deletes. The availability of this procedure has as a consequence that whenever movement over a head H can be accomplished in two steps, the first of which involves a position lower than H which is deletable, no ECP violation will ensue. This consequence opens the door to unwelcome HMC violations.

Consider, for example, the possibility of “leapfrogging” one auxiliary element over another, which is quite impossible. For example, only the highest auxiliary element or main verb may move to I:

- (94) a. Bill [<sub>INFL</sub> HAVE<sub>i</sub>] NEG t<sub>i</sub> BE READ the book.  
 b. \*Bill [<sub>INFL</sub> BE<sub>j</sub>] NEG HAVE t<sub>j</sub> READ the book.

The system does not rule this out without additional mechanisms. Consider the structure in (95):

- (95) NP INFL ( $\mu_1$  MODAL) NEG  $\mu_2$  HAVE EN  $\mu_3$  BE ING  $\mu_4$  [<sub>VP</sub> V ...]

If  $\mu$  may be an intermediate landing point for movement to INFL over negation by the highest verbal element, there is no reason why it should not be an intermediate landing point for movement to INFL by any member of the auxiliary system. For example:

- (96) a. Be moves to  $\mu_3$  and then over HAVE and NEG to INFL; the trace in  $\mu_3$   $\gamma$ -marks the trace of be and deletes.  
 b. V moves to  $\mu_4$ , and then over BE, HAVE and NEG to INFL; the trace in  $\mu_4$   $\gamma$ -marks the trace of V and deletes.

On the other hand, Lasnik (1981) notes that this sort of movement might create problems for the assignment of affixes like past participle -en or present participle -ing. On certain plausible assumptions about these affixes, for example, the derivation in (96a) would lead to outputs like (97a) (where be moves to INFL before -en can be affixed to it) or (97b) (where be moves to INFL after -en has been affixed to it):

- (97) a. Bill i-s<sub>i</sub> have t<sub>i</sub> read-ing-en the book.  
 b. Bill be-en<sub>i</sub>-s have t<sub>i</sub> read-ing the book.

Lasnik suggests that independent properties of English word-structure filter out forms like be-en-s or read-ing-en. This suggestion might conceivably be developed into a reasonable theory. Nonetheless, other examples can be created that are not amenable to this solution. Consider, for example, (98a-b):

- (98) a. Bill ha-s<sub>i</sub> not t<sub>i</sub> seemed [<sub>IP</sub> [<sub>I</sub> to] have enjoyed himself for many years now].  
 b. \*Bill ha-s<sub>i</sub> not have seemed [<sub>IP</sub> [<sub>I</sub> to] t<sub>i</sub> enjoyed himself for many years now].

Example (98b) shows verb raising from an embedded clause to the INFL of the matrix clause. Example (98a) shows normal movement within a single clause. By the hypothesis in Chomsky (1989), both examples involve two steps: (A) movement to the  $\mu$  minimally c-commanding the moved have, (B) movement to INFL. The first step creates the conditions necessary for  $\gamma$ -marking of the original trace of have. Thus, no considerations of Economy of Derivation distinguish the two cases. Additionally, neither example (98a) nor example (98b)

violates any conceivable principles of English word-structure. Each ends up with one instance of unaffixed have and one instance of have adjoined to INFL.

Example (98b) might be taken to violate the subjacency condition of Chomsky's Barriers, depending on the barrier status of the embedded  $\mu$ P, the embedded IP, and the auxiliary-verb projections of the higher clause. In the spirit of Barriers, however, it should be immediately obvious that (98b) is considerably worse than any subjacency violation.<sup>27</sup>

The particular examples discussed above were chosen in order not to beg the question of the nature of  $\mu$ . Depending on what conclusions we draw about  $\mu$ , other examples can make the same point as (98). For example, an auxiliary or main verb should be able to raise through the  $\mu$  that minimally c-commands it to some other  $\mu$ -projection, as in (99):

(99) **Bill must [ $\mu$  been<sub>i</sub>] have<sub>i</sub> [ $\mu$  t<sub>i</sub>] t<sub>i</sub> leaving.**

This example would not violate the ECP if the trace in  $\mu$  could delete. The example might be taken to violate some sort of morphological constraint if  $\mu$ , contrary to my claims, is some sort of affix. A restriction in the spirit of Lasnik's suggestions concerning (97) might filter out a verb that has acquired two  $\mu$ -affixes in the course of a derivation.

In any case, the status of (98) should suffice to make the desired point:  $\mu$ -deletion after  $\gamma$ -assignment allows too many HMC violations to satisfy the ECP. If we are to follow Chomsky's (1986) suggestion that all desirable instances of Travis's HMC reduce to the ECP, and if no plausible alternative principles take care of the troublesome cases, then we have an argument against  $\mu$ -deletion before  $\gamma$ -assignment.

If this is correct, then we need to find an alternative account of why S-structure movement across negation is possible, while LF movement is impossible. My best suggestion has in fact already been given in (8b), which I repeat below as (100):

(100) **Not, like Pollock's pas, is a modifier of NegP, not its head. The head of NegP is in fact empty at S-structure, allowing movement through it, but is filled at LF, blocking movement through it.**

This suggestion makes sense if adjunction to a filled  $\text{Neg}^\circ$  is not allowed and if  $\gamma$ -assignment precedes LF filling of  $\text{Neg}^\circ$ .<sup>28</sup> S-structure movement across NegP is possible because on this account — not because movement to some lower deletable position provides the  $\gamma$ -marking necessary for the ECP, but because movement to  $\text{Neg}^\circ$  itself provides the necessary  $\gamma$ -marking:

(101) **Mary has<sub>i</sub> [ $\text{NegP}$  not [ $\text{Neg}^\circ$  t<sub>i</sub>] [t<sub>i</sub> left the room]]**

As I noted in connection with (8b), there is at least one other known case of a head that behaves as at S-structure but filled at LF: this is the case of verbs that select irrealis infinitival complements — verbs like desire. These verbs behave at S-structure (e.g. for purposes of Exceptional Case Marking) as if the COMP of their object were empty or missing, but behave at LF (e.g. with respect to the ECP) as if this same COMP were filled. Let us assume that at S-structure and PF, the notion “empty category” refers to phonetic emptiness, so that a lexical item (like null for) with no phonetic matrix counts as empty for movement purposes — the position can be moved through. At LF, on the other hand, the notion “empty category” refers to categories that are both phonetically and semantically contentless. Thus, a null version of for will count as filled at LF, as will a null version of not.

Actually, Pollock, in a later section of his paper, proposes that “NegP” is actually “AssertionP”, where “Assertion<sup>o</sup>” can be modified by negation or by emphasis. Chomsky (1957) already noted a strong parallelism between negative sentences and sentences with emphatic do:

(102) **Mary did read the book.**

Pollock adapts Chomsky’s suggestions and posits that NegP and whatever accounts for focus as in (102) are both modifiers of Assertion<sup>o</sup>. If future investigations can invest Assertion<sup>o</sup> with some sort of semantic content, then we shall be justified in proposing an analogy between its behavior and the behavior of null for, and (8b) will be an acceptable account of the S-structure/LF asymmetry in movement over negation.

Taking up once more the main thread of this section, there are still a number of possible alternatives to our hypothesis concerning  $\mu$ , which I wish to mention briefly. For example, one might accept the argument made above against  $\mu$ -deletion before  $\gamma$ -assignment. One could then propose that  $\mu$ -deletion occurs as in (89), but does not counterbleed  $\gamma$ -assignment in the fashion suggested by Chomsky (1989). Instead, one would suggest that  $\mu$  deletes after the ECP has applied (and thus plays no role in explaining movement across negation), but before the version of Case-opacity in (88) applies. Notice that this hypothesis would make of  $\mu$ -deletion merely a device to explain why Case-opacity effects never show up for movement to INFL. In other words, this hypothesis would account for the desired facts, but would be (for now) ad hoc.

One final possibility would be to restrict Case-opacity effects to chains whose heads occupy  $\mu$ . Movement through  $\mu$  would escape Case-opacity effects, since the resulting chain would not be headed by  $\mu$ , but movement to  $\mu$  would be subject to Case-opacity. This hypothesis faces essentially the objections lodged against the ordering hypothesis of the preceding paragraph: the head/non-head distinction relates to nothing else in the system, and is therefore suspicious, though not inconceivable.

I therefore conclude, somewhat cautiously, that solutions that assume  $\mu$  as an obligatory constituent of IP are forced to develop ad hoc mechanisms to explain the lack of any filtering effect of  $\mu$  on movement to INFL. The most attractive such alternative, modeled on Chomsky’s use of  $\mu$ -deletion as an escape from ECP effects, is no different in this regard, since Chomsky’s use of  $\mu$ -deletion is cast in serious doubt by the examples given in (97)-(98).

### **3.4 Summary**

We saw in our discussion of do-inertness that the Earliness Alternative is feasible: there is good evidence that do-insertion applies at its own Level of Representation. We have just seen a somewhat more complicated argument which, if correct, militates against the Economy of Derivation. Finally, in Part Four, we will look at an argument that militates for Earliness. Fortunately, this argument is quite simple.

## **4 Economy is Too Weak: D-linking and WH-movement**

\*\*\*\*\*THIS SECTION TO BE EXPANDED SLIGHTLY\*\*\*\*\*

In earlier work of my own (Pesetsky (1987)), I argued that WH-movement is motivated in English and in languages like Polish by two separate conditions.

The first is a condition on the Q-morpheme in COMP (or INFL) which requires that one WH-phrase move to its SPEC by S-structure: thus every WH-question (particularly an

embedded question) will contain at least one instance of WH-movement to SPEC,CP (or (SPEC,IP), as discussed above)).

The second is the familiar condition on WH-phrases themselves which requires certain of them to move to an appropriate A-bar position by LF.

Why these two overlapping conditions? The evidence concerned a distinction between what I called “D(iscourse)-linked” and “non-D-linked” WH-phrases. Discourse-linked WH-phrases, roughly speaking, ask questions where the range of possible answers is limited to some set given in the prior discourse or “in the air”. WH-phrases of the form which person or which thing are prototypical D-linked phrases — but almost any WH-phrase may have a D-linked usage.<sup>29</sup>

Crucially, D-linked WH-phrases in English show none of the syntactic indications of LF WH-movement. They are, for example, immune from Superiority and ECP effects, as can be seen by comparing the contrasting examples in (103):

- (103) (Non-D-linked WH-phrases)
- a. Mary asked **who<sub>i</sub>** **e<sub>i</sub>** read **what**
  - b. \*Mary asked **what<sub>j</sub>** **who<sub>i</sub>** read **e<sub>j</sub>**
  - c. **who<sub>i</sub>** did you persuade **e<sub>i</sub>** to read **what**
  - d. \***what<sub>j</sub>** did you persuade **who** to read **e<sub>j</sub>**

- (104) (D-linked WH-phrases)
- a. Mary asked **which man<sub>i</sub>** **e<sub>i</sub>** read **which book**
  - b. Mary asked **which book<sub>j</sub>** **which man<sub>i</sub>** read **e<sub>j</sub>**
  - c. **which man<sub>i</sub>** did you persuade **e<sub>i</sub>** to read **which book**
  - d. **which book<sub>j</sub>** did you persuade **which man** to read **e<sub>j</sub>**

Taking the effects in (103) to diagnose LF WH-movement, I suggested that D-linked phrases need not undergo LF WH-movement to take scope, but are assigned scope by coindexation with a Q-morpheme at LF.

This left open the question of why D-linked phrases must undergo S-structure WH-movement in the examples in (105):

- (105) a. I wonder **which book** you read.  
b. \*I wonder you read **which book**.

My answer was that WH-movement in (105a) is forced not by the needs of the WH-phrase, but by the needs of the Q morpheme in COMP. I thus proposed that not only is (106b) a possible motive for WH-movement, but (106a) is as well:

- (106) a. **Q Filter:** Q must be supported by a WH-phrase in its SPEC.  
b. **Scope Filter:** All WH-phrases must be assigned scope:  
(i) either by (LF) movement, or  
(ii) for D-linked phrases, by coindexation with Q.

Important supporting evidence in favor of this approach, and in favor of the analysis of (103) vs. (106) in terms of LF movement vs. coindexation, was supplied by Polish. The relevant data were already presented by Wachowicz (1974), and have been consistently confirmed by most native speakers I have asked. As is well-known, Polish, like all Slavic languages, allows multiple WH-movement in multiple questions.<sup>30</sup> (107) gives an example:

(107) **Zastanawiam sie [kto co przyniesie].**  
**I-wonder who what will-bring**  
**'I wonder who will bring what'**

As has been often noted, the Slavic languages in this respect seem to “wear their LF on their sleeve”. In my earlier paper, I took it as an exciting fact that this “sleeve-wearing” extended to the distinction between D-linked and non-D-linked phrases. Wachowicz had already noted that there were certain circumstances under which WH-phrases could stay in situ at S-structure in Polish (the same facts appear to be true for Czech, Russian and perhaps Romanian), and these appeared to be precisely when the WH-phrases were D-linked. She considered examples like (108), and made the observation in (109):

(108) **W końcu, kto robi co?**  
**finally who does what**

(109)

“[Such] questions are somewhat different from echo questions. We can call them clarifying questions. The speaker could ask [(108)] in the following situation. There are various tasks, and several people to be assigned for them. Proposals have been made how to pair up people and tasks, but no fixed plan has been set up yet. The speaker of [(108)] is confused by the proposals, and wants to have a fixed plan.” (Wachowicz, 1974)

I thus noted that exactly those WH-phrases which, under my analysis of English, could be assigned scope without LF movement, were exempt from S-structure movement in multiple fronting languages like Polish. If one thinks more carefully about the matter, however, it becomes apparent that the contrast between the interpretive possibilities of (107) and (108) was not really explained. To be sure, the contrast between (107) and (108) suggested a link between D-linking and movement that strongly recalls the English data in (103) and (104), but Wachowicz’s fundamental observation actually had no satisfactory explanation.

If co in (108) happens to be D-linked and thus receives scope by coindexation, we know why it need not move and why it cannot move at S-structure or LF. But suppose co is not interpreted as D-linked? Why isn’t LF movement available to non-D-linked in-situ co?

In my earlier paper, I suggested that as a matter of parametric variation, Polish and English differ in two ways:, as seen in (110):

(110) <b>Pesetsky (1987)</b>	Pol	Eng
a. <b>multiple S-structure WH-movement</b>	+	-
b. <b>LF WH-movement</b>	-	+

No connection was made between these two English-Polish differences. Thus, the facts in (107)-(108) did not really follow from the D-linking hypothesis alone, but from D-linking combined with a stipulation about Polish LF. If the difference at LF is to be viewed as a parametric difference, it runs afoul of arguments by Higginbotham that there can in principle be no LF parameters.

In fact, preliminary evidence from Russian (which otherwise seems to behave like Polish; I have not checked the Polish equivalents as of this writing) suggests that the LF parameter is actually wrong. Consider phrases meaning ‘how many X’ and ‘how much’. I argued in Pesetsky

(1987) that these phrases were not automatically D-linked, and thus show Superiority and ECP effects in English. The Russian equivalent of ‘how much’ — skol’ko — behaves as predicted given the theory outlined above. Example (111b) shows only the D-linked reading for skol’ko, and (111a) shows only the non-D-linked reading:

(111)a. Kto skol’ko zaplatil?  
Who how-much paid

b. Kto zaplatil skol’ko?

For unknown reasons, however, full phrases of the form skol’ko N<sup>n</sup> are not allowed to participate in the multiple movement construction:

(112)a. ??Kto skol’ko dollarov zaplatil za ètu knigu?  
Who how-many dollars paid for this book

b. ??Kto skol’ko dollarov komu dal?  
who how-many dollars whom-DAT gave

Only the in situ variant is possible, as in (113):

(113)a. Kto zaplatil skol’ko dollarov za ètu knigu.

b. Kto komu dal skol’ko dollarov

What is important is that (113a-b) is ambiguous between the D-linked and non-D-linked interpretations, unlike the examples in (111). In other words, Russian does show LF WH-movement (yielding the non-D-linked reading of (113a-b)) when S-structure movement is for some reason unavailable.<sup>31</sup>

Clearly this is just the sort of situation predicted by the Earliness Principle: it is a fact about Polish that it allows the Scope Filter in (106a) to be satisfied at S-structure, since Polish has multiple S-structure WH-movement. Given this fact, the Earliness Principle requires that the Scope Filter be satisfied at S-structure whenever possible. The normal absence of LF WH-movement follows as an immediate consequence, as does the possibility of LF movement whenever S-structure movement is unavailable. English, by contrast, lacks multiple S-structure movement, and therefore cannot satisfy the Scope Filter for non-D-linked phrases until LF. Hence, English is allowed to wait until LF.

Thus, the facts in (107)-(108) now really can follow from the D-linking hypothesis: the only WH-phrases that remain unmoved at S-structure are those that never need to move (the D-linked phrases). All WH-phrases that need to move to satisfy the Scope Filter (the non-D-linked phrases) can move at S-structure in Polish, and thus must move at S-structure. Hence (108) can only have the D-linked interpretation for co. This revised picture is exemplified in (114):

(114)	<b><u>This paper</u></b>	Pol	Eng	
	multiple S-structure	+	-	+ <b><u>Earliness</u></b>
	WH-movement	+	-	

Notice now that Economy is too weak to achieve this result: it cannot derive the preference for S-structure WH-movement over LF WH-movement. In particular, a derivation involving S-structure WH-movement for a non-D-linked phrase is exactly the length of the comparable derivation involving LF WH-movement<sup>32</sup> An Economy story would have to fall back on the LF parameter in (110b).

We thus have a case that the Earliness Principle can explain, but not Economy — a case of derivations with identical numbers of steps, where nonetheless the one that “finishes first” is preferred over the other. This is exactly the sort of phenomenon we expect if there is some sort of Earliness Principle, and provides evidence for it.



## NOTES

1. Lasnik's filter (p. 162) actually states that "a morphologically realized affix must be a syntactic dependent at surface structure", where by "syntactic dependent" is meant (fn 8) B in the structure [<sub>B</sub> A B]. The subsequent literature tacitly extends this to B in [<sub>A</sub> A B] as well.

2. I will assume that Do-support inserts do in a position adjoined to INFL. Lexical insertion can then be viewed as thus an essentially transformational operation (cf. Chomsky (1965) pp. 121-123), allowing the normal case of substitution, but also allowing adjunction.

3. This solution raises, in any case, an important question which our shortcut sidesteps: namely, the question of why the "round trip" cannot be completed before S-structure, thereby in effect allowing S-structure V-to-INFL movement in English. Sidestepping a problem is not, of course, the same as solving it.

4. As in Chomsky's system, lexical insertion must have "wide scope" with respect to the Earliness Principle. Neither principle can be allowed to constraint lexical insertion, or else there would be, for example, a preference against passive morphology (which can force movement), the versions of Latin and Irish infinitivals that do not assign S-internal accusative case (forcing Raising of the subject of the infinitive; Rouveret and Vergnaud (1980); McCloskey (1986)), etc. The inclusion of D-structure in (13) thus has no empirical consequences: if Lexical Insertion is not governed by (13), then "nothing can be done" to satisfy the Earliness Principle by D-structure.

5. A third possibility, which follows from considerations explored below, is that the mappings from D-structure to S-structure and from S-structure to LF involve Move  $\alpha$ , but the mapping from S-structure to LP-structure does not.

6. I am indebted to Palmer (198x) for his thorough coverage of many of the properties of English modals, though his work does not deal directly with the problems raised below.

7. Though the factive usage of should (I am shocked that he should speak to you in this manner) may be relevant:

Note that the embedded clause may be interpreted as past with respect to the matrix (though it need not be: I am shocked that I should feel so bad). In my speech, this usage of should is more "literary" than even C-Inv, so that my intuitions are somewhat insecure.

Note as well that the usage of would found in (21b) has a habitual character missing from the should found in the protasis of counterfactuals; I have no account of this difference.

8. The be-to of obligation is an apparent counterexample, but only apparent. If we assume that (ia.) is good, and that be-to is acceptable in a counterfactual, then the fact that were cannot raise to C ((ib)) suggests that it is a  $\theta$ -assigner. But then we are surprised to find it even in INFL. That it is in INFL can be seen by its order with respect to negation ((ic)). The answer lies in the suggestion

that the be of this construction is a modal, base-generated in INFL. Its incompatibility with infinitives, seen in (ii), shows this:

- (i) a. ?If we were to take out the garbage daily, we would have been given instructions to that effect.  
b. \*Were we to take out the garbage daily, we would have been given instructions to that effect.  
c. Our instructions were that we weren't to go near the television.
- (ii) a. We are to take out the garbage daily.  
b. \*For an aristocrat to be to take out the garbage daily would be a serious insult to his dignity.  
c. We are expected to take out the garbage daily.  
d. For an aristocrat to be expected to take out the garbage daily would be a serious insult to his dignity.

9. Thus, in the terms of our speculation, where past tense is just a particular interpretation of a tenseless form, it has a tenseless form.

10. If INFL is “split” as in Pollock, we may assume that do is inserted in T and that AGR moves to provide do with its agreement morphology. The restriction of do insertion to T makes sense, since do only occurs in finite forms. I will not be making crucial reference to the “split INFL” hypothesis here, however.

11. I am using the term “ECM” to refer to Case-marking across a constituent boundary, not necessarily an IP-boundary.

12. Though one might attempt to defend a claim that ungoverned clauses are all CP, while allowing governed clauses to be IP.

13. In the next version of this draft, I will elaborate on this: basically my idea is that A-bar specifier positions are not just a default case, but are determined by the selectional properties of the heads of which they are specifiers. This kind of A-bar-selection feeds into a generalization of Burzio's idea that Case can be assigned by  $\alpha$  only if  $\alpha$  selects ( $\theta$ -marks) its subject: only instances of I that select for their subject, e.g. I containing Q, can assign Case via ECM.

14. The extension to English questions was independently proposed in a March 1988 paper by Akira Watanabe (antedating my own work). Watanabe also gives the topicalization argument presented below (attributed by him to Imanishi (1986a,b)).

15. As an aside, note that C-Inv structures — which involve full CP structure — act like embedded questions, as predicted, with respect to Topicalization:

\*This book [<sub>CP</sub> were [<sub>IP</sub> you to buy]], you would discover...

**16.** Important questions are raised about the nature of the chain: which element will count as a variable at LF, and which elements of an A-bar chain can in general bear Case. Here, the A-bar head of the chain bears Case. Perhaps this will force further A-bar movement at LF, if only a variable may bear Case in an A-bar chain.

**17.** And if INFL and the subject do not form a chain, no grammatical output is possible due to the Case filter.

**18.** One semi-argument in favor of the distinction between the two inversion processes discussed above, involves AUX-NEG contraction. Assume that AUX-NEG contraction applies at LP-structure, cliticizing NEG to INFL. Then it cannot feed I-to-C. We predict contraction in QI but not in CI, correctly:

- (i) a. Why should they not leave?  
b. Why shouldn't they leave?
- (ii) a. Should they not leave, we will get them out.  
b. \*Shouldn't they leave, we will get them out.

This a “semi-argument” at the moment, since I have no principled reason why the contraction process per se should be an LP-structure rule. Perhaps the existence of allomorphy (e.g. won't) is relevant here, but such considerations would also make many INFL+V combinations (are, were) into LP-rules — an undesirable result.

**19.** The situation is considerably more complex if the HMC derives from the ECP, as suggested in Chomsky (1986), and if the gamma-marking mechanisms assumed by Chomsky (1989) are adopted. Section 3.3 discusses a variety of alternatives that might be proposed in the spirit of Chomsky's ideas, and provides some evidence that seems to weigh against these alternatives.

**20.** A problem for my discussion is the fact that adverbs of the scarcely class, which Emonds and Zagana have noted are strictly VP-initial, do not very felicitously allow leftward verb movement of the sort we have been looking at. Intonation does make a difference, however: the adverbs improve in leftward verb-movement constructions if the verb bears focal stress and the adverb is unstressed. Note that this does not make the adverb parenthetical, since no pauses are necessary. However, it is a problem for my approach that similar verb focalization is not necessary for corresponding examples in French like those in (iv), and I have no explanation for this at present:

- (i) a. As for the dictionary, Sue barely relied on it.  
b. ?As for the dictionary, Sue relied barely on it.  
c. \*As for the dictionary, Sue relied on it barely.  
d. \*Sue's reliance barely on it surprised us.
- (ii) a. John scarcely glanced at the students.  
b. ?John glanced scarcely at the students.  
c. \*John glanced at the students scarcely.  
d. \*John's glance scarcely at the students surprised us.
- (iii) a. It was easy to get their attention. John simply shouted to them, and they came.  
b. ?It was easy to get their attention. John shouted simply to them, and they came.  
c. It was easy to get their attention. John shouted to them simply, and they came.  
d. \*John's shout simply to them did the job.
- (iv) a. Comprendre à peine l'italien après cinq ans d'étude dénote un manque de don pour les langues.  
b. Oublier presque son nom ça n'arrive pas fréquemment.  
(Pollock 1988)

(V. Depez (personal communication) notes, however, that even in French there are certain difficulties with examples like (iv)a-b: in certain cases the adverb is actually modifying the direct object; in others, verb class appears to make a difference (e.g. parler presque français vs. ??rencontrer presque Marie.)

21. These occasionally receive one or two question marks in the literature (for example, from Andrews (1983)), but I have not found any objection to them among informants.
22. This observation puts one in mind of Chomsky's identification of  $\mu$  with object agreement (AGR-O), and his suggestion that AGR-O is somehow implicated in objective Case assignment, but I do not see how to connect the two facts, and in any case the discussion below argues against identifying  $\mu$  with a morpheme of this sort.
23. Inter alia, this shows that the restrictions on movement to English INFL cannot be reduced to restrictions on movement to  $\mu$ , as claimed by Pollock.
24. Here I commit a gross oversimplification of Chomsky's approach. The theory of verb movement in Chomsky (1989) is actually a sustained argument that the HMC is only a partially correct generalization, and that those cases that seem correct follow from the ECP. In section 3.3 I deal directly with Chomsky's actual claims.
25. There is a certain parallel with Safir's (1982) claim that expletive elements may be exempt from the ECP of Chomsky (1981).
26. Another argument can be constructed on the basis of subject-oriented adverb interpretation. Assume an subject-oriented adverb takes as an argument the highest A-position in the clause to

which it attaches, and assume that a *by*-phrase is an A-position (in fact, a VP-internal subject position, as hinted at by Fukui and Speas (1986)). Consider the sequence in (i)-(iii) (deliberately totally brainwashed t<sub>i</sub> by the police). If this is structurally ambiguous between “deliberately [<sub>μP</sub> t [<sub>μ</sub> e] [<sub>VP</sub> totally brainwashed t by the police]]” and “[<sub>VP</sub> deliberately completely brainwashed t by the police]” we can explain the ambiguity. But note that the position of totally shows that brainwashed has not moved to μ.

- (i) Sue has been deliberately totally brainwashed t<sub>i</sub>  
by the police.
- (ii) deliberately- [<sub>μP</sub> t [<sub>μ</sub> e] [<sub>VP</sub> totally brainwashed t  
by the police]]
- (iii) deliberately [<sub>VP</sub> completely brainwashed t by the police]

27. This fact is even more striking when we note that the trace of have in (98a) is a “doubtless” gap, given the absence of any “hole” in the upstairs clause and the presence of past-participle morphology in the lower clause. Gaps of adjuncts like why or how are generally not “doubtless” in this way.

28. Or if LF filling of Neg<sup>o</sup> does not eliminate an index on Neg<sup>o</sup> created by S-structure substitution for that position.

29. The term “D-linked” is somewhat misleading, since no actual discourse need to have taken place. It is merely sufficient that a class of possible answers be “in the air”.

30. Though see Rudin (1989) for arguments that the phenomenon is syntactically rather different in Bulgarian and Romanian than in the other multiple-fronting languages.

31. The relevant notion of “unavailable” does not seem to include island violations in this case: when a WH-phrase is buried in a complex NP or tensed S (an island in Russian), S-structure movement is impossible, but the only LF interpretation available seems to involve D-linking. Additionally, such examples are not that easily accepted by my consultant. Thanks to Maria A. Babyonyshev and Olga Brown for the Russian data cited here, but these results are the result of very shallow questioning, and should be treated with caution at present.

32. Indeed, if those who have argued that subjacency fails to hold at LF are correct, an LF derivation might involve fewer steps than the S-structure derivation.

# Contents

1	Earliness vs. Economy	1
1.1	Economy of Derivation	1
1.2	The Earliness Principle	3
1.3	Differences	5
2	LP Structure: The Inertness of <u>do</u>	5
2.1	Argument #1 for the Inertness of <u>do</u> : Conditional Inversion (C-Inv)	5
2.1.1	<u>Non-inverted Counterfactuals</u>	6
2.1.2	<u>Counterfactuals and Past Tense</u>	7
2.1.3	<u>Do</u> in counterfactual conditionals	7
2.1.4	Inversion	8
2.1.5	Non-inversion with <u>do</u>	10
2.1.6	A Necessary Aside: Question Inversion	10
2.2	Argument #2 for the Inertness of <u>do</u> : Adverbs	14
3	Economy is Too Strong: $\mu$ or AGR?	16
3.1	Movement to $\mu$	17
3.1.1	<b>Prolegomena to main verb movement to <math>\mu</math> in English</b>	17
3.1.2	<b>Evidence for main verb movement to <math>\mu</math> in English</b>	18
3.2	Consequences of English Main Verb Movement to $\mu$	21
3.3	<b><u>Alternatives to the argument that <math>\mu</math> is not an affixal position.</u></b>	23
3.4	Summary	28
4	Economy is Too Weak: D-linking and WH-movement	28
	NOTES	33