Some Consequences of Feature Inheritance

GOTO Nobu

1. Introduction

In recent minimalist theorizing, the notion of phase is indispensable for a theory of language. Chomsky (2008: 143) explicitly expresses this point as follows: “along with Transfer, all other operations will also apply at the phase level.” Whereas this phase-based theory of language brings a number of conceptual and empirical advantages over the previous approaches advanced by Chomsky (2000, 2001, 2004), it also poses a new challenge: how C (a phase head) and T (a non-phase head) interact with each other in the course of derivation. Specifically, Chomsky (2007, 2008) argues that the phase head C is the locus of Agree and Tense-features, and that apparent A-properties observed at the T level such as subject-agreement and subject-raising are derivatively determined by virtue of its relation to C, via a mechanism of feature inheritance, and hence the immediate question to be investigated is the precise mechanism behind such a relation between C and T. The following discussion is devoted to clarifying the mechanism of C-to-T feature inheritance and elaborating Chomsky’s feature inheritance system further.

The paper is organized as follows: In Section 2, I first review Chomsky’s (2007, 2008) feature inheritance system and Richards’ (2007) argument for it. I then show that Richards’ argument entails that C-to-T feature inheritance operates differently for the matrix clause and the embedded clause, and propose that C-to-T feature inheritance is unnecessary in the matrix clause, whereas it is necessary in the embedded clause (Asymmetrical Feature Inheritance System). In Section 3, I demonstrate that the proposed feature inheritance system plays a crucial role in explaining some asymmetrical properties observed in English and French wh-questions. Section 4 concludes the paper.
2. Feature Inheritance System


In Chomsky (2007, 2008), a new mechanism called *feature inheritance* is introduced. Under this mechanism, Agree and Tense-features are inherited from C to T, as shown below:

\[
\text{C ... T}_{[\text{Agree}][\text{Tense}]}
\]

After this, the derivation proceeds as in (2) to derive A-properties at the TP level (where angled brackets enclose “copy” of the moved element):

\[
\begin{align*}
\text{(a) Agree-probe} \\
\text{[C [TP DP T}_{[\text{Agree}][\text{Tense}]} [v{\text{P}} < \text{DP}_{[\text{Agree}]} > ... ]]}
\end{align*}
\]

\[
\text{(b) A-movement}
\]

(2a) shows that the Agree-feature on T functions as a probe to enter into an Agree relation with the subject DP in the specifier position of \(v{\text{P}}\) (Spec-\(v{\text{P}}\)). According to Chomsky (2008: 142), since this probing is subject to minimal search, the closest element to T is chosen as the appropriate goal. This Agree relation makes it possible to value uninterpretable features; hence the Agree-feature on T and the Case-feature on DP are valued at the point of (2a). On the other hand, (2b) shows the step of derivation in which the DP is raised to the Spec-TP. This raising is naturally motivated by the Extended Projection Principle (EPP), which requires that clauses have subjects, or more precisely, that the specifier of Tense phrase (Spec-TP) be filled (I return to the nature of the EPP in Section 2.4). Since the advent of the probe-goal theory of Agree (Chomsky 2000), this analysis of A-movement has been basically maintained in a series of Chomsky’s works.

In Chomsky (2007, 2008), it is assumed that only an *edge-feature* (EF) on the relevant phase head plays a crucial role in implementing A’-movement. Thus, overt \(wh\)-movement to the Spec-CP is simply triggered by the probing by the EF on C:
Some Consequences of Feature Inheritance

(3a) EF-probe

\[ \text{[CP wh-phrase } C_{[\text{EF}]} \ldots <\text{wh-phrase}> \]

(b) A'-movement

(3a) shows that the EF on C functions as a probe to find an appropriate goal; in this case, a wh-phrase. Importantly, this probing is not subject to minimal search, contrary to the Agree-probe. In fact, Chomsky (2008: 151) argues that EF-probe “can seek any goal in its domain” and it “does not involve feature matching; hence Agree” (cf. ibid.: 161, fn. 49). Thus, no feature valuation takes place at the point of (3a), though a probe-goal relation is established between C and the wh-phrase. After this probe-goal relation, an overt wh-movement takes place; (3b) shows that the wh-phrase is raised to the Spec-CP, triggered by the EF-probe. This analysis of wh-movement under the locality-free probing by the EF on C is radically different from Chomsky’s (2000, 2001) wh-movement system, in which the Q-feature on C searches its c-command domain as a probe to find a wh-phrase under a locality condition like the one in A-movement system.¹


2. 2. Richards (2007)

Focusing on the mechanism of feature inheritance proposed by Chomsky (2007, 2008), Richards (2007) raises an interesting question as to why Agree and Tense-features must be inherited by T from C, or why C cannot keep the features without letting them inherited to T; the relevant mechanism of C-to-T feature inheritance is reproduced below:

(4) \[ \text{C } \ldots \text{T}_{[\text{Agree}] [\text{Tense}]} \]

¹ This EF-based analysis of wh-movement poses a serious question to the superiority effect, which has been captured by a locality condition on the probing by C (Chomsky 1995, 2000). I return to this issue in Section 3.2.
Richards (2007) argues that the mechanism of feature inheritance is deducible from two independently motivated requirements on Agree and Transfer, i.e., (i) the requirement that Agree and Transfer of uninterpretable features must happen together (Chomsky 2007: 18-19) and (ii) the requirement that only the domain of a phase head be transferred to the interfaces (Chomsky 2000: 108 and Chomsky 2001: 13). According to him, the mechanism of feature inheritance is indispensable to satisfy these two requirements at the same time.

To clarify his point, let us consider the derivation of (5), which is obtained if C-to-T feature inheritance does not take place (I put aside a Tense-feature for expository purposes):

\[
(5) \quad \text{Agree} \quad *\{\text{CP } C_{\text{[Agree]}} \{\text{TP} \text{DP}_{\text{[Agree]}} \text{T} \ldots\}\} \quad \text{Transfer}
\]

In (5), the uninterpretable Agree-feature stays on C without getting inherited by T and only the TP domain of C is transferred under the second requirement (ii). It is important to notice here that Agree and Transfer of the Agree-features on C and on DP do not happen together, contrary to the first requirement (i). To satisfy this requirement, the Agree-feature on C must be in the same domain as the Agree-feature on DP, i.e. the TP domain. Richards concludes that this is why C cannot keep the Agree-feature but instead it has to be discharged to T via feature inheritance. With inheritance, Agree and Transfer of the Agree-features on C and on DP can happen together in the same domain of TP, and the two requirements can be satisfied at the same time; consider the following:

\[
(6) \quad \text{Agree} \quad \{\text{CP } C_{\{\text{TP DP}_{\text{[Agree]}} \text{T}_{\text{[Agree]} \ldots}\}\}} \quad \text{Transfer}
\]
Some Consequences of Feature Inheritance

Specifically, given that Transfer cannot identify and delete uninterpretable features once they are valued at a different domain (cf. Chomsky 2001, 2007 and Epstein and Seely 2002), it is expected that in the illicit derivation of (5), the valued uninterpretable Agree-feature on C that is transferred at the next phase domain will cause the crash of the derivation, since the Transfer cannot identify the valued uninterpretable Agree-feature on C as such. When the unidentified valued Agree-feature on C arrives at the interfaces without being deleted, the derivation causes a violation of Full Interpretation at the interfaces, which dictates that uninterpretable features must be deleted before they arrive at the interfaces. On the other hand, in the licit derivation of (6), the uninterpretable Agree-feature on C is valued at the same domain as the interpretable Agree-feature on DP thanks to the inheritance by T; hence Transfer can properly see which Agree-feature needs to be deleted at the interfaces. Thus, to satisfy the two requirements at the same time, or more precisely, not to violate the first requirement under the phase-based Transfer system, the mechanism of feature inheritance is indispensable. In the next section, I try to seek a further consequence of the mechanism of C-to-T feature inheritance by clarifying an unclear point in Richards' argument.

2.3. A Consequence of Richards (2007)

Although Richards's argument for the feature inheritance system is very impressive, there is one unclear point: how is the edge of C in the matrix clause transferred? Given the phase-based Transfer system, it is expected that the edge of C in the matrix clause remains as a residue of Transfer in the narrow syntax and the position will not be transferred throughout the derivation. To illustrate the point, let us consider the cycles of Transfer of the sentence *What did you think that John bought?* given in (7) (where I ignore certain copies of *what* for expository purposes):

\[
(7) \left[ \text{TP} \text{did(C)} \ [\text{TP} \text{you} \ \text{v}^* \ \text{v}] \ \text{think} \ [\text{TP} \text{C} \ [\text{TP} \text{John} \ \text{v}^* \ \text{v} \ \text{buy} \ <\text{what}>]]] \right]
\]

\[
? \quad \text{Transfer 4} \quad \text{Transfer 3} \quad \text{Transfer 2} \quad \text{Transfer 1}
\]

In (7), under the phase-based Transfer system, Transfer 1, 2, 3, and 4 are appropriately transferred, which is signaled by the merge of the above phase head of each domain; i.e. the embedded v* head, the embedded C head, the matrix v* head, and the matrix C head, respectively. However, it is unclear how the final position of the derivation,
that is, the edge of the matrix C is transferred: the position where _what_ and _did_ appear. As we have seen in (5) and (6), since the phase-based Transfer system requires that the edge of C and TP be transferred separately, the matrix CP-edge is expected to remain as a residue of Transfer in the narrow syntax and will not be transferred throughout the derivation.²

The question is thus whether this consequence is appropriate for the interfaces. Clearly, the answer is no. As exemplified by _wh_-questions, there are many clear cases whose interpretations are derived by a proper scan of the matrix CP-edge position, and hence we need to have a means to transfer the edge to the interfaces. How can we make this happen then? To answer this question, I hypothesize the following:

(8) In the matrix clause, CP is transferred along with its domain.

Given (8), it becomes possible to transfer the matrix CP together with TP, as shown below:

(9) [\[CP \text{ what did(C) [TP you [\[\text{v}^* \text{ [VP think [\[CP C [TP John [\[\text{v}^* \text{ [VP buy <what>]]]]]]]]]}]]]]}

Since the matrix edge is now transferred appropriately, we can conclude that something like the manner of Transfer in (8) is needed to satisfy the Strong Minimalist Thesis (SMT) which holds that language is an optimal solution to interface conditions – a consistent guideline in a series of Chomsky’s minimalist research.

Importantly, this newly proposed manner of Transfer yields a significant consequence for the applicability of C-to-T feature inheritance in the matrix clause. As Richards argues, if C-to-T feature inheritance crucially relies on the phase-based Transfer system, then it is expected that the inheritance becomes inactive in the context where such a Transfer system is inactive. Given that such a situation indeed exists in the matrix clause, as indicated just above, it is reasonable to expect that the inheritance can be inactive in the matrix clause. Specifically, if the matrix CP is transferred along with TP under the hypothesis given in (8), then it is expected that the Agree-feature on

² See Nissembaum (2000) and Obata (2009) for relevant discussions on this.
C need not be inherited by T to satisfy the first requirement (i) that Agree and Transfer must happen together; this point is illustrated below:

(10)  
```
   Agree
[CP C[A] [TP DP[A] T ...]]
```
  Transfer

In (10), the matrix CP is transferred together with TP under the hypothesis (8). Of particular importance here is that the first requirement that Agree and Transfer must happen together can be naturally satisfied even though the Agree-feature stays on C. Therefore, given (8), it is unnecessary for the matrix C to pass the uninterpretable Agree-feature to T in order to satisfy the first requirement. In contrast, considering the fact that Transfer in the embedded clause can be properly handled with the phase-based Transfer system, as demonstrated in (7), we can conclude that C-to-T feature inheritance is always active in the embedded clause. Under this reasoning, I propose the following, which we may call *Asymmetrical Feature Inheritance System*:

(11) C-to-T feature inheritance is unnecessary in the matrix clause, whereas it is necessary in the embedded clause.

This *Asymmetrical Feature Inheritance System* can be taken as a theoretical consequence of Richards' argument and a partial answer to the remaining question of "whether inheritance is obligatory or optional" (Chomsky 2008: 149). The purpose of the following discussion is to seek empirical and theoretical consequences stemming from this *Asymmetrical Feature Inheritance System*.

2. 4. A Prediction

The proposed feature inheritance system provides a new feature distribution in the matrix clause that is not available in Chomsky-Richards's system. Under the present system, since C-to-T feature inheritance is unnecessary in the matrix clause, Agree and Tense-features can stay on C, as shown below:
On the other hand, since C-to-T feature inheritance is necessary in the embedded clause, Agree and Tense-features are inherited by T, as follows:

(13) C ... T_{[Agree][Tense]} (Embedded Clause)

The distribution in (13) is what is found in Chomsky-Richards’s system, but the distribution in (12) is a new one arising only from the present Asymmetrical Feature Inheritance System.

Before verifying this feature distribution, I would like to clarify my position on the nature of the EPP-feature on T (as Chomsky (2007, 2008) is unclear about it). In (13), which is widely accepted in the literature, it would not be implausible to reduce the EPP-feature to an inherent property of the Agree-feature on T and to seek an approach to eliminate it (cf. Bošković 2005 and Epstein and Seely 2006), but given the new feature distribution (12), we need to be careful about taking such an approach, since no formal features are now on T. Thus, in this paper, following the previous studies that support the existence of the EPP-feature on T (see Chomsky 1982 and Lasnik 2001, 2003, among many others), I take a modest position about the EPP by assuming (14):

(14) T inherently bears the EPP-feature (T_{EPP}).

Further, to maintain the idea that the EPP-feature on T operates in such a situation as (13) (where Agree and Tense-features are inherited), I hypothesize the following:  

(15) The EPP-feature on T is activated only when inheritance takes place.

Given this, it follows that the EPP-feature on T in the embedded clause is activated, since C-to-T feature inheritance takes place (see (13)), whereas the EPP-feature on T in the matrix clause is inactive, since C-to-T feature inheritance does not take place (see (12)). Under this assumption on the EPP, the feature distributions of (12) and (13) can

3 This idea is advocated by Abe (2009); see the original paper for relevant arguments.
be rewritten as follows, when A’-movement is involved:

(16) \( C_{[EF][Agree][Tense]} \cdots T \) \hspace{1em} (Matrix Clause)

(17) \( C_{[EF]} \cdots T_{[Agree][Tense][EPP]} \) \hspace{1em} (Embedded Clause)

In the matrix clause (16), all the relevant features (the EF for A’-movement, the Agree-feature for A-movement, and a Tense-feature) reside on \( C \) without getting inherited by \( T \), and because of this, the EPP-feature on \( T \) is not activated under the hypothesis in (15). On the other hand, in the embedded clause (17), while the EF stays on \( C \), the Agree- and the Tense-features are inherited by \( T \), and because of this, the EPP-feature on \( T \) is activated under (15).

In the next section, I discuss three asymmetrical properties observed in English and French \( wh \)-questions, and demonstrate that they are naturally derived from the present \textit{Asymmetrical Feature Inheritance System}.

3. Some Matrix/Embedded Asymmetries

3.1. Matrix/Embedded Asymmetry in SAI

English \( wh \)-constructions show a matrix/embedded asymmetry with respect to the availability of subject-auxiliary inversion (SAI); consider the following examples:

(18) a. Who will John visit?
    b. *Who John will visit?
    c. I wonder who John will visit.
    d. *I wonder who will John visit.

SAI occurs in the matrix clause (see (18a, b)), but it does not occur in the embedded clause (see (18c, d)). This matrix/embedded asymmetry in SAI has been accounted for under the assumption that T-to-C movement is applied in the matrix clause but not in the embedded clause (cf. den Besten 1983), but it is quite unclear why that must be so.

However, such an asymmetry follows from the present \textit{Asymmetrical Feature Inheritance System} under the natural assumption that a Tense-related auxiliary element is materialized on the position in which an affixal Tense-feature resides (Lasnik 1995). Under this assumption, let us first consider the grammatical embedded sentence (18c).
The relevant derivation is given in (19) below:

\[(19) \ [\text{CP who } C_{[EF]} [TP \text{ John will-
T}_{[Agree][Tense][EPP]} [\text{v'P <John> visit <who>}]]) = (18c)\]

In (19), the Agree- and the Tense-features are inherited from C by T, and consequently, the EPP-feature on T is activated under the hypothesis (15), and the EF on C raises who to the Spec-CP, while the Agree-feature on T raises John to the Spec-TP to satisfy the EPP on T, with the relevant feature valuation process taking place. Of particular importance here is that the Tense-related element will is directly merged to T to materialize the Tense-feature on T and it is not moved to C. This follows from the assumption that a Tense-related auxiliary element is materialized on the position in which an affixal Tense-feature resides, as I have mentioned just above. Hence, under our analysis, the impossibility of SAI in the embedded clause is attributed to the obligatory C-to-T feature inheritance in the embedded clause.

Taking the sentence (18a) for illustration, let us then consider why SAI occurs in the matrix clause. Under the present system, we can analyze (18a) as follows:

\[(20) \ [\text{CP who } C_{[EF][Agree][Tense]} \text{ will} [TP <will>] [\text{v'P John visit <who>}]]) = (18a)\]

In (20), it is important to notice that neither the Agree- nor the Tense-feature is inherited by T, and C keeps all the relevant features, and the EF on C raises who to the Spec-CP and the Agree-feature on C enters into an Agree relation with John in the Spec-v'P to value the uninterpretable Agree-features.\(^4\) Remarkably, given that the position of the Tense-feature determines the position of an auxiliary element, it follows that SAI occurs in the matrix clause, since C keeps the Tense-feature without getting it inherited by T and hence the auxiliary element will needs to be moved to C to materialize the Tense-feature on C. Thus, under our analysis, the availability of SAI in the matrix clause can be attributed to the non-application of C-to-T feature inheritance.

---

\(^4\) In (20), John stays in the Spec-v'P without undergoing movement to the Spec-CP. We can attribute the reason for it to the fact that who has already occupied the Spec-CP. However, if there is no such element at the Spec-CP, then it should in principle be possible for the agreeing subject to move to the Spec-CP by making crucial use of the Agree-relation with C. In Section 3.3, I argue that there is indeed such a case in French.
Some Consequences of Feature Inheritance

in the matrix clause.

In this way, the Asymmetrical Feature Inheritance System provides us with a “principled” reason for the availability of SAI in English. In the next subsection, I further argue that the present system enables us to take a new look at the superiority effect, one of the most illuminating properties of English \textit{wh}-movement.

3. 2. Matrix/Embedded Asymmetry in the Superiority Effect

The superiority effect, illustrated by the contrast in (21), has been explored since Chomsky (1973) and attracted a lot of attention from many linguists as revealing something important about the mechanism of \textit{wh}-movement:

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

The empirical generalization is that in a question involving more than one \textit{wh}-phrase, it is the “superior” \textit{wh}-phrase, i.e., the one that asymmetrically c-commands the other \textit{wh}-phrases, that undergoes overt \textit{wh}-movement to the clause-initial position.

Chomsky (1995) argues that this generalization is deducible from the following locality condition, which we may call Attract Closest:

(22) K attracts $a$ only if there is no $\beta$, $\beta$ closer to K than $a$, such that K attracts $\beta$.

Under this condition, the unacceptability of (21b) is correctly captured by virtue of the fact that \textit{what} is not the closest \textit{wh}-element to C and therefore cannot be attracted by C.

It is important to note, however, that such an Attract Closest-based analysis of the superiority effect does not hold anymore in Chomsky’s (2007, 2008) \textit{wh}-movement system, where since \textit{wh}-movement is triggered just by the EF on C that is not constrained by locality conditions (cf. Section 2.1), we can no longer attribute the superiority effect to a locality condition such as (22). Therefore, how to derive the superiority effect remains to be an open issue for the current minimalist approach (cf. Chomsky 2008: 152).

In relation to this, Grebenyova (2006) has reported an interesting fact about the superiority effect. Consider the following paradigm:
(23) a. Who bought what?
    b. ??What did who buy?
    c. I wonder who bought what.
    d. *I wonder what who bought.

Departing from the well-accepted judgment in (21), she claims that superiority violations in embedded questions are stronger than in matrix questions, and that there is a matrix/embedded asymmetry in this effect, as witnessed by the contrast in (23). Thus, the problem is what kind of mechanisms underlies such an asymmetry.

The Attract Closest-based mechanism of wh-movement is too strong to deal with (23): since it is always the wh-subject (the closest wh-element to C) that is attracted by C, (23b) is wrongly expected to be ungrammatical, contrary to Grebenyova's judgment (though the contrast between (23c) and (23d) can be correctly captured under the Attract Closest). Also, it is important to note that the EF-based mechanism cannot give a unified account to (23) either. In this case, the inherent property of the EF-probe by C is too weak to explain the fact of the embedded clause: since the EF-probe "can seek any goal in its domain," (23d) is wrongly expected to be grammatical, contrary to fact (though the fact of the matrix question in (23a, b) can be correctly captured).

To provide a unified approach to the matrix/embedded asymmetry in (23), Grebenyova notes another independently observed matrix/embedded asymmetry in the C-T level and makes the following generalization with respect to the superiority effect:

(24) Superiority effects are stronger in contexts without T-to-C movement.

As we have seen in (18), T-to-C movement occurs in the matrix clause but not in the embedded clause. Under this generalization, she argues that T-to-C movement creates an "equidistance" effect for wh-movement, which effectively makes the Attract Closest condition inert to C that attracts a wh-phrase to its Spec. Thus, in such a matrix clause as in (23a, b), application of T-to-C movement makes who and what equidistant from the matrix C, so that both wh-phrases are freely attracted to the Spec-CP without violating the Attract Closest. On the other hand, in such an embedded clause as in (23c, d), since there is no T-to-C movement, such an equidistance effect is not triggered, so that the downstairs what cannot be attracted to the Spec-CP beyond the closest who.
Grebenyova's system successfully captures the asymmetry in (23), but I would like to seek an alternative account for it, since the notions her system relies on (e.g. equidistance and Attract Closest) do not hold anymore in the current minimalist framework. Further, since her system does not answer the question of why T-to-C movement occurs in the matrix clause but not in the embedded clause, we cannot take her analysis as a principled one. By contrast, recall that we now have a principled reason why there is such a matrix/embedded asymmetry in T-to-C movement (see Section 3.1 for detailed explanation of it). Hence, in the following, I propose an alternative account for the asymmetry in (23) in terms of a principle of computational efficiency.

Keeping to Grebenyova's intuition, we can have the following generalization:

(25) Superiority effects are stronger in contexts where C-to-T feature inheritance takes place.

Given this, it is important to recall that the present system assumes that the inheritance takes place in the embedded clause, but not in the matrix clause, and further that the availability of the EPP-feature on T is crucially tied up with that of the inheritance; the relevant hypothesis is the following:

(26) The EPP-feature on T is activated only when inheritance takes place.

Thus, under this assumption, we can assign the following feature distributions to the matrix question (23a, b) and the embedded question (23c, d), respectively:

(27) \([C_{EF}] \ldots \text{who} \ldots \text{what}] \) (for (23a, b))
(28) \([C_{EF}] - T_{EPP} \ldots \text{who} \ldots \text{what}] \) (for (23c, d))

Here, comparing (27) with (28), we notice an important difference between them with respect to the availability of the EPP-feature on T. In (27), the EPP-feature on T is not activated, but in (28) it is activated under (26).

With this difference in mind, let us first consider why the superiority effect does not arise in the matrix clause. Under the present system, we can easily capture the fact by making crucial use of the inherent property of the EF-probing by C, as illustrated
below:

\[(29) \quad [C_{[EF]} \ldots \text{who} \ldots \text{what}]\]

Since the EF can seek any goal in its domain and establish a probe-goal relation without causing any intervention effects, \textit{what} as well as \textit{who} can be a goal for EF-satisfaction on C, and thus if the EF-probe selects \textit{who} as a candidate for a legitimate goal, then the derivation proceeds in the following way and yields the sentence (23a):

\[(30) \quad [\textit{CP} \textit{who} C_{[EF]} \ldots \langle \textit{who} \rangle \ldots \text{what}] (= (23a))\]

On the other hand, if the EF-probe selects \textit{what}, then the derivation proceeds in the following way and yields the sentence (23b):

\[(31) \quad [\textit{CP} \textit{what} C_{[EF]} \ldots \text{who} \ldots \langle \textit{what} \rangle] (= (23b))\]

Thus, under the present system, the lack of the superiority effect in the matrix clause is simply derived from the locality-free probing by the EF on C.\(^5\)

In contrast to this analysis of the matrix clause, it is quite significant to notice that the C head of the embedded clause (28) should take a different way of probing: since the C head renders the EPP-feature on T active with feature inheritance, it should take care of not only EF-satisfaction on C but also EPP-satisfaction on T. Given Chomsky’s conception that only phase heads can undergo relevant operations, it is natural to assume that the C head takes care of operations at the T level. It follows from this reasoning and a principle of computational efficiency that probing by C-T as a whole accesses the goal that can satisfy each requirement maximally. Thus, in a context like (28), it should be the case that probing by C-T targets \textit{who}, since it can satisfy not only the EF on C but also the EPP-feature on T maximally:

\(^5\) I assume, following Grebenyova (2006), that the mild deviance of (23b) results from a semantic factor. See her original thesis for this.
As a consequence of the probing for who (see (32)), the EF on C and the EPP-feature on T can be satisfied maximally and efficiently (see (33)), which gives rise to the sentence (23c). Noticeably, this analysis correctly excludes the derivation that leads to the ungrammatical sentence (23d), since even if what is probed by C-T, it cannot satisfy the EPP-feature on T as well as the EF on C. Hence, under our analysis, the existence of the superiority effect in the embedded question is reasonably attributed to the efficient-probing by C-T that maximizes feature satisfaction.

Above, we have limited our discussion to the superiority effect involving a wh-subject, but how can we treat the cases that do not involve the wh-subject, as illustrated below:

(34) a. Who(m) did you give what?
   b. *What did you give who(m)?

(35) a. Who(m) did John persuade to buy what?
   b. *What did John persuade who(m) to buy?

It is well-known that the superiority effect shows up in the double object construction (34) (Barss and Lasnik 1986) and the 'pure superiority' construction (35) (Pesetsky 1982).

Under the present approach and the assumption that the v*-V level is crucially involved in the derivation of a (wh-)object (Chomsky 2007, 2008), it is expected that the existence of the superiority effect in (34) and (35) also results from an efficient-probing at the v*-V level. This approach seems to be on the right track given that the probing by the EF, which triggers A'-movement, and transmission of the Agree-feature, which triggers A-movement, are properties of phase-heads in general, as proposed by Chomsky (2007, 2008). Under these assumptions, Agree-feature is inherited from v* (a phase head) to V (a non-phase head), just like the C-T relation; thus, the relevant feature distribution at the v*-V level for (34) and (35) is the following:
If only phase heads can undergo relevant operations, it is natural to assume again that in (36) the v* head takes care of operations at the V level, as the C head does for T, and to expect that probing by v*-V as a whole accesses the goal that can satisfy each requirement maximally in terms of the principle of computational efficiency. Thus, in (36), it should be the case that probing by v*-V targets who(m), but not what, since it can satisfy not only the EF on v* but also the Agree-feature on V maximally and efficiently:

\[
\begin{align*}
(36) \quad [v^*[EF] V_{[Agree]} \ldots \text{who(m)} \ldots \text{what}] & \quad \text{(for (34) and (35))} \\
\end{align*}
\]

Importantly, given this derivation at the v*-V level, the ungrammaticality of (34b) and (35b) follows immediately under the Phase-Impenetrability Condition (PIC), which states that no element inside the complement of a phase (CP and v*P) can be probed by a phase head outside that phase (see Chomsky 2000, 2001). The point is illustrated by (39) below:

\[
\begin{align*}
(37) \quad [v^*[EF] V_{[Agree]} \ldots \text{who(m)} \ldots \text{what}] \\
(38) \quad [\text{VP who(m)} C_{[EF]} [\text{VP who(m)} T_{[Agree]} \ldots \text{<who(m)> \ldots \text{what}}] \\
\end{align*}
\]

As is clear from (39), the EF-probing by C to what induces a violation of the PIC, hence only who(m) is allowed to move to the Spec-CP at the next phase level, triggered by the EF-probe:

\[
\begin{align*}
(40) \quad [\text{CP who(m)} C_{[EF]} \ldots \text{<who(m)> \ldots \text{what}}] & \quad (= (34a) \text{ and (35a)}) \\
\end{align*}
\]

This derivation correctly yields the grammatical sentences in (34) and (35) while excluding the ungrammatical ones. So, the present system can properly treat the superiority effect in the non-subject wh-questions as well, with the efficient-probing.
Some Consequences of Feature Inheritance

at the v*-V level that maximizes feature satisfaction and the EF-probing by C that is subject to the PIC.

Interestingly, we can account for the following contrast too:

(41) a. *What did you tell who about?
   b. What did you talk to whom about?

The above contrast has been accounted for by making crucial reference to the "c-command" relation between the wh-phrases (cf. Fiengo 1980); cf. the relation between who and what in (41a), where the c-command relation is necessary for who to block the movement of what. In his recent work, however, Chomsky (2008: 141) suggests: "whether c-command plays a role within the computation to the C-I interface is an open question." This suggestion makes sense in our approach, and the contrast is correctly derived without recourse to such a dubious notion as c-command.

Specifically, in (41b), it is natural to claim that since the v* head does not have the Agree-feature to inherit to V (unlike the v* head in (37), which is available for (41a)), it can seek any goal in its domain under the locality-free probing by the EF on v* and freely establish a probe-goal relation with what as well as whom, as follows:

(42) \[ v^{*}_{[EF]} \ldots [\text{PP to whom}] \ldots [\text{PP about what}] \]

Here both whom and what can be an appropriate goal for EF-satisfaction on v*. If the EF-probe selects what as an appropriate candidate, then the derivation proceeds as follows:

(43) \[ [\text{CP what } C_{[EF]} \ldots [v_{P} \langle \text{what} \rangle v^{*}_{[EF]} \ldots [\text{PP to whom}] \ldots [\text{PP about } \langle \text{what} \rangle]] \]
   \[ (= (41b)) \]

In (43), what is moved to the Spec-v*P and the Spec-CP to satisfy the EFs on v* and C, respectively, in accordance with the PIC. Thus, the present system correctly captures the grammaticality of the sentence (41b), irrespective of the notion of c-command.

In the next subsection, I further argue that the proposed *Asymmetrical Feature
Inheritance System can provide a new approach to a distinguished property of French wh-questions too.

3.3. Matrix/Embedded Asymmetry in French Wh-In-Situ Strategy

Consider the following paradigm of French, which is taken from Bošković (2000):

(44) a. **Qui** as-tu **vu**?
    whom have-you seen
b. **Tu** as **vu** **qui**?
    you have seen who
    ‘Who did you see?’
c. Pierre a demandé **qui** tu as **vu**
    Pierre has asked whom you have seen
d. *Pierre a demandé tu as **vu** **qui**
    Pierre has asked you have seen whom
    ‘Pierre asked whom you saw.’

To express non-echo questions in French, a wh-phrase may or may not undergo overt wh-movement to Spec-CP in the matrix clause (see (44a, b)); however, in the embedded clause, a wh-phrase must undergo overt wh-movement to Spec-CP (see (44c, d)).

Bošković (2000) extensively discusses this peculiar property of French wh-questions in terms of Chomsky’s (1995) minimalist framework. According to him, the reason why no overt wh-movement takes place in (44b) is straightforward: the CP projection that triggers overt wh-movement, or more precisely, the strong wh-feature in the sense of Chomsky (1995) is not present in the overt syntax, though it is present in (44c, d) where overt wh-movement takes place. In his system, the in-situ wh-phrase can be licensed in LF by its formal features undergoing movement to the C head that is inserted in LF, and hence whether the in situ wh-phrase is properly licensed or not is crucially dependent on the possibility of C-insertion taking place in LF. He claims

---

One might argue that the ungrammaticality of (44d) just results from a failure of the selectional requirement of demander ‘ask’ that requires a wh-clause in its complement, but Bošković (2000) argues against this possibility; see his original paper for this.
insertion is performed in such a way that it always expands structure. Therefore, the reason why the in situ *wh-*phrase is not allowed in (44d) is trivial: the LF C-insertion to license the in situ *wh-*phrase does not expand the tree; hence the in situ *wh-*phrase is not licensed. But in (44b), the LF C-insertion can apply to the top of the sentence without violating the condition, and hence the in situ *wh-*phrase is licensed.

Bošković's analysis is very insightful, but I would like to seek an alternative account for (44) in terms of the present *Asymmetrical Feature Inheritance System*, for two reasons: one is simply that the framework he adopts is crucially different from ours and the other is that our system is superior to Bošković's in that it can properly deal with the matrix/embedded asymmetry given in (45), which the latter cannot, in the same way as English SAI, reproduced in (46):

\[(45) \begin{align*}
a. & \text{Qui } \underline{\text{as-tu}} \text{ vu?} \\
& \text{`Who did you see?'} \\
b. & \text{Pierre a demandé qui } \underline{\text{tu}} \text{ as vu.} \\
& \text{`Pierre has asked whom you have seen.'} \\
c. & \text{*Pierre a demandé qui } \underline{\text{as-tu}} \text{ vu.} \\
& \text{`Pierre has asked whom you have seen.'}
\end{align*}\]

\[(46) \begin{align*}
a. & \text{Who } \underline{\text{will John}} \text{ visit?} \\
b. & \text{I wonder who } \underline{\text{John will}} \text{ visit.} \\
c. & \text{*I wonder who } \underline{\text{will John}} \text{ visit.}
\end{align*}\]

I have argued in Section 3.1 that the matrix/embedded asymmetry in English SAI in (46) is automatically derived from *Asymmetrical Feature Inheritance System*, repeated in (47).

\[(47) \begin{align*}
a. & C_{[EF][Agree][Tense]} \ldots T \text{ (Matrix Clause)} \\
b. & C_{[EF]} \ldots T_{[Agree][Tense]} \text{ (Embedded Clause)}
\end{align*}\]

Given that French subject-clitic inversion (SCI) in (45) is an instance of T-to-C movement, just like English SAI (cf. den Besten 1983, Kayne 1983, Rizzi and Roberts 1989), it is natural to claim that the asymmetry in (45) also results from (47): in the matrix clause, the Tense-feature resides on C without getting inherited by T, and hence the Tense-related element *as* `have’ is moved to C to materialize the Tense-feature (see
(45a)); on the other hand, in the embedded clause, the Tense-feature is inherited to T from C, and hence such a movement does not take place (see (45b, c)). Hence, (47) crucially operates on French questions too, just like English, and I offer an alternative account to (44) in terms of it.

Let us then first consider (44a, b). Under the present system, the case involving overt _wh_-movement (44a) can simply be derived from the EF-probing by C:

\[
\begin{array}{c}
\text{[CP qui } C_{\text{EF}} \ldots \text{ tu } \ldots <\text{qui}>] (= (44a)) \\
\end{array}
\]

Given this EF-based analysis of _wh_-movement, it is natural to conjecture that in the sentence without _wh_-movement (44b), _wh_-movement does not take place because the EF that triggers _wh_-movement has already been deactivated or satisfied by another element at the Spec-CP. Thus I propose the derivation in (50) for (44b) with the additional assumption (49):

\[
\begin{array}{c}
\text{[CP tu } C_{\text{Agree}} \ldots \text{ <tu } \ldots \text{ qui}] (= (44b)) \\
\end{array}
\]

By positing (49), I intend that in French the EPP-feature can work in conjunction with the Agree-feature, independently of (15), which, by hypothesis, regulates the availability of the EPP-feature in terms of the feature inheritance. This independent assumption in French seems to be natural given the fact that in this language, agreement is strongly connected to movement (see Kayne 1989 for the validity of this assumption based on the data from past participle agreement). Thus, in (50), since

---

7 It is important to notice that the English counterpart of (44b) is ungrammatical for expressing non-echo questions (cf. *You have seen who*?), and hence it is natural to consider that the grammaticality of the French sentence (44b) has something to do with a unique property of French that is not attested in English.

8 In relation to this, the following difference between English and French may be relevant:

(i) a. There seem to the general to be two soldiers missing from the barracks.
the Agree-feature stays on C, it can attract the agreeing tu ‘you’ to its Spec to satisfy the EPP-feature on C, thereby deactivating the EF on C and leaving qui ‘who’ in-situ. I take this to be the fundamental reason for why wh-movement may not take place in the matrix clause: wh-movement is blocked because the agreeing subject occupies the Spec-CP. Since this derivation is crucially based on the claim that the C head in the matrix clause keeps the Agree-feature without handing it to T, this analysis lends further support to the present Asymmetrical Feature Inheritance System.

Here, it is crucial to notice that the same thing will never happen in the embedded clause; since in the embedded clause, the Agree-feature is inherited by T, the subject moves to the Spec-TP, but not to the Spec-CP, and hence the EF that triggers wh-movement is never deactivated by the agreeing subject as in (50). Consequently, wh-movement must take place in the embedded clause to satisfy the EF on C (see (44c)). If it did not take place, then the derivation would crash at the interfaces due to a violation of EF-satisfaction (see (44d)).

If this analysis is on the right track, the above discussion will lend further support to the EF-based wh-movement system, proposed by Chomsky (2007, 2008), and to Asymmetrical Feature Inheritance System developed so far.

4. Conclusion

In this paper, I attempted to clarify the mechanism of C-to-T feature inheritance and elaborate Chomsky’s (2007, 2008) phase-based theory of language further. Specifically, as a consequence of Richards (2007), I proposed Asymmetrical Feature Inheritance System, according to which C-to-T feature inheritance is necessary in the embedded clause, but unnecessary in the matrix clause. I supported this system by mainly examining some matrix/embedded asymmetries in English and French wh-constructions. Though I have not yet found out good empirical reasons to apply the present system to non-wh-constructions in English, German seems to provide a

b. Il semble au général y avoir deux soldats manquants à lacasarne.

(ib) is the French counterpart of the English expletive construction (ia). What is important to note is that in (ib) the same element satisfies the Agree-feature on T as well as the EPP-feature on T (i.e. the expletive il’), whereas in (ia) different elements satisfy the features: the EPP-feature is satisfied by there and the Agree-feature is satisfied by two soldiers. Arguably, this difference will follow from the assumption given in (49).
good empirical reason to motivate such an extension; consider the following examples:

(51) a. Der Mann hat den Hund gesehen.
    the man have the dog seen
    'The man has seen the dog.'

b. Er sagte daß der Mann den Hund gesehen hat.
    he said that the man the dog seen have
    'He said that the man has seen the dog.'

It is well-known that German declarative sentences show a matrix/embedded asymmetry with respect to the V2 effect: a finite verb occurs on C in the matrix clause (see (51a)) and on T in the embedded clause (see (51b)). Given the possibility that this fact is also treated in the same way as English SAI (cf. den Besten 1983), it is quite reasonable to extend the present *Asymmetrical Feature Inheritance System* to the German V2 effect and to reason that it holds true for non-*wh*-constructions as well. Thus, an appropriate empirical generalization that will emerge in dealing with such German declarative sentences seems to be as follows:

(52) C-to-T feature inheritance does not take place in the V2 environment.

While English is exceptional among other Germanic languages in that it does not show a V2 effect in declarative sentences, it is also true that such an effect is still retained in *wh*-constructions. Thus, it is promising to extend the present system so as to accommodate V2 effects. A further consequence of the present system awaits future research.

---

9 This conclusion is independently reached by Obata (2009).

10 Although we have not discussed so far how English declarative sentences like *John kissed Mary* is derived under the present system, in which C-to-T feature inheritance is unnecessary in the matrix clause, i.e., Agree and Tense-features can stay on C, it follows from (52) that C-to-T feature inheritance does take place in the English declarative sentence, as is often assumed, and the relevant derivation is analyzed as having the familiar Agree relation between the Agree-feature on T, which is inherited from C, and the Agree-feature on a DP, which is merged to the Spec-vP. This is because the declarative sentence clearly falls within a non-V2 environment. Thus, a natural conclusion is that while C-to-T feature inheritance does not take place in the English *wh*-sentence (V2 environment), as we have argued, it does take place in the English declarative sentence (non-V2 environment), as it has been assumed in the standard analysis.
Some Consequences of Feature Inheritance

(135) 49

Works Cited


dissertation, University of Maryland.