THE MINIMAL LINK CONDITION
AND THE THEORY OF MOVEMENT

TORU ISHII

Kitami Institute of Technology

This paper presents new empirical evidence for Chomsky’s (1995, 1998, 1999) featural view of movement, which claims that the core part of movement is an operation on formal features, Attract/Move-F or Agree, with displacement of a category α being a consequence of pied-piping. Empirical support for the featural view of movement comes from the presence/absence of superiority effects, which has resisted any minimalist account. It is shown that superiority facts straightforwardly follow from the featural Minimal Link Condition (MLC), which is based on the featural view of movement, but not from the categorical MLC, which is based on the traditional Move-α view of movement. It is also shown that the wh-island constraint, which prima facie undermines the credibility of the featural MLC, can be subsumed under the Phase Impenetrability Condition (PIC).*

Keywords: minimal link condition, superiority effects, wh-island constraint, phase impenetrability condition

1. Introduction

Chomsky (1995, 1998, 1999) argues that movement of a category α is not a primitive operation but a complex operation consisting of an operation on formal features of α, Attract/Move-F (Chomsky (1995)) or Agree (Chomsky (1998, 1999)), and subsequent displacement of α. Under this featural view of movement, the core part of movement is Attract/Move-F or Agree and displacement of a category α, what Chomsky (1998, 1999) calls Merge, takes place only if required by a condition which must be satisfied before Spell-Out for convergence.

* I am grateful to Brian Agyan, Hiroshi Aoyagi, Yuji Takano, and anonymous EL reviewers for insightful suggestions and criticisms. The usual disclaimers apply.

English Linguistics 17: 2 (2000) 305-329 — 305 —
© 2000 by the English Linguistic Society of Japan
This is in contrast with the traditional Move-a view of movement, where the primitive form of movement is displacement of a category a. Although Chomsky presents conceptual arguments for the featural view of movement from the standpoint of the minimalist program (MP), it is more desirable to find supporting empirical evidence.

This paper investigates the formulation of the Minimal Link Condition (MLC), a locality constraint which subsumes the Relativized Minimality (RM) effects in the sense of Rizzi (1990). Under the Move-a view of movement, the MLC is formulated as a constraint on category movement (see, among others, Chomsky (1993) and Chomsky and Lasnik (1993)). Under the featural view of movement, on the other hand, the MLC is formulated as a constraint on Attract/Move-F or Agree (see, among others, Chomsky (1995, 1998, 1999)). Let us refer to these two formulations of the MLC as the categorical MLC and the featural MLC, respectively. I will argue that the featural MLC is empirically superior to the categorical MLC, which presents empirical evidence for Chomsky’s (1995, 1998, 1999) featural view of movement.

The organization of this paper is as follows. Section 2 investigates superiority facts. I will argue that they straightforwardly follow from the featural MLC, thereby presenting empirical evidence for the featural view of movement. Section 3 discusses the wh-island constraint, which prima facie undermines the credibility of the featural MLC. I will argue, however, that the wh-island constraint can be subsumed under the Phase Impenetrability Condition (PIC) and thus does not constitute evidence against the featural MLC. Section 4 makes concluding remarks.

2. Superiority Effects

2.1. Superiority Effects and Their Cancellation

It has been observed that multiple wh-interrogatives exhibit contrasts like (1) and (2), which are known as superiority effects:

(1) a. Who, t_i bought what?
b. *What, t_i did who buy t_i?

(2) a. Whom, did you persuade t_i to buy what?
b. *What, did you persuade whom to buy t_i?

While (1a) and (2a), where the higher wh-phrase undergoes movement, are acceptable, (1b) and (2b), where the lower wh-phrase undergoes movement, are less so.

An important fact about the superiority effects is that they are cancelled in certain unexpected cases (see, among others, Kayne (1984), Pesetsky (1987), and Hornstein and Weinberg (1990)):

(3) a. Which boy, t_i bought which book?
b. Which book, t_i did which boy buy t_i?

(4) a. Which boy, t_i did you persuade t_i to buy which book?
b. Which book, t_i did you persuade which boy to buy t_i?

Pesetsky (1987) argues that the contrast between cases like (1) and (2) on the one hand and those like (3) and (4) on the other comes from the fact that which-phrases are discourse-linked (D-linked) whereas wh-phrases like who and what are normally not D-linked. In order for a speaker to be able to felicitously ask a question containing a D-linked wh-phrase, both the speaker and the hearer must have in mind a particular set of entities from which the hearer is to choose a felicitous answer. Such a set of entities is established by the discourse. When a speaker asks a question containing a non-D-linked wh-phrase, on the other hand, neither the speaker nor the hearer need have a particular set of entities in mind. Putting technical details aside, Pesetsky claims that non-D-linked wh-phrases must undergo overt/covert movement to the domain of C for their interpretation, which induces the superiority effects. D-linked wh-phrases, on the other hand, are able to receive an interpretation without movement, which makes them immune from the superiority effects.

As pointed out by Hornstein and Weinberg (1990), however, Pesetsky’s D-linking analysis is untenable:


(6) a. Whose friend, t_i reviewed whose book?
b. Whose book, t_i did whose friend review t_i? (Hornstein and Weinberg (1990: 150))

In (5) and (6), neither a speaker nor a hearer need have in mind a particular set of entities from which the hearer chooses a felicitous answer. In other words, neither what in (5) nor whose in (6) is D-linked in the sense of Pesetsky (1987). Nonetheless, (5) and (6) do not exhibit any superiority effects.1

1 Pesetsky (1987: 109) presents (i) as further evidence in support of his D-linking the
One might argue that the contrast between (1) and (2) on the one hand and (3)-(6) on the other resides in the fact that while the wh-element solely makes up the wh-phrase in the former, the wh-element appears in the prenominal position within the wh-phrase in the latter. This view is untenable, however, since when how many-NP is used, the superiority effect remains (see Pesetsky (1987: 107)):

(7)  a. I need to know how many people, t_i voted for whom.
    b. *I need to know whom, how many people voted for t_i.

Among various proposals that have been made regarding the superiority effects (see, among others, Chomsky (1973, 1995), Hendrick and Rochemont (1982), Cheng and Demirdash (1990), Hornstein and Weinberg (1990), Lasnik and Saito (1992), Chomsky and Lasnik (1993), Hornstein (1995), Kitahara (1997), and Epstein (1998)), it is only Hornstein and Weinberg (1990) that has given an account of the above-mentioned presence/absence of the superiority effects. Hornstein and Weinberg (1990) assumes the theory of generalized binding proposed by Aoun (1985a, 1985b), arguing that superiority violations should be treated as binding theory violations. Although their analysis adequately explains the superiority facts, it is incompatible with the Minimalist Program (MP), where binding relations with extensive use of indexing and its percolation are not available. The next section proposes a minimalist account of the superiority facts. Specifically, I will argue that the featural MLC can straightforwardly explain the presence/absence of the superiority effects. For expository purposes, the following discussion assumes Chomsky’s (1995) view of movement. It should be noted, however, that arguments to follow also hold under Chomsky’s (1998, 1999) Agree view of movement.

2.2. The Featural MLC and the Superiority Effects

Chomsky (1995) proposes the operation Attract-F, which incorporates the featural MLC as part of its definition:

(8) Attract-F
    K attracts F if F is the closest feature that can enter into a checking relation with a sublabel of K.

(Chomsky: 1995: 297)

In (8), K is the target of the operation. A sublabel of K is a feature of the zero-level projection of the head of the target K. The notion of “closeness” in (8) is defined in terms of the notion of c-command. Informally, a is closer to K than β if K dominates both a and β and a asymmetrically c-commands β. I assume the standard notion of c-command, which states that a c-commands β iff a does not dominate β, and every γ that dominates a also dominates β (see, among others, Chomsky (1986)). According to the Attract-F view of movement, movement for checking purposes always involves attachment of a feature F to the head H of K regardless of whether Attract-F applies in the covert component or in the overt component. If Attract-F applies covertly, nothing more happens. If Attract-F applies overtly, on the other hand, PF requirements trigger the application of an operation in the overt component that places the category a which used to contain F “close enough” to the attracted F. Specifically, a is raised to the Spec of H to which F is attached. The attracted F is then subsequently put back into the category a in the overt component so that they form a single unit that is interpretable by the rules of the PF-component.

Chomsky (1995) argues that the featural MLC, which is incorporated into the definition of Attract-F (8), explains the core cases of the super-

---

3 Note that arguments to follow also hold under the Move-F view of movement if the MLC is formulated as in (i):

(1) a can raise to target K only if there is no legitimate operation Move-β targeting K, where β is closer to K than a (where α and β are formal features).
riority effects. Let us consider (1a, b) as examples. Let us assume with Chomsky (1995) that the Q-feature of C in English is “strong” in the sense that it must be eliminated (almost) immediately upon its introduction to the phrase marker in terms of a checking operation. Let us also assume that wh-phrases have an interpretable Q-feature, which may enter into a checking relation with the Q-feature of C. During the derivation of (1a, b), we come to the stage where the “strong” Q-feature of C is to be eliminated through entering into a checking relation with the Q-feature of a wh-phrase:

\[ (9) \quad [\text{cf } \text{C}_{Q} \text{[who}Q \text{[buy what}Q\text{]]}] \]

Since who asymmetrically c-commands what, the former is closer to the target CP than the latter. The featural MLC requires that the CP, which has a “strong” Q-feature as its sublabel, should attract the Q-feature of who, but not that of what. The Q-feature of who attaches to C and enters into a checking relation with the Q-feature of C. Since Attract-F applies in the overt component, who undergoes pied-piping to the Spec of C for PF-convergence:

\[ (10) \quad [\text{cf who}Q \text{[C [t [buy what}Q\text{]]]} \]

This yields (1a). There is no way of deriving (1b). The contrast between (1a) and (1b) follows.\(^4\)

I will argue that the featural MLC explains not only the core cases of the superiority effects but also their unexpected presence/absence exemplified by (3)-(7). Let us first consider their unexpected absence, taking (3) (repeated here as (11)) as an example:

\[ (11) \quad \begin{align*} 
& \text{a. Which boy,} \ t \text{ bought which book?} \\
& \text{b. Which book} \text{ did which boy buy} \ t \text{?} 
\end{align*} \]

During their derivation, we come to the stage where the “strong” Q-feature of C is to be eliminated:

\[ (12) \quad [\text{cf C}_{Q} \text{[which}Q \text{[boy [buy which}Q \text{book]]]} \]

In (12), it is the D which that has a Q-feature, which may enter into a checking relation with the Q-feature of C. According to the featural MLC, the CP, which has a Q-feature as its sublabel, may attract either the Q-feature of which in the subject position or that of which in the object position. This is because neither which in the subject position nor which in the object position is in the c-command domain of the other. Either which boy or which book may undergo movement to the Spec of C. Hence, both (11a) and (11b) are acceptable. The absence of the superiority effect in (4) can be explained in the same way.

The important point to note in this analysis is that Attract-F applies to the Q-feature of the D which, which is the head of the DP which boy/book. This is because in (12), the two occurrences of which are the only syntactic entities that have a Q-feature which may enter into a checking relation with the Q-feature of C. To see this point clearly, it is necessary to explicate the theory of bare phrase structure, which strictly derives from minimalist assumptions (see, among others Chomsky (1995, 1998, 1999)). The theory of bare phrase structure claims that phrase structure representations should be “bare” in the sense that they exclude anything beyond lexical features and syntactic objects constructed from them. Phrase structures are set-theoretic objects recursively constructed by Merge. There are two types of Merge, substitution and adjunction. Suppose that Merge applies to \( \alpha \) and \( \beta \), where \( \alpha \) and \( \beta \) are syntactic objects. Under the theory of bare phrase structure, the constructed object is of the form \( \{ \gamma, \{ \alpha, \beta \} \} \) (substitution) or \( \{ \gamma, < \alpha, \beta > \} \) (adjunction). \( \alpha \) and \( \beta \) are the constituents of the constructed object. \( \gamma \), the label of the constructed object, is the head of the projecting element. Essentially following Chomsky (1998), I claim that substitution takes place when there is a selector. The notion of selector includes selectional restriction features, i.e. the categorial selection properties of functional heads and the thematic selection properties of lexical heads. It also includes the feature attached to the head H of K, the target of an operation, in terms of Attract-F. Recall that the attached F requires the category \( \alpha \) which used to contain F to be merged in the Spec of H for PF-convergence. Adjunction, on the other hand, takes place when there is no selector. Adjunction is only motivated by a condition on phrase markers proposed in Chomsky (1993). In the MP, the computation proceeds in parallel, and thus at each point in the derivation, we have \( \Sigma \), a set of phrase markers. We may apply Spell-Out at any point, but if \( \Sigma \) is not a single phrase marker at Spell-Out, the derivation crashes at PF.

\(^4\) Essentially following Cheng (1991), I assume that a clause is interpreted as interrogative at LF only when its head or Spec position is occupied by an element with a Q-feature. Otherwise, it is interpreted as noninterrogative. In (10), who stays in the Spec of C and thus the clause is interpreted as interrogative. Under this analysis, matrix yes-no questions in English have an empty operator with a Q-feature in the Spec of C. See, among others, Larson (1985) for such an analysis.
This is because PF rules cannot apply to a set of more than one phrase marker and no legitimate PF representation is generated. For example, merger of a nominal with an adjective, though not triggered by any selector, is motivated by the above condition on phrase markers, since if it does not apply, Σ is not a single phrase marker at Spell-Out and thus the derivation crashes at PF.

Let us consider the internal structure of which boy/book under the theory of bare phrase structure. Merger of which with boy/book has a selector, i.e. the selectional restriction feature of D, which requires that D should take a projection of N as its complement. Hence, substitution takes place, yielding (13):

(13) \{ which, \{ which, boy/book \} \}

In (13), which and boy/book are the constituents of the constructed object. Since the constructed object is a projection of which, its label is which. In (13), the only syntactic entity that has a Q-feature is the D which, which appears as a constituent and as the label of the constructed object (13). It should be noted that, in (13), there is no syntactic entity corresponding to the DP node in the representation of the traditional phrase structure theory (14):

(14) \[
\begin{array}{c}
\text{DP} \\
\text{D} \\
\left. \right| \\
\text{NP} \\
\left. \right| \\
\text{N} \\
\left. \right| \\
\text{which} \\
\left. \right| \\
\text{boy/book} \\
\end{array}
\]

Under the traditional phrase structure theory, it might be possible to claim that the Q-feature of the head D which percolates up to the DP node which boy/book, both the D which and the DP which boy/book thereby having a Q-feature. It would follow that Attract-F could apply to either the D which or the DP which boy/book. Such an analysis is impossible under the representation of the bare phrase structure theory (13), where there is no DP node. Hence, in (12), Attract-F applies to the Q-feature of the head D which of which boy/book.\(^5\)

\(^5\) See Takano (2000) for an analysis of remnant movement based on this view of

Let us next consider (5) and (6) (repeated here as (15) and (16)):

(15) a. What type of man, \( t \) generally reads what type of book?
b. What type of book, \( t \) does what type of man generally read \( t \)?

(16) a. Whose friend, \( t \) reviewed whose book?
b. Whose book, \( t \) did whose friend review \( t \)?

In (15), what, which has a Q-feature, is the head of the DP what type of man/book and thus does not c-command beyond what type of man/book. Turning to (16), let us assume with, among others, Abney (1987) and Chomsky (1995) that the possessive element 's, which is D, requires a nominal element to appear in its Spec. Merger of who with the rest of the structure has a selector, i.e. the selectional restriction feature of the possessive element, and thus substitution takes place:

(17) [who [\( 's(=D) \) [friend/book]]]

In (17), who, which is in the Spec of D, does not c-command beyond whose friend/book. Then, when the "strong" Q-feature of C is to be eliminated during the derivations of (15) and (16), neither the wh-element in the subject position nor the one in the object position is in the c-command domain of the other. Hence, the Q-feature of C may enter into a checking relation with the Q-feature of either of the wh-elements. The absence of the superiority effects in (15) and (16) follows.

Let us finally consider (7) (repeated here as (18)):

(18) a. I need to know how many people, \( t \) voted for whom.
b. *I need to know whom, how many people voted for \( t \).

Within the subject wh-phrase how many people, the degree element how modifies the quantifier many and how many as a whole further modifies the nominal element people. Since modification relations have nothing to do with selectional restriction features, neither merger

Attract-F under the bare phrase structure theory. For expository purposes, I depict a constructed set-theoretic object as a more complex configuration involving additional elements like nodes, bars, XP, and so on, unless subtle clarifications are directly relevant.

It has been claimed by, among others, Chomsky (1964) that wh-phrases like who and what should be analyzed as complex entities consisting of wh-elements and indefinite pronouns. Our analysis is compatible with such an analysis if we assume Tsai's (1994) implementation of this idea, where wh-phrases like who and what are constructed below the X\(^*\)-level.
of how with many nor merger of how many with people involves any selector. Hence, adjunction takes place. This yields a multi-segmented category, which is informally represented as in (19):

(19) \([\text{Q} \rightarrow \text{N} \rightarrow \text{many} \rightarrow \text{Deg} \rightarrow \text{how} \rightarrow \text{Q} \rightarrow \text{people} \rightarrow \text{N}]\)

In (19), Q/N many represents the maximal zero-level projection of the head Q/N formed by adjunction to Q/N. Let us assume the standard notion of domination, which states that \(\alpha\) dominates \(\beta\) if every segment of \(\alpha\) dominates \(\beta\) (see, among others, May (1985) and Chomsky (1986)). Then, how in how many people can c-command beyond how many people. When the “strong” Q-feature of C is to be eliminated during the derivation of (18), the CP attracts the Q-feature of how but not that of whom. This is because how asymmetrically c-commands whom and thus the former is closer to the CP than the latter. The superiority effect in (18) follows.\(^6\)

\(^6\) As correctly pointed out by an EL reviewer, the following paradigm presents prima facie evidence against our featural MLC analysis of the superiority effect:

(i) a. Who, \(t_i\) said to whom(m) that I bought a lemon?
   b. *To whom(m), did who say \(t_i\) that I bought a lemon?

(ii) a. Who, \(t_i\) did you say that I bought what?
   b. *What, did you say to whom(m) that I bought \(t_i\)?

Specifically, in (ii), since who(m) is contained within the PP to whom(m), the Q-feature of whom(m) apparently does not c-command that of what. (ii), however, is deviant due to a superiority violation. The same paradigm can be observed with other examples involving a preposition followed by a wh-phrase like say of whom(m), hear from whom(m), and say about what.

It should be noted, however, there is independent evidence to suggest that the DP within the PP complement of the verb say c-commands beyond the PP:

(iii) *Mary said to him, that Susan loves John.

(ii) is deviant due to a violation of Condition C of the binding theory, which suggests that him, though it is contained within the PP-complement, c-commands John.

The question now arises as to how to ensure that the DP within the PP-complement of verbs like say c-commands beyond the PP. One possibility is to assume with, among others, Larson (1988) that verbs like say assign their \(\theta\)-role to the DP. The preposition which is semantically compatible with that \(\theta\)-role appears as a Case assignor. Given this assumption, let us consider (ii). Recall that while substitution involves a selector, adjunction does not. In (ii), whom(m) is assigned its \(\theta\)-role by the verb say, and to appears as a Case assignor. Since to does not assign any \(\theta\)-role to whom(m), merger of these two items has nothing to do with selection. Then, to is adjoined to whom(m), creating the adjunction structure \([\text{PP} \rightarrow \text{Q} \rightarrow \text{whom(m)}]\). Hence, the Q-feature of whom(m) c-commands that of what; the superiority effect follows. See Pesetsky (1995) for a different approach to this problem. I leave the fuller study of this important subject for future research.

\(^7\) For some speakers, (22b) is not perfect. Even for those speakers, however, (22b) is definitely better than (i):

(i) *John needs to find out how many people, whom killed \(t_i\) on the scene.

(20) a. Which boy, \(t_i\) bought what?
   b. What, did which boy buy \(t_i\)?

(21) a. I need to know what type of people, \(t_i\) voted for whom.
   b. I need to know whom, what type of people voted for \(t_i\).

(22) a. John needs to find out which person, \(t_i\) killed how many people on the scene.
   b. John needs to find out how many people, which person killed \(t_i\) on the scene.

(23) a. Which boy, did you persuade \(t_i\) to buy what?
   b. What, did you persuade which boy to buy \(t_i\)?

In (20)–(23), the “superior” wh-phrase is of the which/whose/what-NP type. The superiority effect does not appear, despite the fact the other wh-phrase is of the how many-NP/what type. When the “superior” wh-phrase is of the how many-NP/what type, on the other hand, our analysis predicts that the superiority effect appears even if the other wh-phrase is of the which/whose/what-NP type. This prediction is also borne out:

(24) a. Who, \(t_i\) bought which book?
   b. *Which book, did who buy \(t_i\)?

(25) a. I can’t remember whom, \(t_i\) reviewed whose book.
   b. *I can’t remember whose book, who reviewed \(t_i\).

(26) a. I can’t remember how many people, \(t_i\) voted for which person.
   b. *I can’t remember which person, how many people voted for \(t_i\).

(27) a. Who, did you persuade \(t_i\) to buy which book?
   b. *Which book, did you persuade who to buy \(t_i\)?

Second, our analysis predicts that when how many-NP as a whole
appears in the prenominal possessive position within a DP, the superiority effect is canceled. This is because in such a case, how may not c-command beyond the DP. This prediction is also borne out.\(^8\)

\[(28)\] a. I need to know how many students’ mothers, \(t_i\) voted for whom.

b. I need to know for whom, how many students’ mothers voted \(t_i\).

Third, our analysis can explain the contrast in (29):

\[(29)\] a. *Who, did you see [which picture of \(t_i\)]?

b. [Which picture of who], did you see \(t_i\)?

Since which c-commands who within which picture of who, which is closer to the target CP than who. The CP attracts the Q-feature of which, but not that of who. Which picture of who undergoes movement to the Spec of C, excluding the raising of who to that position. The contrast between (29a) and (29b) follows.

It should be noted that the categorical MLC (30), which is based on the traditional Move-a view of movement, cannot fully explain the presence/absence of the superiority effects:

\[(30)\] a can raise to target K only if there is no legitimate operation Move-\(\beta\) targeting K, where \(\beta\) is closer to K than a (where a and \(\beta\) are categories).

(adapted from Chomsky (1995: 296))

Especially, the categorical MLC cannot explain the absence of the superiority effects in (11), (15), (16), and (20)-(23). Let us consider (11) as an example. Since which boy asymmetrically c-commands which book, the former is closer to the target CP than the latter. Which boy raises to the Spec of C, which excludes the raising of which book to that position. The categorical MLC would wrongly predict that (11b) is deviant. In the same way, the categorical MLC would rule out (15b), (16b), and (20)-(23b), contrary to fact.\(^9\) It should also be noted that the categorical MLC cannot explain the contrast between (29a) and (29b). Specifically, it cannot rule out (29a). This is because neither which picture of who nor who is in the c-command domain of the other. It is clear that who does not c-command which picture of who. Which picture of who does not c-command who either, since the former dominates the latter.

This section has shown that the presence/absence of the superiority effects constitutes empirical evidence in favor of the featural MLC and against the categorical MLC, thereby presenting evidence for the featural view of movement.\(^10\) The next section discusses the wh-island boy does not c-command the head which of the wh-phrase which book, the categorical MLC would allow which book to move to the Spec of C; the absence of the superiority effect in (11) would follow.

It should be noted, however, that the MLC is a constraint on the primitive form of movement operations. The categorical MLC, which is based on the Move-a view of movement, does not compare heads or phrases but legitimate category movement operations. In (11), although the questioned element is the head which, the rest of the wh-phrase, i.e. boy/book, must be carried along by pied-piping. This is because extraction of which out of which boy/book is illegitimate. In other words, there is no legitimate category movement operation which only applies to the head which of which boy/book. Then, the categorical MLC can only compare the raising of which boy and that of which book, both of which are legitimate category movement operations. Hence, the categorical MLC cannot explain the absence of the superiority effect in (11). I would like to thank an EL reviewer for bringing my attention to this issue.

\(^{10}\) There are remaining problems to be solved regarding the superiority effects. First, it has been observed by, among others, Lasnik and Saito (1992) and Epstein (1998) that in (i), the embedded wh-subject in-situ which takes matrix rather than embedded scope:

(i) Who wonders what who bought?

As pointed out by Chomsky (1995: 387), however, this might reflect a preference for association of similar wh-phrases. In (ii), where what appears in the matrix clause and whom appears in the embedded clause, the embedded subject who takes embedded rather than matrix scope:

(ii) What determines to whom who will speak?

I leave this subject for future research.

Second, it has been observed by, among others, Pesetsky (1987) and Cheng and Demirdash (1990), that there is a contrast in unacceptability between standard superiority violations like (1b) and pure superiority violations like (2b). While the former is completely impossible, the latter is less severely deviant. Our analysis cannot explain this contrast. Such a problem, however, is inherent to all approaches which assume the MP, where there is no principled way of accommodating the degree of unacceptability. I also leave this subject for future research.

---

\(^8\) For some speakers, (28b) is awkward for some unknown reason. It should be noted, however, that even for those speakers, (28b) is much better than (18b), which is completely impossible.

\(^9\) One might argue that the categorical MLC (30) can explain the absence of superiority effects in (11), (15), (16), and (20)-(23) if we assume that it compares the heads of the wh-phrases rather than the wh-phrases. Let us consider (11) as an example. According to such a view, since the head which of the wh-phrase which
constraint, which presents prima facie evidence against the featural MLC and in favor of the categorical MLC.

3. The Wh-Island Constraint

3.1. Against the Featural MLC Analysis of the Wh-island Constraint

Chomsky (1995, 1998) argues that the wh-island constraint, a Relativized Minimality (RM) effect, can be explained by the featural MLC. Let us consider (31) as an example:

(31) *What do you wonder [who$_0$ [ti fixed ti$_1$]]?

In the derivation of (31), we come to the stage where the “strong” Q-feature of the matrix C is to be eliminated:

(32) [CP C$_{Q0}$ [you wonder [CP who$_0$$_Q$ [C [ti fixed what$_{Q0}$]]]]]

In (32), the “strong” Q-feature of the embedded C has already been eliminated by the raising of who to the Spec of the embedded C. Since who asymmetrically c-commands what, the former is closer to the matrix CP than the latter. There is no way of raising what to the Spec of the matrix C; the deviancy of (31) follows. Note in passing that in (32), who in the Spec of the embedded C may raise to the Spec of the matrix C in order to check the “strong” Q-feature of that C, yielding (33):

(33) [CP who$_Q$$_0$ [C [you wonder [CP ti$_1$ [C [ti fixed what$_{Q0}$]]]]]]

I claim that this derivation crashes. Let us assume with Lasnik and Saito (1992) that the intermediate traces of a wh-phrase do not bear Q-features (i.e. +WH-features in their terms). Recall that we are assuming that a clause is interpreted as interrogative at LF only when its head or Spec position is occupied by an element with a Q-feature (see note 4). Then, the embedded clause in (33), where neither its head nor its Spec is occupied by any element with a Q-feature, is interpreted as noninterrogative. The selectional property of the matrix verb wonder, which requires that its complement clause should be interrogative, is not satisfied. This makes this derivation crash.11

11 Chomsky (1998) also claims that there is no way of generating a convergent derivation from (32). See Chomsky (1995) for a different view that (32) converges, though it is gibberish.

There is, however, empirical evidence to suggest that the featural MLC account of the wh-island constraint is untenable. Let us consider the following example, where which-phrases are used instead of wh-phrases like who and what:

(34) *Which book$_0$ didn’t John remember [which boy$_0$ [Mary expected ti to read ti$_1$]]?

(34) exhibits the wh-island effect. The featural MLC, however, cannot explain the deviancy of (34). During the derivation of (34), we come to the stage where the “strong” Q-feature of the matrix C is to be eliminated:

(35) [CP C$_{Q0}$ [John couldn’t remember [CP which$_{Q0}$ boy$_0$ [C [Mary expected ti to read which$_{Q0}$ book$_0$]]]]]

Recall that under the theory of bare phrase structure, the two occurrences of the D which are the only syntactic entities that have a Q-feature which may enter into a checking relation with the Q-feature of C. In (35), neither which of which boy nor which of which book is in the c-command domain of the other. According to the featural MLC, the matrix CP may attract the Q-feature of which of which book without being intervened by the Q-feature of which of which boy. Which book may undergo movement to the Spec of the matrix C. Hence, the featural MLC cannot rule out (34).12

It should be noted that the categorical MLC (30) can correctly rule out cases like (34). According to the categorical MLC, there are two candidates for raising to the Spec of the matrix C, i.e. which boy and which book. Since the former c-commands the latter, the former is closer to the matrix CP than the latter. The raising of which book to

12 This also casts doubt on the featural MLC account of the crossing constraint advocated by Kitahara (1997) and Ishii (1997a). This is because the crossing effects are observed not only with wh-phrases like who and what but also with which-phrases:

(i) a. *Which book$_0$, did you decide [which boy$_0$, to persuade ti, to buy ti$_1$]?
   b. *Which boy$_0$, did you decide [which book$_0$, to persuade ti, to buy ti$_1$]?

As correctly pointed out by an EL reviewer, the categorical MLC can accommodate the crossing effects in (i) (see Oka (1993) for a detailed discussion). See Pesetsky (1982) for a different syntactic approach. Alternatively, the crossing effects might be due to a constraint on parsing as argued by, among others, Fodor (1978). I leave this important subject for further research.
the Spec of the matrix C is prohibited due to the existence of which boy. The wh-island effect in (34) follows. This presents prima facie evidence against the featural MLC and in favor of the categorical MLC. In the next subsection, however, I will argue that the wh-island constraint can be subsumed under the Phase Impenetrability Condition (PIC), an independently motivated condition proposed by Chomsky (1998, 1999). The wh-island effect therefore does not count as evidence against the featural MLC.

3.2. A PIC Analysis of the Wh-Island Constraint

Chomsky (1998, 1999) proposes the Phase Impenetrability Condition (36), which ensures that derivations proceed phase by phase, thereby reducing computational burden:13

(36) In phase $a$ with head H, only H and its edge are accessible to operations outside $a$. (adapted from Chomsky (1998: 22))

Following Chomsky (1999: 10), I take the edge of H to be the residue of H outside H', i.e. specifiers and elements adjoined to HP.

Let us consider how the PIC explains the wh-island constraint, taking (34), which remains unexplained under the featural MLC, as an example. During its derivation, we construct the embedded CP (37):

(37) $P_1 = [cP \text{ which}_{Q_0} \text{ boy}_1 \ [C \ [\text{Mary expected} \ t, \ i_{i_1} \ \text{to read which}_{Q_0} \ \text{book}]])$

Although Chomsky (1998, 1999) claims that $vP$ as well as CP constitutes a phase, the discussion to follow assumes for expository purposes that only CP functions as a phase. It should be noted that our analysis is still valid under the assumption that $vP$ is also a phase. In the embedded CP phase (37), while which boy, which is in the Spec of C, is in its edge, which book is not. As the derivation proceeds, we come to the stage where the “strong” Q-feature of the matrix C is to be eliminated:

(38)

a. $[cP \text{ which}_{Q_0} \text{ John couldn't remember P,}]$

b. $P_1 = [cP \text{ which}_{Q_0} \text{ boy}_1 \ [C \ [\text{Mary expected} \ t, \ i_{i_1} \ \text{to read which}_{Q_0} \ \text{book}]])$

In (38), the Q-feature of the matrix C cannot enter into a checking relation with the Q-feature of which of which book. This is because which book is not the head of the embedded CP phase or in its edge and thus given the PIC (36), it is not accessible to operations in the matrix CP phase. There is no way of raising which book to the Spec of the matrix C: the deviancy of (34) follows. The wh-island constraint can be subsumed under the PIC.14

It should be noted that in the derivation of (34), if which book were also raised to the edge of the embedded C as in (39), it would become accessible to operations in the matrix CP phase and thus the derivation would converge, which is undesirable:

(39) $[\text{which}_{Q_0} \text{ book}_1 \ [\text{which}_{Q_0} \text{ boy}_1 \ [C \ [\text{Mary expected} \ t, \ i_{i_1}]]])$

There are two logically possible ways of merging which book to the edge of C, i.e. substitution and adjunction. We have to exclude these two possibilities. Let us first consider how to exclude substitution of which book to the edge of C, which creates a multiple-Spec construction. Recall that substitution takes place when there is a selector, including the feature attached to the head H of K, the target of an operation, in terms of Attract-F. Then, the multiple-Spec construction is formed only if the CP, which has a “strong” Q-feature as its sublabel, attracts not only the Q-feature of which of which book but also that of which of which book. Let us assume with Chomsky (1995) that whether a language allows the multiple-Spec construction is a parameterized property. It is conceivable that in languages like English where multiple wh-fronting within a clause is not allowed as shown in (40), the CP with a “strong” Q-feature as its sublabel does not attract more than one Q-feature:

(40) a. *Where$_1$ what$_1$ did you put $t_1$ $t_1$?

b. *What$_1$ where$_1$ did you put $t_1$ $t_1$?

Hence, there is no way of merging which book to the edge of C in

---

13 See Uriagereka (1999) for a similar proposal. For expository purposes, the present discussion adopts Chomsky’s (1998) definition of the PIC. Note, however, that arguments to follow also hold under Chomsky’s (1999) definition of the PIC. See note 15 for a detailed discussion of Chomsky’s (1999) definition of the PIC.

14 Recall that the raising of which boy from the Spec of the embedded C to the Spec of the matrix C violates the selectional restriction property of the matrix verb.
terms of substitution.

Let us next consider how to exclude adjunction of which book to the CP. Recall that adjunction takes place when there is no selector. It then follows that movement of which book to the CP-adjointed position is not driven by any formal feature. Here, I crucially assume with Chomsky (1998, 1999) that derivations are evaluated locally, specifically at each phase level. Movement of which book to the CP-adjointed position takes place not to satisfy any formal feature in the embedded CP phase but only to ensure convergence in the matrix CP phase. Such movement violates the economy condition which bans superfluous steps of a derivation if we evaluate this derivation locally, i.e. at the embedded CP-phase level. It should also be noted that adjunction of which book to the CP cannot be motivated by the condition on phrase markers mentioned in section 2.2, which states that Σ, a set of phrase markers, must be a single phrase marker at Spell-Out. This is because Σ is a single phrase marker at Spell-Out even if adjunction of which book to the CP does not take place. Hence, adjunction of which book to the CP is excluded. If derivations were evaluated globally, on the other hand, we would wrongly exclude (38) and choose the adjunction structure given that economy conditions only compare convergent derivations (see, among others, Chomsky (1993, 1995)). This is because (38) crashes in the matrix CP phase.15,16

15 It should be noted that adjunction of which book to the embedded CP can also be excluded under Chomsky’s (1999) framework. Before we come to that, it is necessary to explicate Chomsky’s (1999) condition on interpretation/evaluation of a derivation. Chomsky (1999) proposes condition (i) which all operations are subject to:

(i) Interpretation/evaluation for PH₁ is at PH₂ (where PH₁ is a strong phase and PH₂ is the next higher strong phase).

(adapted from Chomsky (1999: 10))

A strong phase is either CP or v*P, where v* is a light verb with a complete set of ph-features.

Chomsky argues that the PIC (ii) can be derived from the assumption that Spell-Out falls under (i):

(ii) In [ge Z ... [ge a [H YP]]], YP, the domain of H, is not accessible to operations at ZP, but only H and its edge a (where HP is a strong phase and ZP is the next higher strong phase).

(adapted from Chomsky (1999: 10))

Given that Spell-Out falls under (i), it applies to HP at ZP. Chomsky assumes that when Spell-Out applies to HP, only YP, the domain of H, is spelled-out. H and its edge a, on the other hand, belong to ZP for the purpose of Spell-Out. Once YP is spelled-out and thus handed over to the PP-component, it becomes no longer accessible to syntactic operations. It is important to note that in order to derive the PIC (ii) from condition (i), Spell-Out, which is a case of interpretation/evaluation, is assumed to take place after ZP is formed but before operations at ZP apply. Otherwise, operations at ZP would be able to access YP.

Bearing the above discussion in mind, let us now consider how adjunction of which book to the embedded CP can be excluded by (i). Although Chomsky (1999) claims that CP/v*P functions as a strong phase, the following discussion assumes for expositional purposes that only CP functions as a strong phase. According to (i), we evaluate the embedded CP after the matrix CP is formed but before operations at the matrix CP including Attract-F triggered by the “strong” O-features of C apply, i.e., when we construct (ii):

(iii) [CP [CQ] couldn’t John remember [CP which is book, [CP which is boy, [C Mary expected t, to read p]]]]

Adjunction of which book to the embedded CP, which is not triggered by any formal feature, is excluded due to a violation of the economy condition which bans superfluous steps of a derivation. The crucial assumption in this analysis is that we evaluate the embedded CP before operations at the matrix CP take place. This is because if which book were raised to the Spec of the matrix C when we evaluate the embedded CP, it might be possible to claim that movement of which book to the Spec of the matrix C, which is a legitimate operation, would license adjunction of which book to the embedded CP. I would like to thank an EL reviewer for bringing my attention to this subject.16

16 This drives us to the question whether the PIC (36) is responsible for the RM effects other than the wh-island constraint, i.e. superraising (ia) and the Head Movement Constraint (HMC) (ib):

(i) a. *John seems [that it is likely [t to win]].
   b. *Read John will t the book.

Let us first consider superraising (ia). Given that structures are constructed phase by phase, we come to (ii) during its derivation:

(ii) [CP [that it is likely [John to win]]]

Note that since John does not have any peripheral feature (P-feature) in the sense of Chomsky (1998), which ensures successive-cyclic A′-movement, there is no way of raising John to the edge of the embedded C. In the embedded CP phase (ii), John is not in the head or edge of the phrase and thus not accessible to operations in the matrix CP phase. There is no way of raising John to the Spec of the matrix T in the matrix CP phase. Hence, the PIC (36) explains superraising. Unlike the wh-island constraint, however, it is not clear at this point whether the PIC (36) has empirical advantages over the featural MLC. We can say at least that given the PIC (36), superraising can be explained without recourse to the featural MLC.

As correctly pointed out by an EL reviewer, Chomsky’s (1999) definition of the PIC mentioned in note 15 cannot exclude superraising. Chomsky’s (1999) PIC only claims that the domain of the embedded C is not accessible to operations at the matrix CP. It does not prevent operations at the matrix TP from applying to an element in the domain of the embedded C. Hence, if we assume Chomsky’s (1999) PIC, we still need the featural MLC to exclude superraising.

Turning to the HMC, it is not entirely clear whether the PIC (36) or the featural MLC is relevant to the standard cases of the HMC like (ib). As pointed out by,
Our analysis predicts that in languages where more than one *wh-* phrase may be merged in the edge of an embedded interrogative clause, there is no *wh*-island effect. The next subsection shows that this prediction is borne out.

3.3. Multiple-*Wh*-Fronting Languages

Rudin (1988) argues that there are two types of multiple-*wh*-fronting languages, i.e. MULTIPLY-FILLED SPEC-CP ([+MFL]) languages like Bulgarian and Rumanian and [−MFL] languages like Serbo-Croatian, Polish, and Czech. In the [+MFL] languages, all *wh*-phrases are fronted to the Spec of C. In the [−MFL] languages, on the other hand, only one *wh*-phrase is in the Spec of C and the others occupy a TP-initial position.

Rudin observes that while [−MFS] languages exhibit the *wh*-island effects as in (41b), [+MFS] languages do not as in (41a):

(41)  a. **Bulgarian**

    Koja ot tezi knigi se čudiš [koj znæ [koj which of these books wonder-2s who knows who provadaj]]?

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

    (Rudin: 1988: 457)

b. **Polish**

    *Co on zapylał [kto wynalazł]?

    what he asked who invented

    Lit. ‘What did he ask who invented?’

    (Rudin: 1988: 459)

Our PIC analysis explains this difference between the [+MFS] and [−MFS] languages.

Let us first consider the lack of the *wh*-island effect in the [+MFS] languages, taking the most embedded CP phase of (41a) as an example:

(42)  [ce koj o[kojaq ot tezi knigi] C [tu prodava ti]]]

    who which of these books sells

Let us assume with, among others, Koizumi (1995) and Richards (1997) that both koj ‘who’ and koja ot tezi knigi ‘which of these books’ may move to the Spec of C to check its ‘strong’ Q-feature. Specifically, let us assume with Richards (1997) that koj ‘who’ first moves to the Spec of C and then koja ot tezi knigi ‘which of these books’ moves to the inner Spec of C. Since koja ot tezi knigi ‘which of these books’ is in the edge of the most embedded C, it is accessible to operations in the intermediate CP phase; there is no *wh*-island effect in the most embedded CP. The *wh*-island effect in the intermediate CP phase can also be nullified in a similar way. The lack of the *wh*-island effect in (41a) follows.

The question arises as to why the featural MLC does not prevent koja ot tezi knigi ‘which of these books’ from moving across koj ‘who’ despite the fact that koj ‘who’ asymmetrically c-commands koja ‘which.’ Chomsky (1993, 1995, 1998) proposes the notion of equidistance (43):

(43) Terms of the same minimal domain are equidistant to the target of an operation that the minimal domain of a head H is defined as the set of terms immediately contained in projections of H). (adapted from Chomsky: 1998: 38)

I claim that the theory of bare phrase structure virtually ensures that if two terms are equidistant to the target of an operation, so are the features of their heads. This is because given the theory of bare phrase structure, when Attract-F applies to a term, it in effect applies to the features of the head of the term. Let us consider (42). Under the theory of bare phrase structure, since the term koja ot tezi knigi ‘which of these books’ is a projection of koja ‘which,’ its label is the head koja ‘which.’ When Attract-F applies to the term koja ot tezi knigi ‘which of these books,’ it only ‘sees’ the features of the head koja ‘which,’ which appears as the label of the term. Hence, Attract-F in effect applies to the features of the head koja ‘which.’ According to the definition of the notion of equidistance (43), since koj ‘who’ and koja ot tezi knigi ‘which of these books’ are in the minimal domain of C, these two terms are equidistant to the intermediate CP. Attract-F may apply to koja ot tezi knigi ‘which of these books’ and thus the Q-feature of its
head koja 'which' without violating the MLC.\textsuperscript{17}

Let us next consider the wh-island effect in the [−MFS] languages, taking (41b) as an example. During its derivation, we construct the embedded CP phase:

\[(44) \quad [cP \ kto_{QP} \ [C \ [TP \ co_{QP} \ f\_wyna\_laz\_t_{i}]]]\]

who what invented

In (44), kto 'who' moves to the Spec of C and co 'what' occupies the TP-initial position. According to the PIC (36), co 'what,' which is not the head of the embedded CP phase or in its edge, is not accessible to operations in the matrix CP phase. There is no way of raising co 'what' to the Spec of the matrix C. The wh-island effect in (41b) follows.

4. Conclusion

This paper has first shown that the presence/absence of the superiority effects, which has resisted any minimalist account, presents evidence for the featural MLC and against the categorial MLC, thereby constituting empirical support for the featural view of movement. It was then shown that the wh-island constraint, which prima facie undermines the credibility of the featural MLC, can be subsumed under the PIC and thus does not count as evidence against the featural MLC.

Finally, I will briefly point out theoretical issues raised by our PIC analysis of the wh-island constraint. First, if our PIC analysis is on the right track, it constitutes support for a derivational approach to language and against a representational approach. Recall that our PIC analysis of the wh-island constraint crucially makes use of the notion of phase. Let us assume Chomsky’s (1998, 1999) model, where phases are assembled to form unified linguistic levels, PF and LF; otherwise, no legitimate derivation is formed. The notion of phase is available during structure-building but obscured in the output representation.

\textsuperscript{17} I would like to thank EL reviewers for bringing my attention to this issue. It should be noted that under the light verb analysis of a clause proposed by Chomsky (1995), where subjects appear in the Spec of v, subjects and objects are not in the same minimal domain. Hence, in structures like (9), wh-subjects and wh-objects are not equidistant to the target of an operation.

where phases are put together. Our analysis therefore is only valid under a derivational approach but not under a representational approach, presenting support for the former.

Second, recall that our PIC analysis crucially assumes that derivations are evaluated locally, specifically at each phase level. Within the theory of computational complexity, it is generally agreed that local considerations induce less computational burden than global ones (see, among others, Chomsky (1995), Fukui (1996), and Ishii (1997b)). As argued by Chomsky (1998), however, it is not clear whether computational complexity matters for a cognitive system like language, which does not involve any processing but only stores information. In other words, there is no a priori reason to claim that language should be local to avoid the problem of computational complexity, but we need to seek a resolution of this local versus global issue on empirical grounds. Our analysis gives a local analysis of the wh-island constraint, providing support for the language design that language is local in nature.

REFERENCES


