

ELIMINATING THE STRONG/WEAK PARAMETER ON T*

NOBU GOTO
Toyo University

1 Introduction

This paper consists of two parts: the first part (section 2) identifies potential flaws in the labeling theory proposed in Chomsky (2015a); the second part (section 3) suggests a system that eliminates the flaws. More specifically, I argue that the strong/weak parameter on T regarding labeling, which is proposed by Chomsky (2015a) to capture the parametric difference in Extended Projection Principle (EPP)/Empty Category Principle (ECP) between non-null subjects languages (non-NSLs) like English and null subject languages (NSLs) like Italian, should be eliminated in favor of the optimal design of language. I then propose a new way of labeling by Merge of a head H to SPEC (*Merge H-to-S*) just as an instance of the general application of the simplest Merge: $\text{Merge}(\alpha, \beta) \rightarrow \{\alpha, \beta\}$. As a consequence of this proposal, it is shown that the strong/weak parameter on T is eliminable, the spirit of Chomsky's (2015a) labeling analysis of EPP/ECP is maintained, with significant empirical advantages, and feature inheritance (Chomsky 2008, Richards 2007) is provided with a simple principled explanation under the labeling theory.

* Portions of this paper have been presented at Kent Ridge Linguistics Colloquia held at National University of Singapore (NUS) (March 2, 2016), at the workshop on "Current Issues in Comparative Minimalist Syntax" held at the 34th annual meeting of the English Linguistic Society of Japan at Kanazawa University (November 12, 2016), at the workshop on "Exploring the Notion of Parameters in the Minimalist Program" held at the 153th annual meeting of the Linguistic Society of Japan at Fukuoka University (December 4, 2016), at Aoyama Gakuin University (February 28, 2017), and at Sophia University (March 10, 2017). I would like to thank Jun Abe, Andreas Blümel, Noam Chomsky, Mary Dalrymple, Yoshihito Dobashi, Michael Yoshitaka Erlewine, Naoki Fukui, Ken Hiraiwa, Mayumi Hosono, Hisatsugu Kitahara, Masayuki Komachi, Takashi Munakata, Hiroki Narita, Masashi Nomura, Masayuki Oishi, Soonhyuck Park, Marc Richards, Mamoru Saito, Yosuke Sato, Daniel Seely, Jae-Young Shim, Yushi Sugimoto, Koji Sugisaki, Sandhya Sundaresan, Shigeo Tonoike, Takashi Toyoshima, and all the audience for invaluable comments and discussions.

2 EPP, ECP, and Labeling

Chomsky (2015a) drops the assumption that EPP is a universal property of language, i.e., the requirement that clauses must have a subject, or more precisely, that SPEC-T be filled with a subject (see, among others, Chomsky 1981, 2000, 2008). To ensure this principle, it has been assumed that in non-NSLs like English, EPP is satisfied by an overt subject (i.e. **(He) speaks*), whereas in NSLs like Italian, EPP can be satisfied by a null pro expletive subject (i.e. *pro parla* ‘*_ speaks*’). However, dropping the long-standing assumption, and adopting Rizzi’s (1982) idea that the parametric difference is attributed to richness of agreement, Chomsky (2015a) proposes to parameterize the strength of T regarding a labeling algorithm (LA) proposed in Chomsky (2013): T in English-type non-NSL, with poor subject-verb agreement, is too “weak” to serve as a label, hence there must be an overt subject in SPEC-T to label the SPEC-TP construction as $\langle \phi, \phi \rangle$ by the agreeing features, whereas T in Italian-type NSLs, with rich subject-verb agreement, is “strong” enough to serve as a label, hence there is no need to have an overt subject in SPEC-T to label the SPEC-TP construction. Under the present theory, therefore, whether SPEC-T needs to be filled with an overt subject is attributed to the strength of T: if T is weak, the SPEC-T requires an overt subject, but if T is strong, it does not. It follows that EPP is no longer a universal property of language; SPEC-T does not have to be filled with an overt subject nor with a null pro expletive subject if T is strong (see Zushi 2005 for a similar view).

The strong/weak parameter on T captures the long-standing observation that further *wh*-extraction from SPEC-T is impossible in non-NSLs like English, but possible in NSLs like Italian (see Perlmutter 1971) (cf. **Who_i do you think that t_i will leave* versus *Chi_i credi che t_i partirà* ‘*Who_i do you think that t_i will leave*’). The classical explanation for the difference was attributed to ECP, i.e., the requirement that traces be properly governed (Chomsky 1981): in non-NSLs like English, subject traces fail to be properly governed, so that further *wh*-extraction is impossible, whereas in NSLs like Italian, subject traces are properly governed by the null pro expletive subject, so that further *wh*-extraction is possible. Dropping this ECP-based explanation in the GB era, Chomsky (2015a) offers a unified account of the parametric difference in ECP under the strong/weak parameter on T: in non-NSLs like English, T is too weak to serve as a label, hence the *wh*-subject must stay in SPEC-T to keep the SPEC-TP construction labeled as $\langle \phi, \phi \rangle$, but in NSLs like Italian, T is strong enough to be able to label the SPEC-TP construction, hence the *wh*-subject does not have to stay in SPEC-T and can raise further.

Given that memory is phase-level (see Chomsky 2008 and much related work), Chomsky (2015a) also provides an argument for why *wh*-extraction from SPEC-T is allowed in English when the phase head C undergoes deletion (see Perlmutter 1971) (e.g. *Who_i do you think t_i will leave*): if the SPEC-TP construction has already been labeled as $\langle \phi, \phi \rangle$ at a phase level, then the information is available and stays labeled at the phase level, so that even if the *wh*-subject raises further to SPEC-C, no labeling failure occurs. In this particular case, Chomsky (2015a) notes that phasehood of C is inherited by T along with ϕ -feature inheritance and accordingly activated on T by C-deletion (cf. Chomsky 2008, Richards 2007). Thus, T, not C, functions as a phase head here, and as a result, unlike the usual way, the interior of T is identified as a transferred domain, observing the Phase Impenetrability Condition (PIC). For this reason, when the phase head C undergoes deletion, a PIC violation by further *wh*-extraction from SPEC-T can be circumvented. When C is not deleted, on the other hand, C retains phasehood, so that the interior of C, TP, informally speaking, is transferred as usual, along with a *wh*-subject in SPEC-T, therefore

yielding the so-called *that*-trace effect. In this way, EPP, ECP, and their parametric differences can be unified under the labeling theory once we adopt the strong/weak parameter on T.

Although these arguments are suggestive, unfortunately, there is a major hurdle to overcome before we accept the parameter as a principled property of the language system. The Minimalist Program (MP) for the past 20 years tells us that the notion of “strength” runs the risk of stipulation (see, among others, Broekhuis 2000). In fact, Chomsky (1998: 127) himself has suggested the following research agenda for the MP, pointing out that such a notion should be eliminated in favor of the optimal design of language:

- (1) “[the] optimal design [of the computational system of language faculty] should eliminate such strange and difficult properties as strength”

Also, Chomsky (2001: 2) has put forward the following Uniformity Principle, suggesting that languages be uniform:

- (2) “In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances.”

In terms of these, the strong/weak parameter on T should be eliminated.

In relation to this, one of the conceptual problems of the parameter in question is why it is only T that undergoes such parameterization. In contrast to T, adopting the idea that roots R are unspecified as to category in the lexicon and their category derives from merger with a functional element (see, among others, Marantz 2013), Chomsky (2015a: 10) claims that “R is universally too weak to label.” This asymmetry between T and R with respect to the strength of labeling itself may be an interesting question, but if it is dispensable, it should be abandoned, in favor of (1) and (2).

A more serious problem in this regard is the fact that Chomsky’s (2015a) system has nothing to say about robustly attested correlations between rich object-verb agreement and null object phenomena, and even makes a wrong prediction regarding the possibility of null objects. If R (more specifically, V) in all languages is universally too “weak” to serve as a label, as Chomsky argues, then it should be predicted that there are no languages that allow null objects, because an overt object must be present in SPEC-V to label the SPEC-VP construction as $\langle \phi, \phi \rangle$ by the agreeing features, for the same reason as the case of the weak T in English-type non-NSLs.¹ However, this prediction is not borne out. Null objects are indeed found in languages with rich object-verb agreement. See, for instance, the following Pashto example (taken from Huang 1984):

- (3) a. ma **mana** wə-xwar-a
 I apple PRF-eat-3.F.SG.
 ‘I ate the apple.’
 b. ma _____ wə-xwar-a
 I _____ PRF-eat-3.F.SG.
 ‘I ate _.’

¹ V here is a synonym for R. Hereafter I will use it without distinction from R.

In (3a), the verb *xwar* ‘eat’ agrees with the object *mana* ‘apple’, manifesting the third person feminine singular form *-a*. Note here that due to the agreement morphology on the verb, the object *mana* ‘apple’ can be null, as shown in (3b) (see also Welsh, Swahili, Georgian, Arabic, etc. for null objects). Unfortunately, Chomsky’s (2015a) treatment of V is not consistent with this fact, and more seriously with Rizzi’s (1982) idea he adopts and applies to T. If the parametric difference is attributed to richness of agreement, then it should be more reasonable and consistent to parameterize the strength of V as follows: V in English-type non-null object languages, with poor object-verb agreement, is too “weak” to serve as a label, hence there must be an overt object in SPEC-V to label the SPEC-VP construction as $\langle \phi, \phi \rangle$ by the agreeing features (just like T in English-type non-NSLs), whereas V in Pashto-type null object languages, with rich object-verb agreement, is “strong” enough to serve as a label, hence there is no need to have an overt object in SPEC-V to label the SPEC-VP construction (just like T in Italian-type NSLs). However, Chomsky does not do so, completely ignoring the correlation between the richness of agreement and the possibility of null objects. The null object phenomenon therefore remains unclear in the labeling theory he assumes.

The parameterization of Italian T as “strong” also poses a fundamental question for Chomsky’s (2013) elegant argument for why the external argument (SUBJ) must raise from the v^*P -internal position: the syntactic object $SO = \{SUBJ, v^*P\}$, neither a head, cannot be labeled; hence the subject must raise (Chomsky 2013: 44). In other words, by raising the subject to SPEC-T, what is visible to LA in the SO becomes v^* alone, as in $\{t, v^*P\}$, t the lower copy of the subject, so that the SO is labeled as v^* , as required. The reason for the invisibility of the lower copy of the subject to LA, Chomsky argues, is that it is part of a discontinuous element, which is created when some element (here, SUBJ) in a domain D1 (here, SPEC- v^*) goes to a different domain D2 (here, SPEC-T). This argument is suggestive, but once we accept the strong/weak parameter on T, unfortunately, the impact of the argument is undermined. That is, if Italian T alone can label the SPEC-TP construction due to its strength, then it is not necessary for the subject to raise to SPEC-T, but is allowed to stay in situ. The question is, then, how the predicate-internal subject construction is labeled in such a situation. Chomsky (2015a: 9) recognizes this point as a potential problem of the labeling analysis of Italian EPP, putting aside about just where the subject and v^*P are when the subject does not raise. However, according to Belletti’s (2001, 2004) series of studies of the relevant structure (i.e. “subject inversion structures” in her terms), she concludes that when the subject does not raise, “either the subject remains in its original v^*P -internal position or it raises to a very low position in the clause structure, lower than the lowest SPEC hosting an adverb” (cf. “Cinque’s hierarchy”). If this is the case, it follows that the subject that does not undergo raising to SPEC-T amounts to staying around in its base position, i.e. the v^*P -internal position, and is hence still visible to LA. The construction in question is therefore expected to result in a labeling failure, contrary to fact: as we have seen above, for it to be labeled as v^* , the subject must raise (at best up) to SPEC-T so that the lower copy of the subject is invisible to LA. Then, the question remains how the predicate-internal subject construction is labeled when the subject does not raise to SPEC-T, or more relevantly, when T is strong.

What remains, then, is the impression that the strong/weak parameter on T may be a stipulation, and so should probably be eliminated from the theory if possible. Having said that, Chomsky’s (2015a) fine unification of EPP and ECP in terms of labeling should be defended, in light of recent developments of the labeling theory (for arguments for Chomsky’s 2013, 2015a labeling theory, see, among others, Epstein, Kitahara, and Seely 2014, 2015, Goto 2013, 2016,

Hiraiwa 2017, in press). Especially, given that everything unfortunate seems to stem from the radical assumption that EPP is not universal, it is tempting to ask whether it is indeed on the right track and the traditional assumption that EPP is a universal property of language cannot be maintained in a way that fits well into the present theory. Considering that EPP has received a significant amount of attention in the literature to date, there should be a breakthrough solution. If there is an independently motivated way to satisfy EPP, or to ensure $\langle \phi, \phi \rangle$ labeling without recourse to the strength of T, then it would be worth considering the possibility of the theory that incorporates the mechanism.

In what follows, I explore such a possibility, attempting to learn new things from the past, with appropriate updating of the relevant assumptions. My claim is, thus, as has traditionally been assumed, that EPP is universal, holding for both phases, CP and v*P, and for all languages (see Chomsky 2008), and, consequently, the strong/weak parameter on T can be, and hence should be, eliminated from the theory. More specifically, I argue that it is indeed possible to eliminate the stipulation if we adopt and elaborate on the widely accepted view that EPP can be satisfied by V-raising (for a theoretical precursor, see Alexiadou and Anagnostopoulou 1998 and Zushi 2003). Then, in section 2, I adopt the idea of EPP-satisfaction by V-raising, and elaborate on it, particularly in terms of the simplest Merge-based system, advocated by Chomsky (2013, 2014, 2015a, b, 2016, 2017a, b, c) and Chomsky, Gallego, and Ott (to appear), under which an SO is constructed by simplest Merge, defined as follows:²

$$(4) \text{ Merge}(\alpha, \beta) \rightarrow \{\alpha, \beta\}$$

Just as an instance of (4), I argue that Merge of a head to SPEC (henceforth, *Merge H-to-S*) is possible, and demonstrate that the labeling theory that allows this possibility is able to not only maintain the spirit of Chomsky's labeling analysis of EPP and ECP but also bring significant consequences for the halting problem, the feature inheritance mechanism, and *wh*-in-situ phenomena.

3 Labeling by Merge H-to-S and Its Consequences

Alexiadou and Anagnostopoulou (1998) offer an interesting approach to EPP-satisfaction: EPP can be satisfied by either DP-raising to SPEC-T or V-raising to T. They basically assume, as Chomsky (2015a) does, that the parametric difference is attributed to richness of agreement, proposing that in non-NSLs like English, with weak ϕ -agreement marking on V, EPP is satisfied by DP-raising to SPEC-T, whereas in NSLs like Italian, with rich ϕ -agreement marking on V, EPP can be satisfied by V-raising to T (as well as by DP-raising to SPEC-T). Since it is easy to find morphological evidence for the postulation of such a difference, I assume their approach to EPP-satisfaction in the labeling theory as a principled way to satisfy EPP, or more precisely, to

² Incidentally, Chomsky (2017b) proposes seven conditions on Merge – which he dubs “Seven Desiderata,” trying to eliminate notions like *late merge*, *multidominance*, *sideward movement*, *parallel merge*, *extension condition*, etc.: [1] satisfy empirical conditions; [2] recursion; [3] NTC, Inclusiveness; [4] binary; [5] minimize comp resources; [6] derive SO with interpretation at interfaces; [7] stability of interpretation. For the desiderata, see also Chomsky (2017c) and Chomsky, Gallego, and Ott (to appear), and for a possible mechanism that can meet the desiderata [4]-[5], see Goto (2015, 2016).

ensure $\langle \phi, \phi \rangle$ labeling of the SPEC-TP construction.³ Then, following Chomsky's (2013: 45) suggestion that so-called SPEC-Head agreement be mandatory for $\langle \phi, \phi \rangle$ labeling, I assume Fukui and Takano (1998) and Toyoshima's (2001) Head-to-SPEC movement hypothesis, according to which all movement is to a SPEC position, whether the movement is of a head or of a phrase (see also Matushansky 2006).⁴ This hypothesis also seems to me to make sense conceptually, in particular in the context of the simplest Merge-based system: if Merge applies freely, as in (4), and if Merge H-to-S is just an instance of the general application of Merge, then there is nothing to preclude the possibility of Merge H-to-S, as is the case of Merge of a phrase to SPEC. Once we accept the simplest Merge-based system, anything goes; rather, any stipulation that blocks it carries a burden of proof.⁵

Only with these assumptions can Chomsky's (2015a) labeling analysis of EPP be maintained: in non-NSLs like English, the SPEC-TP construction is labeled as $\langle \phi, \phi \rangle$ by raising DP to SPEC-T, as Chomsky argues, whereas in NSLs like Italian, it is done by raising of V to SPEC-T with Merge H-to-S, as shown in (5a) and (5b), respectively:

- (5) a. $\langle \phi, \phi \rangle$ labeling via DP-raising to SPEC-T (cf. non-NSLs like English):
 $[\alpha \text{ DP}_\phi [\text{T}_\phi [\dots t_{\text{DP}} \dots]]]$ ($\alpha = \langle \phi, \phi \rangle$)
 b. $\langle \phi, \phi \rangle$ labeling via V-raising to SPEC-T (cf. NSLs like Italian):⁶
 $[\alpha \text{ V}_\phi [\text{T}_\phi [\dots t_V \dots]]]$ ($\alpha = \langle \phi, \phi \rangle$)

Note that raising of V to SPEC-T as in (5b) is not blocked for English: in the simplest Merge-based system, what is possible in a language X is also possible in another language Y, and vice versa. Let us suppose that raising of V to SPEC-T applies in English, the acknowledged way of $\langle \phi, \phi \rangle$ labeling in Italian. Then, V raises to SPEC-T, and in this case, a labeling failure occurs: since V in English has no rich ϕ -agreement, the raised V does not qualify as an element to trigger $\langle \phi, \phi \rangle$ labeling. Thus, in English-type languages, the absence of rich ϕ -agreement marking on V makes $\langle \phi, \phi \rangle$ labeling on T impossible.⁷ In relation to this, it is important to note that raising of V to SPEC-T is not sufficient to trigger $\langle \phi, \phi \rangle$ labeling. This is clear from the fact that raising of auxiliaries AUX to T in English does not cancel EPP. In this case too, the reason is simple: raising of AUX to T is in principle possible, but it lacks rich ϕ -agreement, and hence, it results in a labeling failure.

The conclusion, then, is that the parametric difference in EPP is attributed to visible richness of ϕ -agreement morphology on V (or AUX), and Merge per se is universal but only its lexical

³ This assumption is compatible with Chomsky (2001) and Fukui and Sakai's (2003) approach to parametric variation, according to which parametric variation is restricted to visible morphological (inflectional) elements.

⁴ Goto (2017c) argues that labeling by SPEC-Head agreement can be reduced to labeling by a head-head relation, in accord with the general principle that "minimal search is better than deeper search" (Chomsky 2014: 5).

⁵ Note that Merge H-to-S is not excluded by "Seven Desiderata" introduced by Chomsky (2017b, c) and Chomsky, Gallego, and Ott (to appear). See footnote 2.

⁶ V in SPEC-T must be a more complex element like [V-v*]. The nature of the amalgam has to be clarified, but the argument presented here is perfectly compatible with Chomsky's (2015a: 12) assumption that "although R cannot label, the amalgam [R-v*] can." Whether the amalgam is formed by head-to-head adjunction, as in Chomsky (2015a), or by Head-to-Spec movement, as in Toyoshima (2001), is a question for future research.

⁷ It is well known that French is also a non-NSL like English but has V-raising to T like Italian. This follows given that ϕ -agreement marking on French V is in fact too weak to satisfy EPP, as Biberauer and Roberts (2008) argue.

input varies. The point is whether or not Merge targets an element with rich ϕ -agreement, and whether there is such an element in the lexicon to begin with.⁸

One of the important differences between Chomsky's (2015a) analysis and ours is that the universality of EPP is not maintained in the former, but in the latter it is a consequence of labeling necessity: in our analysis, EPP is just a universal property of T, and, therefore, must be observed in all languages by means of free Merge. Thus, in our theory, there is no need to stipulate the strong/weak parameter on T to satisfy EPP, or to ensure $\langle\phi, \phi\rangle$ labeling, which would be a welcome development, in favor of (1) and (2).

Given this consideration, the potential problem of the asymmetry between T and V with respect to the strength of labeling disappears: T is inherently weak, just like V, and hence need not be parameterized to begin with. Note that "weak" here is nothing other than a cover term for the universal property of the nonphase heads (T, V). That is, the term "weak" suggests that T and V are inherently unspecified as to category and their categorial status is derivative from the phase heads C and v^* . It follows that the nonphase heads lack their labels unless they are selected by the phase heads. This, in turn, allows a uniform characterization of the phase heads and the nonphase heads: the phase heads are the functional elements that are specified as to category, whereas the nonphase heads are the substantive elements that are unspecified as to category. This would be a desirable result for feature inheritance to which I will return.

We conclude, then, that EPP is a universal property of language, holding for both phases CP and v^*P .

Now, the null object phenomenon in (3b) can be accounted for in the same way as in Italian EPP: since the verb *xwar* 'eat' has rich ϕ -agreement (i.e. the third person feminine singular *-a*), the SPEC-VP construction is labeled as $\langle\phi, \phi\rangle$ by raising the verb to SPEC-V with Merge H-to-S. Thus, in Pashto-type languages, the presence of rich ϕ -agreement marking on V makes $\langle\phi, \phi\rangle$ labeling on the SPEC-VP construction possible even without an overt object.⁹

Let's return to the crucial difference between Chomsky's (2015a) analysis and ours, i.e., the way of satisfying EPP in Italian-type NSLs: for Chomsky, strong T has the key, but for ours, it is V with rich ϕ -agreement. Consider the scheme in (6), *t* a copy of V:

(6) [β V [T [α DP, *t*]]] (e.g. "verra Gianni." 'Gianni will come.')

In (6), as we have just said above, β is labeled as $\langle\phi, \phi\rangle$ by raising V to SPEC-T. Note here that the normal labeling procedure can apply to α : by raising V to SPEC-T, what is visible to LA in α becomes D alone, hence α is labeled D, as required. Thus, in our analysis, the problem of how the predicate-internal subject construction is labeled when the subject stays in situ can be circumvented. One may wonder what would happen if the construction in question is transferred to the interfaces with the label of D, not that of v^* , contrary to the standard assumption. However, there is no conceptual argument against this; hence there is no reason to reject what seems to be different from the standard. Rather, labeling the structure in question as D may shed new light on an important observation by Belletti (2001, 2004) that a postverbal subject is interpreted as new

⁸ This implies that Merge is free, labeling is universal, but lexicon formation varies depending on whether the primary linguistic data (i.e. the lexical input) are visible or not, as argued in Fukui and Sakai (2003) (cf. footnote 3). The final point of visibility-sensitive lexicon formation would hold the key to explicating problems of parameters (i.e. language variation).

⁹ Under the proposed analysis of null subjects and null objects, agreement morphemes are treated just like phrases, which would be compatible with Baker (2001). See Goto (2017b) for more details.

information focus: as is clear from the scheme in (6), what is prominent in the remaining structure is the label of the postverbal subject D, so that the postverbal subject might be showing a strong tendency to be interpreted as new information focus at the interfaces.

An immediate consequence of our analysis of Italian EPP is that it naturally explains the fact that the VSO order is impossible in Italian (when focal stress is put on the sentence-final element, which is indicated with capital letters), as illustrated by (7) (cited from Belletti 2001):¹⁰

- (7) *Capirà Gianni IL PROBLEMA
 will-understand Gianni the problem
 ‘Gianni will understand the problem.’

In terms of labeling theory, the impossibility of (7) must be due to a labeling failure arising in the v*P phase. Consider the following v*P phase, t_V a copy of the raised verb *capirà* ‘will-understand’; t_{DP} a copy of the raised object *il problema* ‘the problem’:

- (8) [V(capirà)-v* [α DP(il problema) [t_V [t_{DP} ...]]]]

In (8), α is labeled as $\langle \phi, \phi \rangle$ through relevant procedures, including raising of the object to SPEC-V, ϕ -feature inheritance from v* to V, and V-raising to v*.¹¹ As mentioned above, since V is universally too weak to serve as a label, there must be an overt ϕ -bearing element in SPEC-V to label the SPEC-VP construction as $\langle \phi, \phi \rangle$. Note that if V raises further to SPEC-T, as in (6), then α in (8) will be de-labeled: if V raises to SPEC-T along with ϕ -features, the lower V copy (t_V) becomes invisible to LA, since it is part of a discontinuous element, so therefore, α loses the label of $\langle \phi, \phi \rangle$. To keep the SPEC-VP construction labeled as $\langle \phi, \phi \rangle$, V must stay around v*, in accord with the requirement that SPEC-Head agreement be mandatory for $\langle \phi, \phi \rangle$ labeling. Thus, the impossibility of (7) is ruled out as a labeling failure in the v*P phase level.¹²

Another consequence of our theory gives a straightforward answer to the observation that the *that*-trace effect in English is obviated by a presence of the expletive *there*. Consider (9) below (cited from Rizzi and Shlonsky 2007):

- (9) a. *What_i do you think that [α t_i is in the box]?
 b. What_i do you think that [α there is t_i in the box]?

The ungrammaticality of (9a) has already been explained by Chomsky’s (2015a) theory: the *wh*-subject (here, *what*) must stay in SPEC-T to keep α labeled as $\langle \phi, \phi \rangle$; otherwise, the derivation results in a labeling failure or a PIC violation (see section 1). He does not provide an explanation

¹⁰ Note that the possibility of VSO order in Italian rather depends on certain prosodic properties (cf. Antinuci and Cinque 1977), and the VSO order occurs rather freely in Spanish (cf. Zubizarreta 1998). I leave the details of these properties for future work.

¹¹ Note that these ϕ -features are for abstract object agreement, irrelevant to subject agreement. Here I assume that the amalgam V-v* is formed by head-to-head adjunction, as in Chomsky (2015a). But even if we assume that it is formed by Head-to-Spec movement, as in Toyoshima (2001) (cf. footnote 6), the point will not be affected.

¹² Belletti (2001) notes that VOS is slightly more acceptable than VSO in Italian:

- (i) ?? Capirà il problema Gianni
 will-understand the problem Gianni
 ‘Gianni will understand the problem.’

This would follow if v*P raises to SPEC-T (cf. Richards and Biberauer 2005).

for the grammaticality of an example like (9b), but once we incorporate Merge H-to-S, it becomes possible to account for it without extra stipulation: since α stays labeled as $\langle \phi, \phi \rangle$ thanks to Merge of *there* to SPEC-T, further *wh*-extraction is allowed. This analysis is compatible with Hornstein and Witkoś's (2003) proposal that the expletive *there* is indeed a head, and shares the values of its ϕ -features with its associate by forming a constituent underlyingly, as in $\{D, NP\}$. Details aside, Merge H-to-S actually permits a loophole for further extraction from the criterial position, SPEC-T.¹³

Significantly, this is reminiscent of Rizzi and Shlonsky's (2007) "skipping strategies" – that is, XP can raise further, if the criterial position is occupied by another element. Thus, without recourse to the strength of T, we can now reinterpret Chomsky's (2015a) labeling analysis of ECP in Italian-type NSLs as follows (cf. *Chi_i credi che t_i partira* 'Who do you think that t_i will leave'): the SPEC-TP construction stays labeled as $\langle \phi, \phi \rangle$ thanks to Merge of V to SPEC-T (see β in (6)), hence further *wh*-extraction is allowed. The same analysis is true of the following *que-qui* contrast in French (taken from Rizzi and Shlonsky 2007):

- (10) a. *Quelle étudiante_i crois-tu que t_i va partir?
 which student believe-you that will leave
 'Which student do you believe that is going to leave?'
 b. Quelle étudiante_i crois-tu qui t_i va partir?
 which student believe-you who will leave
 'Which student do you believe QUI is going to leave?'

(10a) is ungrammatical for the same reason as (9a): a labeling failure or a PIC violation. Also, the grammaticality of (10b) can be accounted for in the same way as (9b), particularly in collaboration with Taraldsen's (2001) proposal that French *qui* consists of *que* and expletive *-i*: thanks to Merge of the expletive *-i* to SPEC-T, the relevant SPEC-TP construction stays labeled as $\langle \phi, \phi \rangle$, and hence further *wh*-extraction is allowed.

In this way, once we adopt the possibility of labeling by Merge H-to-S, EPP, ECP, and their parametric differences in Chomsky (2015a), as well as the phenomena that can be captured by skipping strategies in Rizzi and Shlonsky (2007) can be unified under the labeling theory.

Furthermore, the system developed so far makes it possible to capture the parametric difference between overt *wh*-movement languages like English and *wh*-in-situ languages like Japanese under the labeling theory (cf. Fukui 1986, Kuroda 1988, among many others). See (11):

- (11) a. English has *wh*-movement:
 What_i did Mary buy t_i?
 b. Japanese does not have *wh*-movement:
 Mary-ga nani-o kai-masi-ta ka?
 Mary-Nom what-Acc buy-polite-past Q
 'What did Mary buy?'

If *wh*-interrogatives must be labeled as $\langle Q, Q \rangle$ with feature sharing, as argued by Chomsky (2013, 2015a), then the question is: how the SPEC-CP construction is labeled in (11a) and (11b), respectively. In Chomsky (2013, 2015a), it is proposed that the SPEC-CP construction in English is labeled as $\langle Q, Q \rangle$ by raising the *wh*-phrase (Wh) to SPEC-C, as in (12a), but it is unclear how

¹³ See Goto (2017a) for a detailed analysis of *there*-constructions in the labeling theory.

it can be in *wh*-in-situ languages like Japanese. Under the proposed theory, in which head-like elements such as Italian V, Pashto V, English *there*, and French *-i* can ensure labeling by Merge H-to-S, the difference can be derived as follows: in English (11a), the SPEC-CP is labeled as $\langle Q, Q \rangle$ by raising Wh to SPEC-C, as in Chomsky, whereas in Japanese (11b), it is done by raising of the Q-head (particle) *-ka* (Q) to SPEC-C from a *wh*-phrase via Merge H-to-S, as in (12b):

- (12) a. $\langle Q, Q \rangle$ labeling by *wh*-raising to SPEC-C (cf. *wh*-movement languages like English):
 $[_\alpha \text{ Wh}_Q [\text{C}_Q [\dots t_{\text{Wh}} \dots]]]$ ($\alpha = \langle Q, Q \rangle$)
 b. $\langle Q, Q \rangle$ labeling by Q-raising to SPEC-C (cf. *wh*-in-situ languages like Japanese):
 $[_\alpha \text{ Q}_Q [\text{C}_Q [\dots t_Q \dots]]]$ ($\alpha = \langle Q, Q \rangle$)

Thus, in our analysis, the parametric difference in *wh*-interrogatives is attributed to the very existence of the Q-head (particle) (see, among others, Cheng 1991, Tonoike 1992, Hagstrom 1998, Hasegawa 2005 for the precursors of this analysis): particularly in Japanese, thanks to Merge of the Q-head (particle) *-ka* to SPEC-C, the SPEC-CP is appropriately labeled, and a *wh*-phrase can stay in situ.

What is noteworthy is that this analysis predicts that in Japanese, unlike English, raising from *wh* criterial position is allowed. This prediction is borne out. Compare (13) ((13b) is based on Saito 1989 and Takahashi 1993):

- (13) a. English obeys the *wh*-criterion:
 *What_i do you wonder $[_\alpha t_i [\text{C}_Q [\text{John likes } t_i]]]$?
 b. Japanese does not obey the *wh*-criterion:
 Nani-o_i Ken-wa $[_\alpha t_i [\text{C}_Q [\text{Tom-ga } t_i \text{ kata ka}]]]$ siritagetteiru no?
 what-Acc Ken-Top Tom-Nom bought Q want-to-know Q
 ‘What does Ken want to know whether Tom bought?’

In English (13a), since the *wh*-phrase *what* raises from the criterial position, α loses the label of $\langle Q, Q \rangle$, and thus the sentence is ruled out as a labeling failure in the embedded CP. In contrast, in Japanese (13b), since α stays labeled as $\langle Q, Q \rangle$ thanks to Merge of the Q-head (particle) *-ka* to the embedded SPEC-C, further *wh*-movement of *nani-o* ‘what’ from the criterial position is allowed without inducing such a failure.

In this way, if we invoke simple visibles (i.e. head-like elements such as Italian V, Pashto V, English *there*, French *-i*, and Japanese *-ka*) in the simplest Merge-based system, then complex visibles (i.e. parametric variations with respect to overt subjects, *wh*-movement, etc.) can straightforwardly be explained by interaction of the simple invisibles (viz. Merge and labeling).

Let us turn finally to the indispensable device, feature inheritance, which has been a well-accepted mechanism in phase theory ever since Richards (2007) provided a clear rationale for it. He claims that the necessity of feature inheritance from C to T and v^* to V (“inheritability,” in his terms) follows by conceptual necessity from two basic assumptions about the phase-based derivational system: “interpretability” (14a) and “impenetrability” (14b) (= his (1) and (2), respectively):

- (14) a. Value and Transfer of uFs must happen together.
 b. The edge and nonedge (complement) of a phase are transferred separately.

(14a) is the interface condition and (14b) is the motive of PIC. By way of illustration, consider the following CP phase:

(15) [C [α DP [T [...]]]]

Here, since C is a phase head, it follows from (14b) that what is transferred is α . In this situation, if C retains ϕ -features without feature inheritance, Value and Transfer of ϕ -features of C and those of DP cannot happen together at α , failing to satisfy (14a); on the other hand, if C hands ϕ -features onto T with feature inheritance, Value and Transfer of ϕ -features of T and those of DP can happen together at α , being able to satisfy (14a). Hence, feature inheritance is necessary, for Richards, to satisfy these two assumptions.

Bearing this argument in mind, let us turn to Chomsky's (2015a) explanation of why *wh*-extraction from SPEC-T is allowed when C is deleted, as illustrated in (16):

(16) Who_i do you think \emptyset [α t_i [T [β read the book]]] (“who do you think read the book”)

Of particular relevance here is the shift of transfer domain: to circumvent a PIC violation, Chomsky notes that phasehood of C is inherited by T along with ϕ -feature inheritance and is activated on T due to C-deletion. Thus, in (16), T is a phase head, and β is transferred, in conformity with PIC/(14b), so that a PIC violation by further *wh*-extraction (cf. the trace of *who*, t_i in α) is circumvented.¹⁴ It is important to note here that Richards' (2007) argument for feature inheritance from PIC is untenable: in (16), Value and Transfer of ϕ -features of T and those of DP (here, *who*) cannot happen together, just because what is transferred is β , for the reason just given. Thus, unfortunately, Richards' argument for feature inheritance is undermined.

However, this does not mean that feature inheritance is an unwarranted stipulation in the labeling theory; rather it receives a general rationale under our theory. Recall that in the previous section, I have suggested a uniform characterization of the phase heads and the nonphase heads: the phase heads (C, v*) are category-specified elements, and the nonphase heads (T, V) are category-unspecified elements. This characterization suggests that the phase heads C and v* inherently have the ability of labeling (“labelability”) but the nonphase heads T and V do not:

(17) C and v* have labelability but T and V do not.

Note that this view presents a new rationale for feature inheritance in collaboration with Chomsky's (2013: 45-46) interpretation of interpretability in terms of labeling, defined as follows:

(18) “all SOs that reach the interfaces must be labeled” (p. 45); “a label is required for interpretation at the interfaces” (p. 46)

Given (18), T and V must be labeled before they are sent to the interfaces. It follows that feature inheritance is necessary for them to be labeled in conformity with (18): without feature inheritance, they would be unable to qualify for labeling, but with it, they could qualify for it. To put it differently, feature inheritance is an indispensable go-between between the internal property of the nonphase heads and the external requirements from the interfaces. Thus, in our

¹⁴ See section 1 for labeling of α .

theory, the inheritability follows by the two basic principles about the labeling theory – that is, “labelability” (17) and “interpretability” (18):

(19) Feature inheritance is required for labeling for interpretation at the interfaces.

In this way, the status of feature inheritance is maintained and enhanced, provided with a simple principled explanation under the labeling theory. This would be a desirable result for the phase-based system (Chomsky 2017a).

These considerations lead us to the following question: why is it agreement features, not categorial features, which undergo feature inheritance? To understand the nature of the problem, let us consider the CP phase level, and compare a case in which the categorial feature “T” is inherited by T from C with a case in which agreement ϕ -features are inherited by T from C:

(20) [γ C [β DP [α T [...]]]]

In (20), if the categorial feature “T” is inherited by T from C, then α and γ will be labeled as “T” and “C,” respectively, but the intermediate label of β will not be determined because no relation is established between DP and T; on the other hand, if agreement ϕ -features are inherited by T from C, then γ will be labeled as “C,” and α and β will be appropriately labeled as $\langle\phi, \phi\rangle$ and $\langle\phi, \phi\rangle$, respectively, because DP and T can enter into a “relation” in terms of ϕ -agreement. Thus, in order to fully label the structure in question (in accord with the principle of Full Interpretation), features that should undergo feature inheritance must be agreement features, not categorial features, as only the former can establish a “relation” in the course of derivation (see Chomsky 2015b: 81 for relevant discussion). Insofar as this is true, it carries us part of the way towards a principled explanation for why languages should have uninterpretable agreement features at all. That is, uninterpretable agreement features are necessary for labeling SOs fully and completely.

4 Summary and Further Implications

If we adopt Merge H-to-S, a natural consequence of free Merge, then the strong/weak parameter on T, a sharp departure from the minimalist assumptions (1) and (2), can be eliminated, and not only the phenomena Chomsky (2013, 2015a) discusses (i.e. EPP and ECP) but also the ones he does not (i.e. the null object phenomenon, the subject inversion, the *que-qui* alternation, the *wh*-in-situ phenomenon, the feature inheritance, etc.) can be unified under the simplest Merge-based system with labeling theory in a way that fits well into the minimalist assumptions.¹⁵

¹⁵ A question remains how null argument phenomena are derived in languages without overt ϕ -particles (cf. Japanese) and how *wh*-in-situ phenomena are derived in languages without overt Q-particles (cf. Chinese). I have to leave this important question for future research, but see Saito (2016) and Goto (2017b) for relevant discussion on the former.

References

- Alexiadou, Artemis and Elema Anagnostopoulou. 1998. Parameterizing Agr: word order, V-movement and EPP checking. *Natural Language and Linguistic Theory* 16:491-539.
- Antinuci, Francesco and Guglielmo Cinque. 1977. Sull' ordine delle parole in italiano: l'emarginazione. *Studi di Grammatica Italiana* 6:121-146.
- Baker, Mark. 2001. *Atoms of language*. New York. Basic Books.
- Belletti, Adriana. 2001. "Inversion" as focalization. In *Inversion in Romance and the theory of universal grammar*, ed. by Aafke Hulk and Jean-Yves Pollock, 60-90. Oxford: Oxford University Press.
- Belletti, Adriana. 2004. Aspects of the low IP area. In *The structure of IP and CP: the cartography of syntactic structures*, ed. by Luigi Rizzi, vol 2:16-52. New York: Oxford University Press.
- Biberauer, Theresa and Ian Roberts. 2008. Subjects, tense, and verb movement in Germanic and Romance. *Cambridge Occasional Papers in Linguistics (COPiL)* 3:24-43.
- Broekhuis, Hans. 2000. Against feature strength: The case of Scandinavian object shift. *Natural Language and Linguistic Theory* 18:678-721.
- Cheng, Lisa. 1991. On the typology of wh-questions. Doctoral dissertation, MIT, Cambridge, MA.
- Chomsky, Noam. 1981. *Lectures on government and binding: the Pisa lectures*. Berlin: Mouton de Gruyter.
- Chomsky, Noam. 1998. Some observations on economy in generative grammar. In *Is best good enough? Optimality and competition in syntax*, ed. by Pilar Barbosa et. al, 115-127. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89-155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: a life in language*, ed. by Michael Kenstowicz, 1-52. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory: essays in honor of Jean-Roger Vergnaud*, ed. by Robert Freiden, Carlos P. Otero and Maria Luisa Zubizarreta, 133-166. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2013. Problems of projection. *Lingua* 130:33-49.
- Chomsky, Noam. 2014. Minimal recursion: exploring the prospects. In *Recursion: complexity in cognition*, ed. by Thomas Roeper and Margaret Speas, 1-15. New York: Springer.
- Chomsky, Noam. 2015a. Problems of projection: extensions. In *Structures, strategies and beyond – studies in honor of Adriana Belletti*, ed. by Elisa Di Domenico, Cornelia Hamann and Simona Matteini, 3-16. Amsterdam/Philadelphia: John Benjamins.
- Chomsky, Noam. 2015b. A discussion with Naoki Fukui and Mihoko Zushi (March 4, 2014). In *The Sophia lectures (Sophia Linguistica 64), Chomsky (2015)*, 67-97. Tokyo. Sophia Linguistic Institute for International Communication, Sophia University.
- Chomsky, Noam. 2016. Talk given at Universitat Autònoma de Barcelona, November 6. Available at: <https://m.youtube.com/watch?v=NaeTdYnxgMo>.

- Chomsky, Noam. 2017a. Puzzles about phases. In *Linguistic variation: structure and interpretation – a festschrift in honour of M. Rita Manzini*, ed. by Ludovico Franco and Giulia Bellucci. Berlin and Boston: Mouton de Gruyter.
- Chomsky, Noam. 2017b. Talk given at the University of Arizona, February 22.
- Chomsky, Noam. 2017c. Talk given at the University of Reading, May 11. Available at: <https://www.facebook.com/theuniversityofreading/>.
- Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. To appear. Generative grammar and the faculty of language: insights, questions, and challenges. In *Generative syntax: questions, crossroads, and challenges. Special issue of Catalan Journal of Linguistics*, ed. by Ángel J. Gallego, and Dennis Ott.
- Epstein, Samuel. D., Hisatsugu Kitahara, and Danie Seely. 2014. Labeling by minimal search: implications for successive cyclic A-movement and the conception of the postulate “phase.” *Linguistic Inquiry* 45:463-481.
- Epstein, Samuel. D., Hisatsugu Kitahara, and Danie Seely. 2015. *Explorations in maximizing syntactic minimization*. New York and London: Routledge.
- Fukui, Naoki. 1986. A theory of category projection and its applications. Doctoral dissertation, MIT, Cambridge, MA.
- Fukui, Naoki and Hiromu Sakai. 2003. The visibility guideline for functional categories: verb raising in Japanese and related issues. *Lingua* 113:321-375.
- Fukui, Naoki and Yuji Takano. 1998. Symmetry in syntax: merge and demerge. *Journal of East Asian Linguistics* 7:27-86.
- Goto, Nobu. 2013. Labeling and scrambling in Japanese. *Tohoku: essays and studies in English language and literature* 46:39-73, Sendai, Japan: Tohoku Gakuin University. Available at: <https://sites.google.com/site/gotounobu/recent-papers>.
- Goto, Nobu. 2015. Restricting *n* to two: when Merge requires search. Ms., Toyo University. Available at: <https://sites.google.com/site/gotounobu/recent-papers>.
- Goto, Nobu. 2016. Labelability = extractability: its theoretical implications for the free-Merge hypothesis. In *The proceedings of NELS* 46:335-348. Amherst, MA: GLSA.
- Goto, Nobu. 2017a. How to label *there*-constructions. In *English Literature, Regional Branches Combined Issue XII*:1-12.
- Goto, Nobu. 2017b. Labeling theory and parametric variation. Paper presented at Sophia University, March 10, Tokyo. Available at: <https://sites.google.com/site/gotounobu/oral-presentations>.
- Goto, Nobu. 2017c. Notes on search and syntactic visibility. Paper presented at the colloquium “On Phase Theory with Labeling Algorithm” held at the 89th annual meeting of The English Literary Society of Japan, Shizuoka University, Tokyo, May 20. Available at: <https://sites.google.com/site/gotounobu/oral-presentations>.
- Hagstrom, Paul. 1998. Decomposing Questions. Doctoral dissertation, MIT, Cambridge, MA.
- Hasegawa, Nobuko. 2005. The EPP materialized first, agree later: *wh*-questions, subjects and *mo* ‘also’-phrases. In *Scientific Approaches to Language* 4:33-88.
- Hiraiwa, Ken. 2017. On the so called “so called “so called pronouns””. In *A pesky set: papers for David Pesetsky*, ed. by Claire Halpert, Hadas Kotek, and Coppe van Urk, 233-242. Cambridge, MA: MITWPL.
- Hiraiwa, Ken. In press. Labeling roots and pronouns. In *The proceedings of NELS* 47. Amherst, MA: GLSA.

- Hornstein, Norbert and Jacek Witkoś. 2003. Yet another approach to existential constructions. In *Grammar in focus: festschrift for Christer Platzack*, ed. by Lars-Olof Delsing, Cecilia Falk, Gunlög Josefsson and Halldór Ármann Sigurðsson, 167-184. Lund: Wallin and Dalholm.
- Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15:531-574.
- Kuroda, Shige-Yuki. 1988. Whether we agree or not. *Linguisticae Investigationes* 12:1-47.
- Marantz, Alec. 2013. Verbal argument structure: events and participants. *Lingua* 130:152-168.
- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69-109.
- Perlmutter, David. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart and Winston.
- Richards, Mark. 2007. On feature inheritance: an argument from the phase impenetrability condition. *Linguistic Inquiry* 38:563-572.
- Richards, Mark and Theresa Biberauer. 2005. Explaining expl. In *The function of function words and functional categories*, ed. by Marcel den Dikken and Christina Tortora, 115-154. Amsterdam: Benjamins.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, Luigi and Ur Shlonsky. 2007. Strategies of subject extraction. In *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, ed. by Hans-Martin Gärtner and Uli Sauerland, 115-160. Berlin: Mouton de Gruyter.
- Saito, Mamoru. 1989. Scrambling as semantically vacuous A'-movement. In *Alternative conceptions of phrase structure*, ed. by Mark R. Baltin and Anthony S. Kroch, 182-200. Chicago: University of Chicago Press.
- Saito, Mamoru. 2016. (A) Case for labeling: labeling in languages without ϕ -feature agreement. *The Linguistic Review* 32:1, 129-175.
- Takahashi, Daiko. 1993. Movement of *wh*-phrases in Japanese. *Natural Language and Linguistic Theory* 11:655-678.
- Taraldsen, Knut T. 2001. Subject extraction, the distribution of expletives and stylistic inversion. In *Subject inversion in Romance and the theory of universal grammar*, ed. by Aafke Hulk and Jean-Yves Pollock, 163-182. New York: Oxford University Press.
- Tonoike, Shigeo. 1992. Operator movements in Japanese. In *Meiji Gakuin University Ronso* 507:79-142. Meiji Gakuin University.
- Toyoshima, Takashi. 2001. Head-to-spec movement. In *The minimalist parameter: selected papers from the Open Linguistics Forum, Ottawa, 12-23 March 1997*, ed. by Galina M. Alexandrova and Olga Arnaudova, 115-136. Amsterdam: John Benjamins.
- Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. Cambridge, MA: MIT Press.
- Zushi, Mihoko. 2003. Null arguments: the case of Japanese and Romance. *Lingua* 113:559-604.
- Zushi, Mihoko. 2005. Deriving the similarities between Japanese and Italian: a case study in comparative syntax. *Lingua* 115:711-725.

