TOWARDS SINGLE OUTPUT SYNTAX

HIROSHI TERADA
Yamaguchi University

The purpose of this paper is to provide support for the Single Output Hypothesis (SOH), arguing against Chomsky’s (1995: Chapter 4) view that weak features attract formal features after Spell-Out. The SOH advanced here is implemented in such a way that movement operations triggered by either strong or weak features take place before Spell-Out. Strong features require pied-piping of a whole category, while weak features attract formal features. This hypothesis makes it possible to improve Chomsky’s proposals surrounding strict cyclicity, derivational c-command, immobility of traces, adjacency effects, and Q-insertion. It also weakens the motivations of certain global economy conditions.

1. Introduction

In Chomsky’s (1995: Chapter 4) version of the minimalist program, although the computation from a numeration N to an LF (logical form) representation \( \lambda \) (hereafter, the \( N \rightarrow \lambda \) computation) is assumed to be uniform, the system introduces some asymmetries into the syntactic computation: pre-Spell-Out operations are claimed to be more costly than post-Spell-Out ones; strict cyclicity is not observed in cases of post-Spell-Out feature-attraction; and the lexical access is to a large extent confined to overt syntax (the only exception is post-Spell-Out merger of a phonologically null Q-feature). This paper attempts to eliminate these asymmetries by endorsing the Single Output Hypothesis (SOH), adumbrated by Bobaljik (1995), Groat and O’Neil (1996) and Holmberg (1997) among others. (See also Brody (1995) for related ideas within a different approach.)

* I would like to thank Prof. Heizo Nakajima and two anonymous EL reviewers for their invaluable comments and suggestions. I am also grateful to my informants, Claire-Nelly Suematsu, Paul Snookes, and Henry Atmore for their help. Needless to say, I am solely responsible for any remaining inadequacies.

© 1998 by the English Linguistic Society of Japan
It is hypothesized within Chomsky’s (1995: Chapter 4) framework that syntactic movement operations occur before and after Spell-Out, which is the split that takes an input and leads to separate PF (phonological form) and LF representations (or, π and λ):

(1) Chomsky (1995: Chapter 4)

Contrary to Chomsky, Bobaljik (1995) posits a model of grammar where the syntax produces a single output representation from a given input, and the output is interpreted by λ and π. Spell-Out is applied after LF-operations, so that there is no syntactic operation after Spell-Out:

(2) Single Output Hypothesis

Hence, both overt and covert operations take place before Spell-Out. What distinguishes the former from the latter is the following condition on pronunciation:

(3) Pronunciation of the highest copy corresponds to “overt” movement and pronunciation of a lower copy corresponds to “covert” movement. (Bobaljik (1995: 354))

In other words, the top copy of an element raised by a strong feature (hereafter, such movement is referred to as overt movement) is pronounced at PF, while the bottom copy of an element raised by a weak feature (hereafter, covert movement) is pronounced.

An alternative implementation of the SOH is provided in Groat and O’Neil (1996: 114), who assume the following:

(4) Overt movement is movement which carries phonological features to the head of the chain being created, while covert movement leaves phonological material behind.

What motivates this assumption is the consideration that strong features
share some properties with affixes, in that they can be licensed only if some phonological material serves as a host. Weak features neither bear this character nor require a phonological host (p. 124). (As for phonological hosts to affixes, see Bobaljik (1995).)

A little thought should show that Groat and O’Neil’s way of implementing the SOH is preferable to Bobaljik’s pronunciation account. Within the latter approach, even weak features raise phonological features, which are not pronounced in the raised position. Obviously, pied-piping of phonological features in such cases counts as an entirely superfluous operation. Furthermore, although (3) requires the top copy to be pronounced for the checking of strong features, these are erased before Spell-Out and hence do not enter the phonological component, where it is not detectable whether the top copy or bottom copy of a chain is pronounced.

Despite its superiority to (3), Groat and O’Neil’s assumption (4) is not without its problems. A statement that a strong feature attracts phonological features will not make sense, because feature-checking has nothing to do with phonological features. For this reason, we will seek another way to maintain Groat and O’Neil’s original intuition that strong features share some properties with affixes. It is plausible to assume that affixes are in general satisfied not by a mere grammatical feature, but by a word(-like element) or a full category (cf. Lasnik (1990: Chapter 5)). Strong features are largely restricted to categorial features (cf. Chomsky (1995: 232)), so that they attract full categories (cf. Brody (1997)). If this is plausible, let us rephrase (4) as something like (5):

\begin{equation}
(5) \text{A strong feature, being affixal in nature, requires a full category; A weak feature only raises formal features.}
\end{equation}

Suppose that FF(V) (a set of formal features of V) rather than V is attached to $\nu^{\text{topmax}}$ (=the zero-level projection of $\nu$) to check its strong V-feature, forming a structure $[\nu^{\text{topmax}} \text{FF(V)} \nu]$. (5) is not met: FF(V) cannot project into a V-head (cf. Chomsky (1995: 270)); Otherwise it would head a chain $\text{CH} = (V^0, t_{\text{FF(V)}})$, which violates the Uniformity Condition (6) (see (ibid.: 253)):

\begin{equation}
(6) \text{A chain is uniform with regard to phrase structure status.}
\end{equation}

Then a full category is adjoined to $\nu$ to satisfy the affixal nature.

Let us next suppose that FF(\text{the man}) rather than $[\text{DP the man}]$ is attached to $T^{\text{topmax}}$ to satisfy its strong EPP (extended projection principle)-feature (or D-feature), forming $[T^{\text{topmax}} \text{FF(\text{the man})} T]$. This
feature requires a full category. Note that FF(*the man*) can project neither a D°-head nor nonminimal DP; otherwise the Uniformity Condition would be violated. In this case a full category DP rather than FF(*the man*) raises to [Spec, T]. DP cannot adjoin to T because of the following morphological property (see (ibid.: 319)):

(7) Morphology deals only with X° categories and their features.

A question naturally arises here: why is it that a strong D-feature overtly raises a whole DP rather than a D-head, whereas a strong V-feature raises a V-head rather than a whole VP? If such pied-piping were attributed to PF-convergence, as in Chomsky (1995: 261ff.), then the application of Move would rest on information that is available only at PF, not at the stage where Move is applied. Such a global approach should not be pursued. Instead, we hope that the pied-piping phenomena in question will be somehow derived from a set of constraints on extraction, i.e. the HMC (Head Movement Constraint), the LBC (Left-Branch Condition) and the CED (Condition on Extraction Domain) among others (see Chomsky (1995: 49, 263, 328), for some discussion). Thus, pied-piping of, say, [DP *whose mother*] rather than *whose* alone is required by the CED or the LBC; raising of [DP *the picture of John*] rather than [D *the*] alone from within VP to the checking domain of T is required by the CED or the HMC; raising of V (rather than VP) to v, for example, is allowed under the HMC. For these reasons, a strong feature requiring a category D attracts a whole phrase DP. Note furthermore that a set of constraints on extraction might be incorporated into the definition of Attract/Move, though the attraction of formal features will be immune from such constraints for reasons to be clarified (cf. Chomsky (1995: 307)). If this move is on the right track, then whether to pied-pipe a whole phrase or its FF is determined not on PF-convergence, but on the basis of these syntactic constraints. This largely reduces computational complexity, although we are not in a position to elucidate the exact status of such con-

---

1 Within Chomsky’s framework a strong feature raises not simply a set of formal features but “whatever extra baggage is required for PF-convergence”: isolated features and other scattered parts of words may not be subject to the morphophonological rules, which causes the cancel of the derivation or a violation of FI (full interpretation) (pp. 262, 263). Within the present system, such reasoning is not maintained, since formal features and the other parts of a word should be necessarily separated before Spell-Out.
A weak feature attracts FF(LI) of a lexical item, rather than its whole category. The idea that movement of a full category is taken to be more costly than that of FF(LI) alone is also motivated by economy considerations (see (22) below).

Before giving conceptual and empirical support to the SOH, let us answer the question as to why there is no syntactic operation after Spell-Out, or why Spell-Out is applied only after all syntactic operations are completed. Within Chomsky's (1995: Chapter 4) approach, it appears that Spell-Out is freely applied at any point of a derivation. In what follows I will provide several arguments in favor of our proposal.

First, as we will see in section 2.1, strict cyclicity is not observed if post-Spell-Out feature-attraction is carried out: a set of formal features is adjoined to a non-root position. As a result, the asymmetry before and after Spell-Out conflicts with the uniformity of the N→λ computation. Under the SOH there is no post-Spell-Out syntactic operation and all syntactic operations are applied in accordance with strict cyclicity. Thereby the uniformity is ensured.

The second argument concerns feature-mismatch. We keep to Chomsky's (1995: 309) assumption that the mismatch of features cancels the derivation. Suppose that a nominative object is generated within VP in a language like English where the accusative Case-feature of V is weak. In Chomsky's system FF(Obj) raises to [\textit{\text{v}}_{\text{max}} \text{FF(V-\nu)} \text{T}] after Spell-Out. Then its nominative Case-feature mismatches the accusative Case-feature of V; it is not until this point that the derivation is canceled. In this derivation Spell-Out has already stripped away those elements relevant to \textit{\text{v}} from the structure formed before Spell-Out, although this expression is not pronounced. This would be a superfluous step from the minimalist viewpoint.

In contrast, under the SOH all feature-checking operations are restricted to overt syntax. Therefore, any mismatch of features cancels the derivation before Spell-Out, which makes it possible to preclude stripping away phonological features etc. This reduces much computational complexity.

Third, it appears as if in Chomsky's framework Spell-Out would be applicable at any point of a derivation, although severe restrictions are actually imposed on its application. Note that if an item bearing no phonological features (say, a null Q-feature) were inserted into a struc-
ture before Spell-Out, then this operation would have no PF-effect and be barred under Chomsky's (1995: 294) economy condition as follows:

\[(8) \quad \alpha \text{ enters the numeration only if it has an effect on output.}\]

Then, Spell-Out should be applied before this element is Selected. In short, whenever Spell-Out is applied we would have to see if the application would not violate (8). Since economy (8) is rather global, Spell-Out conditioned by it will lead to computational complexity.

By contrast, under the present assumptions Spell-Out can be characterized as the operation applied during the course of a derivation when the structure removes illegitimate objects. We therefore need not see whether its application might lead to an economy-violation. As long as any item remains in a numeration, Spell-Out is not applicable by definition. If a numeration is not exhausted no derivation is generated; by this token, if Merge is insufficient no derivation is generated either (see Chomsky (1995: 226)). Then all we have to see is whether a derivation generated contains any illegitimate object (i.e. Uninterpretable feature) or not. Not conditioned by global economy (8), the application of Spell-Out under the SOH does not cause computational complexity.

If the above rationales are correct, we can safely conclude that there is no post-Spell-Out syntactic operation.

In sum, feature-movement, whether it is triggered by a strong feature or a weak one, takes place before Spell-Out. Thus, a syntactic derivation consists of Merge and Attract/Move, terminating in a single final phrase marker $\Sigma$ which is the input to the morphophonological component and LF. Spell-Out strips away from $\Sigma$ those elements relevant to $\pi$, leaving the residue $\Sigma_L$, which qualifies as $\lambda$. There is then no post-Spell-Out syntactic operation prior to $\lambda$.\(^2\)

In what follows, I will provide several arguments for the SOH. First, by introducing the Earliness Principle, the SOH can eliminate

\(^2\) It is possibly the case that instances of Merge driven by a strong feature are overt, whereas merger of empty categories is executed in order to satisfy weak features. Moreover, although Chomsky (1995: 292–294) suggests that insertion of features lacking phonological features is a post-Spell-Out operation, it may well be executed before Spell-Out in the present framework, as we will see below. Whatever feature attracts (formal features of) null operators and empty pronouns (i.e. pro and PRO) might also be weak rather than strong (see Takahashi (1997) and Groat and O’Neil (1996: 125) for discussion). We will not pursue these matters any further here.
Chomsky's (1995b: 233) stipulation that only a strong feature can induce strict cyclicity (section 2.1). In this connection, the present approach makes it possible to define c-command as an explicable derivational construct, rather than an unexplicable representational notion (section 2.2). Second, given (5) we can (almost) dispense with a global economy condition, Procrastinate (section 3). By so doing, the present proposal can reduce the above-mentioned asymmetries between pre- and post-Spell-Out operations. Third, we can confirm Chomsky's (1995: 304) claim that only the head of a chain enters into the operation Attract/Move (section 4). The SOH lends further support to Chomsky's treatment of adjacency effects (section 5). It also detracts from the global economy condition with which Chomsky attempts to bar post-Spell-Out insertion of strong features (section 6). Section 7 suggests that Collins' (1997) argument against the SOH is not maintained. Section 8 concludes this paper.

2. Strict Cyclicity

2.1. Observance of Strict Cyclicity

In this section, we will show that the SOH permits us to dispense with a stipulation that post-Spell-Out derivations as opposed to pre-Spell-Out ones do not obey strict cyclicity, thereby ensuring the uniformity of the N→λ computation.

Within Chomsky's (1995: Chapter 4) framework, the derivation of a sentence like *we build airplanes* proceeds as follows:

(9) [TP we T [TP vP [VP tV build v]] [VP tV airplanes]]

a. raising of *build* to *v*;  
b. merger of Subj *we* and vP;  
c. merger of T and vP;  
d. raising of *we* to [Spec, T]

(10) [TP we [T^0max FF(airplanes) [T FF(Vb) T]] [vP tV [VP tV airplanes]]]

a. raising of FF(Vb) to T^0max;  
b. raising of FF(Obj) to T^0max

(9) is a pre-Spell-Out derivation, where Vb is a complex verb (i.e. [Vb build v]), while (10) is a post-Spell-Out derivation. Note that Chomsky's framework necessarily yields a countercyclic derivation in the covert component. FF(Vb) and FF(Obj) can potentially attach to non-root categories. Then it follows that post-Spell-Out movement is not restricted to a root position.

In contrast, the present framework forces both *overt* and *covert*
movements to take place before Spell-Out so that they rigidly observe strict cyclicity. Consider an alternative derivation as in (11a-f):³

(11) a. \[\text{vp} \text{v} \text{[vp build airplanes]}\]
b. \[\text{vp} \text{[vb build v]} \text{[vp tv airplanes]}\]
c. \[\text{vp} \text{[vb FF(airplanes) [vb build v]} \text{[vp tv airplanes]}\]
d. \[\text{vp} \text{we [v' [vb FF(airplanes) [vb build v]} \text{[vp tv airplanes]}\]
e. \[\text{tp T [vp we [vb FF(airplanes) [vb build v]} \text{[vp tv airplanes]}\]
f. \[\text{tp we [t_{max} FF(Vb) T] [vp t_{we} [vb FF(airplanes) [vb build v]} \text{[vp tv airplanes]}\]

After \(v\) is merged with VP in (11a), the strong V-feature of \(v\) attracts a full category V. This forms a complex verb Vb as in (11b). Then the weak Case-feature of Vb attracts FF(airplanes), as in (11c), to complete object Case-checking. Given that substantial categories lack the checking ability, attraction of FF(Obj) may be postponed until V is raised to v. We is inserted into the Spec of v, as in (11d). After merger of T and vP, the tense-feature of Vb adjoins to T in (11f), which is possibly driven by the weak V-feature of T in English (as suggested in Chomsky (1995: 269, 361) and Collins (1997: 117)). Then the strong EPP-feature attracts a full category viz. subject DP. At this point, (i) the strong EPP-feature and the weak nominative Case- and \(\phi\)-features of T, (ii) the weak \(\phi\)-features of FF(Vb) and (iii) the Case- and

³ We adopt the following definition of Attract and closeness:

( i ) \(K\) attracts \(F\) if \(F\) is the closest feature that can enter into a checking relation with a feature of the zero-level projection \(H(K)^{\text{imax}}\) of the head \(H\) of \(K\).

(Chomsky (1995: 268, 297))

( ii ) \(\beta\) is closer to the target \(K\) than \(\alpha\) if \(\beta\) c-commands \(\alpha\).

(Chomsky (1995: 368))

Given (i)-(ii), overt object-shift to the Spec of vP should precede merger of the subject with vP; otherwise, the subject would be closer to the target than the object so that the latter could not be shifted.

An EL reviewer points out that elimination of the equidistance clause from (ii) leaves us with a question of how locative inversion is accounted for, if Collins (1997) is correct in claiming that equidistance of the theme and locative PP is needed. Note, however, that Levin and Rappaport Hovav (1995) (L&RH) describe unergative verbs such as work, chatter, glitter, and rumble as also being attested in this construction. The subject of an unergative verb, generated in [Spec, v] (cf. Chomsky (1995: 315, 316)) and locative PP, generated within VP (cf. L&RH, pp. 261–263), would not be in the same minimal domain of a head. Then, the MLC would bar raising of the PP across the subject. Accordingly, even an equidistance-based account does not work on this construction. Since this issue is far beyond the scope of this paper, we will leave it open here. See L&RH, for discussion.
\(\phi\)-features of we enter into a checking relation. (Here, we adopt Chomsky's (class lectures, Fall 1995) insight, according to which in languages such as English the checking of \(\phi\)-features is carried out together with that of the nominative Case-feature of T (cf. Ura (1996)). This is the reason why the \(\phi\)-features of FF(Vb) cannot enter into a checking relation with FF(Obj) in (11c).) Since FF(Obj) lacks phonological content, the resultant sentence will be linearized as we build airplanes, as desired. Accordingly, the present system unlike Chomsky's, rigidly observes strict cyclicity throughout a derivation.

One might claim that within Chomsky's framework, post-Spell-Out operations necessarily involve a single phrase marker given requirements of FI etc., so that if an embedded head contains an Uninterpretable feature, a countercyclic operation must take place in order to ensure convergence. If the absence of strict cyclicity of post-Spell-Out operations followed from FI, however, the following problems would necessarily arise. First, such reasoning would necessarily imply that a countercyclic derivation for convergence could be incorrectly tolerated even before Spell-Out; otherwise the uniformity of the N\(\rightarrow\lambda\) computation should be abandoned. It is questionable to assume that post-Spell-Out operations can be countercyclic because of FI.

Moreover, there is empirical evidence that movement driven by a weak feature is subject to strict cyclicity. Consider first the following \(wh\)-island effects in Japanese:

(12) a. (kimi-wa) [dare-ga nani-o katta ka] sitteiru ka?
you-topic who what bought Q know Q
b. Do you know who bought what?
c. (?)Who do you know bought what?
d. Not: What do you know who bought?

Both dare and nani take their scope within the embedded clause, so that (12a) can be interpreted as a yes-or-no question (12b). Some speakers allow for (12c), according to Nishigauchi (1990: 33–40). Note that nani in (12a) cannot extend scope over dare; (12d) is not its correct interpretation. With this in mind, consider the following:

(13) a. (Kimi-wa) [Bill-ga nani-o tabeta ka] oboeteiru ka?
you-topic Bill what ate Q remember Q
b. Do you remember what Bill ate?
c. Not: What do you remember whether Bill ate?

The interrogative complementizer \(ka\) is capable of constituting a yes-or-no question in the absence of a \(wh\)-phrase (cf. Nishigauchi (1990: 30)).
However, (13a) cannot be interpreted as (13c): embedded *ka* does not introduce a yes-or-no question. Rather, (13a) is unambiguously interpreted as (13b).

Let us examine what happens if movement involved in (13a) were applied in a countercyclic fashion. To this end, let us propose that in Japanese Q-features are weak and Uninterpretable. This proposal will be necessary, since the *wh*-feature of a *wh*-phrase is Interpretable and need not raise; however, as Fujita (1996) suggests, no *wh*-island effect as in (12d) would be yielded if *wh*-features were not raised in this language. If the Q-feature of interrogative C were strong, then not only a *wh*-feature but also its whole *wh*-phrase would be overtly raised, which is not the case with *wh*-in-situ languages. Suppose that this proposal is correct. Consider the structure given to (13a):

(14) (Kimi-wa) [Bill-ga nani-o tabeta ka(=Q)] oboeteiru ka(=Q)

If post-Spell-Out operations could be countercyclic, matrix Q would attract FF(*nani*), so that it would introduce a *wh*-question. In the absence of a *wh*-phrase, embedded Q introduces a yes-or-no question. Then, (13a) would be incorrectly interpreted as (13c). Under the present hypothesis, on the other hand, feature-raising is subject to strict cyclicity. Embedded Q attracts FF(*nani*). Thereby the *wh*-phrase takes embedded scope. Then matrix Q, in the absence of a *wh*-phrase, introduces a yes-or-no question. This correctly yields interpretation (13b).4

If this discussion is tenable, movement driven by a weak feature also obeys strict cyclicity: countercyclic derivations will wildly overgenerate. We will discuss below that the derivational c-command in section 2.2 and the adjacency effects as in (36a) and (38) below provide additional empirical evidence for cyclicity of covert movement.

It should be noted here that Chomsky’s (1995: 234) implementation of strict cyclicity crucially rests on feature-strength:

4 Possibly, Q-features of yes-or-no questions are Interpretative.

If FF(*nani*), once raised to embedded Q, were further raised to matrix Q, then the embedded clause would not count as a yes-or-no question, since embedded Q as a *wh*-question has undergone checking. The derivation will converge with possibly a gibberish interpretation, just on a par with the following example:

(i) (guess) which book [they remember [if Q [John gave it to Mary]]]  
(Chomsky (1995: 291))

In (i) *which book* is first raised to check embedded Q, and then raised to check the higher Q. The resultant interpretation will be gibberish.
Given that the derivation D has formed Σ containing a with a strong feature F, D is canceled if a is in a category not headed by a.

This is a stipulation: we cannot answer the question of why only a strong feature must be eliminated before it becomes part of a larger structure by further operations. The present approach allows us to formulate strict cyclicity as in (16), making no reference to feature-strength.5

(16) Strict Cycle Condition
Given that the derivation D has formed Σ containing a with an Uninterpretable feature F of a target, D is canceled if a is in a category not headed by a.

A derivation as in (11) observes strict cyclicity: Uninterpretable features are checked off before embedded in a larger structure. (In this case a root might be understood as a category not embedded in a larger category. If so, even head-movement obeys the strict cycle, since it is safe to say that a head adjoins to a root category.) (16) is superior to (15), in that it will be derived from either a more general principle, i.e. Earliness, or the general property of the computation for human language (CHL) that proceeds to remove illegitimate objects such as Uninterpretable features of a target. In what follows, I will explore the first possibility.

It is worth emphasizing that though not explicitly mentioned, Chomsky's system draws a distinction between Uninterpretable features: those Uninterpretable features that have the ability to attract other formal features (call them Attractors) and those which lack this ability (call them Non-attractors). The former include features of targets (e.g. the Case-features of Vb and T; the D-feature of T, and φ-features of A, etc.), while the latter include others (e.g. the Case-feature of DP, the φ-features of V) which do not attract any formal feature to be checked off. By elucidating this implicit distinction, let us propose the Earli-
ness Principle, an economy condition originated in Pesetsky (1989) and developed in Brody (1995) and Lee (1997): 6

(17) Earliness

A derivation must remove illegitimate objects as early as possible.

Given Earliness, an Attractor must be checked off before it is embedded in a higher structure. This is the earliest possible stage of the derivation in which Attractors can be erased. On the other hand, Non-attractors could never be checked off by attracting other formal features, their checking is postponed until they are attracted. Still, this is the earliest stage.

It is in order here to show that Earliness is preferable to Procrastinate. First, Chomsky’s requirement (15) is in a sense an Earliness requirement: a strong feature has to be erased as early as possible. However, on the other hand, he assumes Procrastinate: checking is postponed as late as possible. It is obviously problematic that these contradictory requirements coexist within a single system. We do not invoke Procrastinate (cf. section 3), so that no such contradiction is involved in our system.

Second, Earliness is a local economy condition: only by discerning whether an Uninterpretable feature is embedded in a larger structure or not is it detectable whether a checking operation has been applied at the earliest point or not. Chomsky’s system adopts global economy, since one would have to see a whole derivation to judge whether checking is applied at the latest possible point or not.

To sum up, both overt and covert movements target a root in accordance with the Strict Cycle Condition or the Earliness Principle, and are invariably cyclic. In our framework we can avoid the unwanted situation which the post-Spell-Out feature-movement approach could not: strict cyclicity is required only before Spell-Out. This is a desideratum; otherwise we would have to abandon the minimalist assumption that the \( \text{N} \rightarrow \lambda \) computation is uniform.

2.2. Derivational C-command

We will suggest in this section that insofar as Epstein’s (1995) deriva-

---

6 It seems that Earliness is inviolable, so that it might be incorporated into the definition of Attract.
tional view of c-command is correct, Chomsky's post-Spell-Out feature-movement approach may not be maintained.

Although the notion of c-command plays a significant role in defining rudimental concepts, before Epstein's (1995) study this notion had been simply stipulated and defined on syntactic representations, not defined derivationally. Assuming that all syntactic relations are formally expressed by the operation Concatenate A and B forming C, Epstein (1995: 16) proposes the notion of derivational c-command:

\[(18) X \text{ c-commands all and only the terms of the category } Y \text{ with which } X \text{ was Merged in the course of the derivation.}\]

In short, c-command is nothing but the relation that holds between X and elements of Y when X is attached to Y by Merge or Move.

Let us examine what potential problems would arise with post-Spell-Out movement. Chomsky (1995: 254) seems to imply that one can reduce strict cyclicity to Epstein's derivational c-command (for a somehow similar proposal, see Collins (1997: 81–89)). Thus, he suggests that if $\alpha$ is attached to an embedded non-root category by an operation, $\alpha$ will enter into no c-command relation with any higher element $\beta$. The countercyclic derivation as seen in (10) above actually runs counter to this suggestion, however. Thus, a covertly raised element such as FF(airplanes) would never be c-commanded by such an element as the subject we. Were this the case, an anaphor raised after Spell-Out would not be c-commanded by its antecedent:

\[(19) \begin{align*} 
  \text{a. They seemed to each other [t to have been angry].} \\
  \text{b. *There seem to each other [t to have been many linguists} \\
  \text{given good job offers].} 
\end{align*}\]  

Chomsky (1995: 275) claims that in (19b) both FF(associate) and (the formal feature of) the anaphor (hereafter, FF(anaphor)) are covertly attached to matrix T to form an illegitimate binding-theoretic configuration. In order for this account to go through, FF(anaphor) should also be attached to matrix T in (19a). Still, given Epstein's derivational c-command, the raised FF(anaphor) will not enter into a c-command relation with any higher element, especially with its antecedent they. This would incorrectly yield an illegitimate binding-theoretic configuration of (19a).

A similar problem would arise with wh-in-situ. In a language with no overt wh-movement, an interpretive strategy like unselective binding is employed so that a wh-phrase in situ, serving as a variable, can be bound by a weak Q-feature (see section 6). In languages like English
weak $Q$ is an Interpretable feature, so that it does not trigger overt $wh$-movement. Languages like Chinese would roughly have a structure like (20a) before Spell-Out and (20b) at LF:

(20)  
   a. We know $[Q [\text{John } T \text{ loves whom}]]$
   b. We know $[Q [\text{John } [T_{\text{omax}} \text{ FF(whom)}-\text{FF(loves)}-T] \text{ loves whom}]]$

Under Chomsky's (1995: 4.5.4) assumptions, an embedded weak $Q$-feature should be inserted into an embedded question only before the matrix clause is constructed: Merge is restricted to the root. Suppose that the Case-feature of $V$ is weak. Then the Case-feature of the $wh$-phrase in situ is attached to embedded $T$ after Spell-Out, carrying along with its all the formal features including its $wh$-feature (cf. Chomsky (1995: 291)). Then, FF($whom$) attached to the embedded $T$ will not enter into a c-command relation with any higher element, especially with the weak $Q$-feature. If so, its $wh$-feature now under the embedded $T_{\text{omax}}$ would fail to be interpreted as a variable bound by the $Q$-feature. This would probably end up as an illegitimate LF-object, violating FI. Then the derivation of (20) would crash, contrary to fact.7

7 A reviewer suggests a possibility that $Q$ unselectively binds $whom$ in the base-position rather than the raised FF($whom$). It might be impossible, however, given Chomsky's (1995: 300-304) assumption that after FF-raising, erasure of the formal features of a trace is obligatory under the principle (28) below.

The other reviewer points out a possibility that in (19)–(20), FF raised in the covert component could have already established a c-command relation in overt syntax. If this were adopted, the LF interpretive rules such as Binding Principles would apply not only to an LF-output but also to a pre-Spell-Out structure, which is an unwanted situation. An alternative to this would be that a c-command relation is regarded as some form of marking (like *-marking proposed in Chomsky (1995: Chapter 1)). This is obviously a departure from the inclusiveness condition which requires that outputs consist of nothing beyond properties of items of the lexicon (cf. ibid.: 225).

Furthermore, if a c-command relation established at a certain stage of a derivation is maintained throughout the derivation, reconstruction effects as in (i) would be unexplained.

(i) Which picture of John did Mary say he saw t? (Marantz (1995: 374))

The reviewer further claims that (19b) might be accounted for without resorting to feature-movement, since the requisite c-command relation does not hold before Spell-Out. This raises the question about binding in exceptional Case-marking (ECM) contexts, as the reviewer suggests. The following example cited from Chomsky (1995: 274) also indicates the necessity of feature-attraction:

(ii) There arrived with their own books three men from England.
It is worth noting at this point that the examples (19) and (20) motivate that movement triggered by a weak feature obeys strict cyclicity in the sense that movement of FF(anaphor) and FF(what) in these examples should be applied before their respective target is c-commanded by a higher element. Thus, the derivational view of c-command provides a further empirical argument for the claim in section 2.1.

Insofar as feature-attraction is executed after Spell-Out, Chomsky’s framework fails to incorporate Epstein’s insight, leaving a c-command relation as a stipulative notion, defined solely in representational terms. Under the SOH, by contrast, each feature attracted in a strictly cyclic fashion successfully enters into a c-command relation with its higher element. This allows us to keep to the minimalist view that $C_{HL}$ is strictly derivational (cf. Chomsky (1995: 223, 224)).

3. Elimination of Procrastinate

This section is allotted to the discussion that given assumption (5), the motivation of Procrastinate, a global economy condition, is largely weakened.

Procrastinate states that overt operations are more costly than covert ones. For this reason, it has been assumed that whenever possible, movement operations are postponed until the covert component. This, however, would detract from the uniformity of the $N \rightarrow \lambda$ computation. Chomsky (1995: 264) attempts to provide a rationale (which originated with Hisatsugu Kitahara and Howard Lasnik) for Procrastinate, with recourse to the economy principle (21):

\[(21) \text{ F carries along just enough material for convergence.}\]

(ibid.: 262)

Under Chomsky’s assumption pre-Spell-Out movement raises phonological and semantic features, which is possible only if otherwise the derivation would not converge at PF. Because of (21), post-Spell-Out feature-movement, which does not raise non-formal features, is cheaper than pre-Spell-Out raising. Then, for Move to be applied at a certain stage of a derivation, it is necessary to know whether the derivation crashes at PF or not. In other words, the application of Move rests on information that is available only at PF, not at the stage where Move is applied. Therefore, Procrastinate counts as a global economy condition in character (cf. Collins (1997), for discussion).

Procrastinate has another inadequacy: unlike other economy princi-
ples, it is violable (cf. Collins (1997: 116)). Last Resort, the Minimal Link Condition (MLC) and economy conditions like (8) and (21) cannot be violated (for detailed discussion, see Chomsky (1995: Chapter 4)). One might claim that economy conditions may well be violable for convergence. That Procrastinate is exceptionally violable unlike other economy conditions is not explained but simply stipulated in Chomsky's framework. Further, violability would invoke a global notion of economy again, another departure from minimalism.8

Under the SOH, effects of Procrastinate follow (at least partially) from Groat and O'Neil's economy condition:9

(22) Minimize the number of features carried along.10

(22) is a local economy condition: only by discerning whether the target has a strong feature is it detectable at a certain point of a derivation whether a full category can be carried along or not: pied-piping of a whole category is required if the target has a strong feature. Unlike Chomsky (1995: Chapter 4), the present system does not assume that pied-piping of a whole category rests on PF-convergence; rather, pied-piping takes place for purely syntactic reasons (see section 1). If the target has a weak feature, formal features alone are raised under (22). Thus, (22) derives the effects of Procrastinate: overt movement is more costly than covert movement.

Let us turn to expletive constructions. Chomsky's (1995: 344–346) treatment of the contrast in (23a, b) is contingent on Procrastinate:

(23) a. There seems to be someone in the room.

b. *There seems someone to be in the room.

At some stage, we have the following structure common to (23a, b).

---

8 A reviewer suggests that the shortest derivation condition (SDC) is also violable. It is questionable, however, that the SDC actually exists. Chomsky (1995: 315) has introduced this condition in accommodating that a violation of the θ-Criterion causes the derivation to crash. However, if the θ-Criterion is ultimately reduced to FI, the SDC is dispensable: FI determines a set of convergent derivations and economy considerations hold only among convergent derivations (p. 220).

9 We do not adopt (21): it is also a global economy condition which refers to convergence.

10 (22) needs further elaboration, if we are confronted with the fact that pied-piping is optional (cf. Which piece of paper did Betty draw on? On which piece of paper did Betty draw (Collins (1997: 120)). We will leave this matter unsolved in this paper.
The next step is to fill [Spec, T] in (24) to satisfy the strong EPP-feature of T. When the numeration N includes expletive *there*, we have two possibilities: either *someone* can raise to [Spec, T], or *there* may be Selected from the N and merged into [Spec, T]. It is hypothesized that the former choice violates Procrastinate while the latter does not, so that *there*-insertion is chosen as in (25):

(25)  \[\text{TP there to be [someone in the room]}\]

At a later stage in the derivation, we obtain (26):

(26)  \[\text{TP T seems [TP there to be [someone in the room]]}\]

The only legitimate operation to satisfy the EPP-feature of matrix T is to raise *there*, forming (23a). Raising of *someone* to matrix [Spec, T] will be barred by the MLC.

This explanation involves at the very least two problems. First, it is crucially contingent on the global economy condition, Procrastinate. The second problem concerns self-contradiction in Chomsky’s framework. He claims that Procrastinate applies not only to Attract/Move but also to Merge (p. 292). If we take this claim seriously, we are led to conclude that either option applied to (24) would violate Procrastinate: the merger of *there* before Spell-Out results in a forced violation of this principle to satisfy the strong EPP-feature of T, as much as the raising of *someone* before Spell-Out. In other words, as far as Procrastinate is concerned, Merge and Move are equally costly.

Chomsky (1995: 226, 235) assumes that Merge is chosen over Move whenever either option is available. Shima (1997) and Lee (1997) provide arguments to the contrary. Thus, apart from the stipulation (15) strict cycle effects would not be derived in Chomsky’s system, simply because at any stage of a derivation movement for checking is over-ridden by merger. Given these arguments, Chomsky’s solution to (23a, b) is controversial.

Note that this contrast does not fall under the economy condition (22), which says nothing of the comparison between Merge and Move. I suggest that there is an additional principle of economy that captures the intuition that merger of expletive *there* is preferred to raising of the associate. We adopt Poole’s (1996: 39) principle Total Checking (hereafter, TC):\(^{11}\)

\(^{11}\) We should impose a premise on this principle: the MLC outranks TC in that
(27) An operation resulting in total checking is more economical. In other words, given two or more possible operations $\alpha_1, \ldots, \alpha_n$ at a given point in a derivation, the most economical operation is the one which enables all of the formal features of the moved or merged element to enter into checking relations. The contrast in (23a, b) is immediately derived from TC. Let us now consider (24), repeated here:

(24) [[TP to be [someone in the room]]

For [Spec, T] in (24) to be filled, it is necessary to consider the features checked in this position. Someone bears Case-, $\phi$-, and D-features, whereas expletive there bears no formal feature other than the D-feature. Non-finite T possesses only a strong EPP-feature, and hence insertion of there is more economical than raising of someone, since the latter bears Case- and $\phi$-features that will not enter into a checking relation in the [Spec, T] position. There can have its D-feature checked against T and this results in total checking.

In this section, we have discussed the inadequacy of Procrastinate, suggesting that it should be superseded by economy conditions such as (22) and TC, although questions surrounding unforced violations of Procrastinate (cf. Chomsky (1995) and Ura (1996)) are left open here.

4. Immobility of Trace

The present approach gives further support to Chomsky’s (1995: 304) conclusion that there is no trace-movement.

Let us first review how Chomsky is led to this conclusion. In a sentence like we are likely $t_3$ to be asked $t_2$ to $t_1$ build airplanes, we have three possible A-chains created by successive cyclic movement of we: CH$_1$=(t$_2$, t$_1$), CH$_2$=(t$_3$, t$_1$), and CH$_3$=(we, t$_1$). Nevertheless, among them only CH$_3$ can serve as a legitimate LF-object, with the Uninterpretable Case-feature erased from t$_1$. It follows that CH$_1$ and CH$_2$ should be eliminated and only CH$_3$ survive; otherwise, the derivation would crash. Then the formal features of the intermediate traces of the MLC cannot be overridden in order for TC to be satisfied. We refer the reader to Poole (1996), for an argument that TC is superior to its alternative condition which states that a derivation that results in checking of more features (rather than total checking) is more economical. (See also Fujita (1997), for related discussion.)
A-chains, which are deleted, should be allowed to be erased if possible. Actually, they must be erased to satisfy the following principle (for detailed discussion, see Chomsky (1995: 300–304)):

(28) Deleted $a$ is erased when possible. (Chomsky (1995: 280))

Chomsky then seeks to extend this argument to traces of A-movement in general. The formal features of A-traces must be erased, so that they are inaccessible to Move.

Let us now turn to traces of A′-movement. Consider the following:

(29) a. what did John see $t$
   b. what did John expect $t$ to be seen $t′$  (ibid.: 302, 303)

It is clear that the wh-phrase heading a categorial chain $CH_{\text{CAT}}=(\text{what}, t)$ cannot have its Case-feature checked in the operator position. The features of A′-traces, unlike those of A-traces, are assumed to remain extant. This permits the Case-feature in the trace position $t$ in (29a, b) to raise after Spell-Out for checking. The Case-feature is then erased in both the head and the tail of the chain $CH_{F}$ created by feature-movement as well as in the operator position of $CH_{\text{CAT}}$. In short, the following operations are involved here:

(30) a. overt A′-movement of $a$ to form $CH_{\text{CAT}}=(a_i, t_{ai})$
   b. raising of the Case-feature of $t_{ai}$ to form $CH_{F}=(F_i, t_{Fi})$
   c. erasure of the Case-feature in $F_i$, $t_{Fi}$, and $a_i$.

Based on this reasoning, Chomsky (1995: 304) proposes the following principle:

(31) The head of CH can attract or be attracted by $K$, but traces cannot attract and their features can be attracted only under narrow conditions reviewed (and left partially open).

Here, “narrow conditions reviewed (and left partially open)” are the qualifications needed for A′-traces (cf. (30b)).

Clearly, the operations involved in (30a-c) appear to include superfluous steps from the minimalist viewpoint: movement of unchecked Uninterpretable features to a position where they cannot be checked off, followed by subsequent erasure of the features in that position. If A′-movement can avoid accompanying such features, then the subsequent erasure will also be done away with. To the extent that this is possible, we can replace (31) with (32) below, crossing out “narrow conditions”:

(32) The head of CH can attract or be attracted by $K$, but traces can neither attract nor be attracted by $K$.

Let us consider first an example (33) and its derivation (34):
(33) (guess) what they think John will fix t
(34) a. $[VP \ V [VP \ fix \ what]]$ raising of V to $v \Rightarrow$
b. $[VP \ [Vb \ fix \ v] \ [VP \ tV \ what]]$ raising of $FF(what)$ to $Vb \Rightarrow$
c. $[VP \ [v \ FF(what) \ [v \ fix \ v]] \ [VP \ tV \ what]]$
d. $[CP \ Q \ [TP \ they \ think \ John \ will \ [VP \ tJohn \ [Vb \ FF(what)-fix-v] \ [VP \ tV \ what]]]]$ raising of $what$ to $[Spec, Q] \Rightarrow$
e. (guess) $[CP \ what \ [C' \ FF(what)-Q \ [TP \ they \ think \ John \ will \ [VP \ tJohn \ [Vb \ tFF(what)-fix-v] \ [VP \ tV \ tObj]]]]$
The weak Case-feature of embedded Vb attracts FF(Obj) as in (34c). Therefore the Case-feature of what is checked and erased before the strong Q-feature of matrix C attracts the whole category what with FF(what), as in (34e). Note that the category what has its wh-feature erased, once FF(what) has raised to $v$; otherwise principle (28) would be violated. (Even if what retained its wh-feature, it would be FF(what) rather than what that is the closest, and raised, to matrix C under the MLC.) Still, the full category what is raised; otherwise the strong feature, being affixal, could not be satisfied.

The A'-trace of a categorial chain of wh-movement has its Case-feature erased and thus is inaccessible to further Case-driven feature-movement. In this manner, Case-checking of A'-moved elements is carried out without trace-movement.

If this is on the right track, then it amounts to saying that formal features are accessible to movement as far as they are unattracted. Thus, a strong feature can attract the full category in a trace position as in (34d, e), so that it forms a categorial chain (i.e. $CH_{CAT} = (\alpha, t_\alpha)$) by pied-piping a whole category; a weak feature can attract a set of the formal features of a trace, forming a FF-chain, i.e. $CH_{FF} = (FF, t_{FF})$.12

Under our assumptions, (32) can be sharpened by relativization with respect to chain-types, as follows:

12 It is unclear why a single movement operation can form two separate chains, i.e. $CH_{FF} = (FF(what), t)$ and $CH_{CAT} = (what, t)$ in (34e). Pending a better understanding of feature-checking, we leave this question open here, by pointing out an argument for such dual chain formation. The following examples are cited from Bošković (1997b: 361):

(i) a. John gave Mary a book and Peter a car.
(ia) demonstrates that gapping is not involved in (ia). Then the Case-feature of gave in (ia) should attract both FF(Mary) and FF(Peter), possibly forming two separate FF-chains.
(35) The head of \( \alpha \) can attract or be attracted by \( K \), but traces of \( \alpha \) can neither attract nor be attracted by \( K \), where \( \alpha = CH_{\text{CAT}} \) or \( CH_{\text{FF}} \).

In sum, the weak Case-feature of a verb attracts the Case-feature of a \( wh \)-phrase before \( wh \)-movement, so that the formal features of \( A' \)-traces are erased and invisible to movement-operations; under the SOH advocated here both \( A \)- and \( A' \)-traces receive exactly the same treatment with respect to immobility.

5. Adverbials and Adjacency Effects

Let us first consider the following paradigm that demonstrates adjacency effects on accusative Case-checking:

(36) a. John often reads books.
    b. *John reads often books.
    c. John often reads to his children.
    d. John reads often to his children.

(Chomsky (1995: 329–332))

Chomsky (1995: Chapter 4) postulates that the adverbs as in (36) bear features that a verb can attract (p. 390, note 104). In the following structure given to (36b), \( often \) c-commands books and hence the former is closer to the target \( T_{\text{Omax}} \) than the latter; raising of \( FF(\text{Obj}) \) across the adverb would not be possible under the MLC:

(37) John \([T_{\text{Omax}} FF(\text{Vb}) T]\ [\text{VP} t_{\text{Subj}} [\text{Vb} \text{reads-}\nu] [\text{VP} \text{often} \ t_{\nu} \text{books}]]\)

It has been suggested (cf. Noji (1997) and Nishioka (1997)) that this account is problematic, for it would also rule out a grammatical sentence like (36a). Suppose that (36a) is given the following structure, in which \( often \) is inserted in the inner Spec of \( \nu \), and the subject in the outer Spec:

(38) John \([T_{\text{Omax}} FF(\text{Vb}) T]\ [\text{VP} t_{\text{Subj}} [\nu \text{often} [\text{Vb} \text{reads-}\nu] [\text{VP} \ t_{\nu} \text{books}]]]\)

Here, the adverb is closer to the target \( T_{\text{Omax}} \) than the object and the MLC would bar raising of \( FF(\text{books}) \). Chomsky’s framework should be modified in such a way that post-Spell-Out movement could be somehow immune from an MLC-violation in (38) but not in (37); otherwise the derivation of (36a) would crash, contrary to fact.\(^{13}\)

\(^{13}\) One possibility is to assume that \( FF(\text{Obj}) \) is checked against \( \text{Vb} \) before \( FF(\text{Vb}) \) moves to \( T \). This is the analysis I will put forth below.
Let us see how our framework resolves this problem. As a first approximation, we keep to the assumption that adverbs like *often* can appear either in [Spec, *v*] or [Spec, *V*], while PPs like *in the garden* appear as a sister to *V* (see Chomsky (1995: 331, 332) and Noji (1997)). Consider the following derivation of (36a):

(39) a. \[
[v_p \ [v_b \ \text{reads } v] \ [v_p \ t_v \ \text{books}]}
\]

b. \[
[v_p \ [v_b \ \text{FF(books)} \ [v_b \ \text{reads } v]] \ [v_p \ t_v \ \text{books}]}
\]

c. \[
[v_p \ \text{often} \ [v_b \ \text{FF(books)} \ [v_b \ \text{reads } v]] \ [v_p \ t_v \ \text{books}]}
\]

d. \[
[v_p \ \text{John} \ [v' \ \text{often} \ [v_b \ \text{FF(books)} \ [v_b \ \text{reads } v]] \ [v_p \ t_v \ \text{books}]}
\]

e. \[
[v_p \ \text{John} \ [\text{T}^{\text{max}} \ \text{FF(Vb)} \ T] \ [v_p \ t_{\text{John}} \ \text{often} \ [v_b \ \text{FF(books)} \ [v_b \ \text{reads } v]] \ [v_p \ t_v \ \text{books}]}
\]

The weak Case-feature of *Vb* attracts FF(Obj) as in (39b). Then, *often* is inserted into the (inner) Spec of *v* via Merge in (39c). Then, the subject is inserted into the outer Spec of *v* as in (39d). The weak *V*-feature of *T* attracts FF(*Vb*) and the strong EPP-feature and weak Case- and \( \phi \)-features of *T* attract the subject to its Spec, as in (39e). These features enter into a checking relation. Note that in (39c) the adverb does not block attraction of FF(Obj), simply because the former is inserted after FF(Obj) is attracted to *Vb*. The derivation converges, as is expected.

If this is on the right track, (36a) provides us with an empirical argument that movement driven by a weak feature obeys strict cyclicity (cf. section 2.1): post-Spell-Out attraction of FF(Obj) across *often* to [FF(*Vb*)-*T*] is not tolerated as in (38); rather, attraction of FF(Obj) for Case-checking should take place at the latest before *vP* is merged with *T* (and FF(*Vb*) is raised to *T*) (cf. (39)).

Consider next (36b), whose (partial) derivation would look like:

(40) a. \[
[v_p \ v \ [v_p \ \text{often} \ [v' \ \text{reads books}]]] \ \text{raising of } \text{reads} \ \text{to } v \Rightarrow
\]

b. \[
[v_p \ [v_b \ \text{reads } v] \ [v_p \ \text{often} \ [v' \ t_v \ \text{books}]]] \ \text{raising of FF(Obj)} \Rightarrow
\]

c. \[
[v_p \ [v_b \ \text{FF(books)} \ [v_b \ \text{reads } v]] \ [v_p \ \text{often} \ [v' \ t_v \ \text{books}]]]
\]

Note that the adverb could not be generated in [Spec, *v*], in which case the order *V-Adv-Obj* in (36b) would not be yielded since *V* raises as high as *v* in English. When *often* is generated in [Spec, *V*] as in (40a), an adjacency effect results. The MLC prevents the weak Case-feature of *Vb* from attracting FF(Obj): *often* c-commands Obj and is closer to *Vb*. The derivation of (36b) does not converge, as expected.

As for (36c, d), the PP-internal DP does not undergo feature-attraction from within PP, with no adjacency effect yielded.
A similar account holds true of the contrasts below.

(41)  a. ?“I am leaving,” abruptly shouted John.
     b. *“I am leaving,” shouted abruptly John.

(Collins (1997: 36))

Here, we assume with Bobaljik (1995), Ura (1996) and Collins (1997) that matrix verbs in (41) are overtly raised to T (for some reason yet to be clarified). In (41a) John remains in [Spec, v] and the FF(John) is raised to \([T_{\text{hmax}} [v/_{\text{vb shouted-v}}] T]\) for the checking of Case- and \(\not\)-features; this movement does not move across the adverb in accordance with the MLC. In (41b) the adverb within the outer Spec of \(v\) c-commands FF(John), whose raising is barred by the MLC.

It is worthwhile to note that the derivation of (41b) cannot violate the MLC in order to satisfy TC in (27). It is not maintained that the derivation that violates the MLC and satisfies TC (i.e., raising of FF(John) to T in (41b) results in total checking of the raised features) is more economical than the derivation that satisfies the MLC and violates TC (i.e., raising of abruptly to T would not result in total checking). As we have discussed in note 11, the MLC cannot be overridden in order for TC to be satisfied.

In this section we have shown that the present approach reinforces Chomsky’s account of adjacency effects on Case-checking.\(^{14}\)

6. Avoidance of a Global Economy Approach to Q-Insertion

In Chomsky’s (1995: 233, 292) analysis, insertion of a Q-feature is hypothesized to take place after Spell-Out. Q is a feature for interrogative clauses. Against this, we will argue here that a global economy perspective to Q-insertion is done away with.

Let us first recap Chomsky’s (1995: 4.5.4) proposal. Q has at least two variants, viz. Q[WH] and Q[V]. The former is satisfied by a wh-feature of interrogatives while the latter is satisfied by a verbal feature via T-to-C movement. First we will focus attention on Q[V], and then on Q[WH].

He claims that Q[V] cannot have a null alternant; otherwise a sen-

\(^{14}\) It is plausible to assume that VP-adverbials lack features that are potentially attracted by the Case-feature of v/T in a cluster of languages (e.g. Dutch, French, etc.) that lack adjacency effects on Case-checking.
sentence like (42a) might be wrongly construed as interrogative:

(42) a. (Q[V]) John will see Bill.
b. \([c \text{FF(will)-Q[V]}]\) John will see Bill.

Suppose that Q[V] without phonological features is inserted. This insertion should be applied after Spell-Out, since Chomsky posits that Merge as well as Move is subject to Procrastinate (p. 292). Then, if Q[V] is strong, FF(will) is raised to check off this feature as soon as it is inserted. Then, mapped onto a structure like (42b) after Spell-Out, (42a) would be incorrectly interpreted as a yes-or-no question, despite the lack of T-to-C. He then stipulates that Q[V] cannot have a null alternant; rather, it is obligatorily provided with phonological features and inserted before Spell-Out. (A covertly inserted feature may not have phonological features; otherwise, they would enter the covert component and the derivation would crash at LF, violating FI (pp. 293, 230).)

Within the present framework, Chomsky's stipulation directly follows from a trivial assumption that Q[V] is strong in English. Consider first what would happen if Q[V] were weak. A weak Q[V]-feature, hypothesized to be Interpretable, triggers neither overt T-to-C movement nor covert raising of FF(T). Still, if Q[V] in (42a) were weak, it would force this sentence to be construed as a yes-or-no question. This is an undesirable situation. Suppose then that Q[V] is strong. Recall that strong features can be erased only when a full category is pied-piped (see (5)). (Note that strong Q[V] is not inserted after Spell-Out: there is no syntactic operation after Spell-Out.) Thus, the assumption that a Q[V]-feature is always strong yields the desirable result: it has no null alternant and hence always triggers overt T-to-C movement. The lack of such movement in (42a) implies that strong Q[V] is not inserted; as a result, this sentence is not construed as a yes-or-no question.

If this is tenable, we can dispense with Chomsky's (1995: 294) economy condition (8): \(\alpha\) enters the numeration only if it has an effect on output. Chomsky resorts to this condition in order to bar post-Spell-Out insertion of strong Q-features; if a strong feature lacking phonological features were inserted after Spell-Out, then the PF-output would be identical to that in which the strong feature has not been inserted. No PF-effect is expected. Without recourse to (8), we can deduce the fact that strong features must be inserted before Spell-Out: there is no syntactic operation after Spell-Out.

Our reasoning is also carried over to Q[WH]. Let us review Chom-
sky's treatment. The following are the examples with which he explains his argument against post-Spell-Out insertion of strong Q[WH]:

(43)  
   a.  Q [IP who will fix the car]
   b.  Q [IP John will fix what]
   c.  Q [IP John will fix the car how (why)]  (ibid.: 293)

(43a) is well-formed; (43b) is dubious at best in English and perfect in French, although (43c) is generally excluded. Suppose that strong Q[WH] is inserted after Spell-Out. This is actually required by Procrastinate under Chomsky's analysis (p. 292). Then, this strong feature should be erased via raising of some wh-feature. If so, the wh-feature of how (why) is covertly raised to Q in (43c). (43c) would then be nondistinct from a structure with overt wh-movement as in (44) below, abstracting away from T-to-C movement:

(44)  How (why) will John fix the car.

Either the overtly raised adjunct in (44) or its covert counterpart in (43c) could give rise to a well-formed operator-variable construction. Therefore, if strong Q[WH] were inserted after Spell-Out, (43c) would be wrongly ruled in, though (43a, b) are correctly predicted to be well-formed in this manner.

If, on the other hand, strong Q[WH] cannot be inserted after Spell-Out, then the language will employ an interpretive strategy like unselective binding. Chomsky (1995: 386, note 65) posits that arguments, unlike adjuncts, introduce a position for a variable, so that in-situ interpretation is available. In (43a, b) weak Q[WH] is inserted after Spell-Out and unselectively binds the wh-phrase in situ, giving rise to a well-formed operator-variable construction. In (43c) even if weak Q[WH] is inserted after Spell-Out, the wh-adjunct lacks a position for a variable, so that unselective binding as well as an operator-variable construction is not available. It ends up as an illegitimate LF-object, violating FI. This is a brief outline of Chomsky's solution to the contrast between (43a, b) and (43c).

In our terms, the fact that strong Q[WH] cannot be inserted after Spell-Out follows directly from (2) again: there is no syntactic operation at LF. The lack of overt wh-movement in (43a-c) is indicative of the fact that inserted Q[WH] is weak rather than strong. (We should then assume that Q[WH] in English is strong except for such particular cases.) Then, Chomsky's solution to the contrast in question can be partially maintained: the wh-adjunct in (43c) unlike the wh-arguments in (43a, b) lacks a position for a variable, ending up as an illegitimate
Before closing this section, let us consider French root questions. Assuming that a strong Q-feature is inserted into a root after Spell-Out in French, Bošković (1997a) attempts to accommodate the lack of overt wh-movement in matrix clauses and the purported obligatoriness of wh-movement from within embedded finite declarative CPs. If Bošković's analysis were correct, the above mentioned treatment for (42) and (43) in English would not be a universal property. On the contrary, there are reasons to believe that the property is not peculiar to English.

First, if French allowed for post-Spell-Out insertion of a strong Q-feature, wh-adjuncts interpreted as causal or manner adverbs could remain in situ, contrary to fact. Consider the following examples:

(45) a. "*Tu es venu pourquoi?" "Parce que je suis malade." 
   Lit. 'You came why (causal)?' 'Because I am sick.'
   b. "*Tu as ouvert la porte comment?" "Lentement."
   Lit. 'You opened the door how (manner)?' 'Slowly.'
   (Aoun (1986: 27))

If strong Q were inserted after Spell-Out, these ill-formed sentences would acquire licit operator-variable constructions just on a par with (44). Then their ill-formedness would remain unexplained.

Second, Bošković tries to derive the observation that a wh-object embedded in a finite CP must overtly raise to the matrix [Spec, C] in French, by assuming that the formal features of a wh-phrase may not raise across an intervening A'-head, i.e. the embedded C-head, due to an HMC-effect. Such an account would be falsified given that wh-in-situ is actually embedded in a finite CP:

(46) a. Tu penses que tu as vu qui?
   you think that you saw who
   'Who do you think that you saw?'
   b. Tu veux que je te presente à qui?
   you want that I you introduce to whom
   'To whom do you want me to introduce you?'

(46a, b) are perfect as non-echo questions, according to my informant to whom I am indebted for these examples (see also Aoun (1985: 62, 1986: 103)). These arguments suffice to undermine Bošković's account.

If our treatment of (42), (43), (45) and (46) is on the right track, then we need not resort to the global economy condition (8) so as to bar post-Spell-Out insertion of strong features. This result seems to
take us further down the path toward a local economy perspective of syntactic derivations.

7. A Remaining Puzzle: Predicate-Fronting

In this section, I seek a solution to a potential problem for our proposal. Collins' (1997) argument against Bobaljik's (1995) SOH might be carried over to our framework. Consider (47a, b):

(47) a. How likely to arrive on time was John.

b. *How likely to be a storm was there. (Collins (1997: 89))

Based on this contrast, he proposes a principle as follows:

(48) Overt movement of a bleeds covert movement from within a.

Thus, once the embedded clause is raised before Spell-Out, FF(a storm) in (47b) is not able to adjoin to matrix T^\text{univ} (and then to there) after Spell-Out. Then the Case-feature of the associate would remain unchecked and the derivation would crash. In contrast, (47a) is well-formed since there is no formal feature that undergoes post-Spell-Out movement from within the raised constituent. Within our framework, on the other hand, FF(a storm) is hypothesized to adjoin to T^\text{univ} (and then to there) before wh-movement. The principle (48) would be met, thereby wrongly ruling (47b) in. If so, the contrast would constitute counterevidence to our framework.

On the contrary, (48) is not without its problems. Idiom chunks and expletive there behave alike with respect to predicate-fronting:

(49) a. Advantage is likely to be taken of John.

b. *How likely to be taken of John is advantage? (Lasnik and Saito (1992: 141))

Since there is unlikely to be post-Spell-Out feature-movement involved here, the contrast in (49) does not readily fall under (48).

Lasnik and Saito's (1992) insight provides us with a promising way out of this problem. They suggest that predicates such as likely, promise and threaten are ambiguous between raising- and control-predicates and that predicate-fronting in question is well-formed only if the predicate is a control- rather than raising-predicate. Therefore well-formed sentences like (47a) involve control-predicates:

(50) *[How likely PRO to arrive on time] was John?

In contrast, ill-formed cases like (47b) involve raising-predicates:

(51) *[How likely t to be a storm] was there?

This analysis is confirmed by the fact that neither expletive there nor an
idiom chunk qualifies as a controller for a PRO-subject:

(52) a. *There tried [PRO to be a riot].
    b. *Advantage wants [PRO to be taken of John]. (ibid.: 141)

For these reasons, the ungrammaticality of (47b) and (49b) should be
ascribed to the fact that fronted raising-predicates involve unbound
subject-traces in violation of some form of the proper binding condition
(PBC), or FI that requires that a trace be bound.

Consider now (53). Fronting of an associated NP is permissible:

(53) A storm, there was likely to be.
    (cf. *To be a storm, there was likely.)

Under the SOH, FF(a storm) heads a chain CH_{FF}=(FF(a storm), i) be-
fore the full category is fronted to a topic position, possibly [Spec, C].
A question naturally arises: why does the PBC not rule out (53)?
Note that a checked feature has to be erased so as to satisfy the prin-
ciple (28). Moreover, erasure of the formal features in the trace posi-
tion does not yield a violation of any principle (e.g. the principle of re-
coverability of deletion). Once erased, the trace within the CH_{FF} is
accessible neither to any operation nor to interpretation at LF. It fol-
lows that no violation of the PBC, or FI is involved in (53). In the
cases of (47b) and (49b), the trace of a full category must not be
erased, due to Chomsky's (1995: 281) following principle:

(54) A term ‡" cannot erase.

Therefore, a violation of the PBC or FI cannot be avoided.\footnote{15, 16}

\footnote{15} This account is not readily extended to the following example:

(i) *[John to be nice], I considered. (Collins (1997: 89))

One might claim that the deviance of (i) shows that an ECM-subject is overtly
shifted to the matrix clause and John cannot be carried along by fronting of the
complement TP. However, my informant finds the following sentence rather de-
graded, which is unexpected under such an analysis:

(ii) *To be nice, I considered John.

Note that (i) also casts doubt on Collins' claim, since FF(John) within the A'-
trace position could be attracted after Spell-Out. Then, the derivation of (i) would
converge, contrary to fact.

A reviewer suspects that trace-movement is responsible for the contrast between
(47b) and (53). Jang (1997: 32) seeks a similar way out of this problem, adding a
qualification to trace-immobility: FF(XP) can be moved from the trace of XP but
not from XP which is structurally internal to the trace of a larger category. If this
were correct, it would not only restore Chomsky's "narrow conditions" on trace-
movement as in (31), but also complicate it with Jang's qualification, which is not a
welcome result.

We will leave these issues unsolved in this paper.
If this move is tenable, the contrast in (47a, b) and the like may not undermine the present proposals. Of course, this sketchy account will need to be elaborated in future study.

8. Conclusion and Further Possibilities

In this paper, we have enumerated several arguments to defend the Single Output Hypothesis. The proposal of pre-Spell-Out feature-checking leads to (i) the rigid observance of strict cyclicity on the basis of Earliness, (ii) the derivational rather than representational definition of c-command, (iii) the abolition of Procrastinate, (iv) a guarantee of the immobility of A'-traces as well as A-traces, (v) the reinforcement of Chomsky's treatment of adjacency effects on Case-checking, (vi) the avoidance of a global economy approach to Q-insertion. To the extent that it is viable, the SOH does a better job of accommodating the data that have been believed to support the minimalist program.

Left open is the question of what status is given to traditional LF-operations like QR (quantifier raising), reconstruction (or a rule akin to QR and subsequent deletion), and anaphor-raising. Our speculations on these questions are sketched here. For reconstruction and anaphor-raising, Chomsky's (1995: Chapter 3) theory will be carried over with slight modifications (cf. (19)); formal (and possibly, semantic) features (but not phonological features) undergo these operations to form legitimate LF-objects (e.g. operator-variable constructions) before Spell-Out, although they have no phonological reflexes. As to QR, one can adopt Chomsky's (1995: 377) and Sato-Zhu's (1996) claim that it is a formal feature [quant(ificational)] that raises to a certain scope-related head such as T0 or Dis0 (=distribution) (see also Terada (1997)). In stark contrast, recent studies (Fox (1995), Wilder (1997),

16 As a reviewer suggests, the PBC needs an explanation in terms of more fundamental principles. Kitahara (1997: 69ff.) attempts to derive this condition from the c-command condition on Move. Consider the following derivation of (49b). The embedded subject advantage is moved out of TP2 in (i). Then α is raised to [Spec, C] in (ii):

(i) \[TP1 \text{ advantage is } [\alpha \text{ how likely } [TP2 \text{ to be taken } t \text{ of John}]]\]
(ii) \[CP \alpha \text{ how likely } [TP2 \text{ to be taken } t \text{ of John}] [\text{TP1 advantage is } t_{α}]\]

The c-command condition on Move is satisfied, so that Kitahara's (1997) approach does not suffice to subsume (49b) as well as (47b). We will leave this issue to future study.
and Kennedy (1997)) convincingly argue that QR raises a whole quantified NP rather than [quant] only. If they are correct, it is not implausible to modify the Attract-F Theory in such a way that some scope-related feature of T/ν, which is weak, pied-pipes all the formal features of a quantified phrase as a whole (e.g. FF(every picture of John) rather than FF(every) only (see also Chomsky (1995: 318, 322 for related discussion)). Full explorations of this issue would take us too far afield here. I will leave it to future research.

REFERENCES


Faculty of Humanities
Yamaguchi University
1677-1 Yoshida, Yamaguchi-shi
Yamaguchi 753–8512
e-mail: terada@po.cc.yamaguchi-u.ac.jp