

## Copy invisibility and (non-)categorial labeling\*

Elliot Murphy · Jae-Young Shim\*\*

(University of Texas Health Science Center · Gyeongsang National University)

**Murphy, Elliot and Jae-Young Shim. 2020. Copy invisibility and (non-)categorial labeling.** *Linguistic Research* 37(2), 187–215. In contrast to dominant views that the labeling algorithm (LA) detects (i) only the structurally highest copy of a moved object, or (ii) detects all copies, we propose and defend a third option: (iii) all copies are invisible to LA. The most immediate consequence of this is that objects formed by Internal Merge cannot serve as labels. We relate this proposal to a particular reinterpretation of LA theory such that LA constructs only categorial labels, barring the construction of  $\langle Q, Q \rangle$  and  $\langle \varphi, \varphi \rangle$  configurations. We then propose an interface condition, Equal Embedding (EE), under which agreeing features must be equally as embedded in order for interpretation to be licensed. We argue that EE appears to fall out of minimal search requirements. We then propose a principled distinction between Agree and LA, based on their sensitivity to copies and interface relations: Both Agree and LA involve minimal search (Probe-Goal for Agree; categorial feature-detection for LA); however, copies are invisible to LA but not to Agree, and LA involves a CI relation (category-specific interpretation) whereas Agree involves an SM relation (the morpho-phonological process of feature-valuation). (University of Texas Health Science Center · Gyeongsang National University)

**Keywords** copy invisibility, criterial position, labeling, minimal search, equal embedding

### 1. Labeling theory

The Labeling Algorithm (LA) in the tradition of Chomsky (2008, 2013, 2015b, 2019a, 2019b) assumes that every syntactic object must be labeled at the interfaces. Syntactic objects formed by Merge are label-less and require a label to be read by the Conceptual-Intentional (CI) interface (Chomsky 2013; Murphy 2015a; Narita 2014; Narita *et al.* 2017). This axiom is formalised in Shim (2014: 463) as the Single Label Condition on Interpretation:

---

\* We would like to thank two anonymous reviewers for their constructive and invaluable comments. The usual disclaimer applies.

\*\* Corresponding author

**(1) Single label condition on interpretation**

An expression must have a single label to be interpreted at the interfaces.

This idea is in keeping with generative assumptions that every syntactic object must have a label to be interpreted (Chomsky *et al.* 2019), with (1) claiming that every constructed set of independent syntactic objects must also have a specific label (see also Fukui 2017; Narita and Fukui 2020).

LA is standardly assumed to search not just for the categorial feature of a unique head that is least embedded in the given structure, but also for any agreeing (prominent) features shared by two equally embedded heads in order to establish the label of an object. From Chomsky (2013) to the present, there is no longer assumed to be a dedicated symbol of representation that denotes categorial status; rather, minimal search detects within a merged syntactic object features determining its categorial status.<sup>1</sup>

The objects that LA searches for, the computational units of syntax, have recently been assumed to be flat or atomic, not being fully-fledged words but rather conceptual representations/roots that syntax assembles (Marantz 1997; Boeckx 2015; Borer 2014). In addition, when these representations are merged they appear to do so across the smallest possible search space (Larson 2015: 60):

**(2) General restriction on Merge**

Merge can only apply to an object in a given space if there is no possible Merge with an object in a more constrained search space.

---

1 Motivations for Minimal Search are not in short supply (Chomsky 1995; see Larson 2015 for comprehensive discussion): For instance, consider Reuland's (2011: 64) argument that a wide variety of binding phenomena, such as the inability of 3rd-person pronominals to enter a chain with their antecedent, can be explained via an economy hierarchy:

(i) Economy of Encoding

Narrow syntax < Logical syntax (C-I interface) < Discourse

Dependencies blocked by some principle at one level of the hierarchy cannot be overridden by operations at a higher level, and since Conditions A and B reduce to Agree and Merge (economy), as Hicks (2009) demonstrates, there is simply no need to resort to discourse-interpretive constraints. As Reuland (2011: 128) puts it: 'If you are in variable binding mode, it is costly to switch to discourse'. For related discussion, see Epstein *et al.* (2014).

This notion is closely related to the “strong hypothesis” entertained in Chomsky *et al.* (2019: 245) that “operations never extend [the workspace]”.<sup>2</sup> The syntactic workspace is inherently ‘flat’ and hierarchical relations between entities within workspaces are only established at the point of interpretation (due to labeling). Merge consequently exhibits a significant degree of optimal design, the apparent centre of a “perfect” (Chomsky 2015a: ix) computational system. Van Gelderen (2018b) even argues that regular patterns of language change can be seen as resolutions to labeling failures, with the syntax-semantics interface imposing major restrictions on how languages develop over the centuries. As such, it is possible - though further cross-linguistic research is needed - to entertain the idea that major features of language change result directly from syntax-semantics and CI requirements. Relatedly, Miyagawa *et al.* (2019) claim, following Saito (2016, 2018), that when Case markers attach to XPs they block labeling and allow the sister element to project instead, with Topic marking having the same function. In contrast, the Q-particle in Japanese allows the C to project. In brief, when a particle attaches to a non-projectable element it induces labeling, and when a particle attaches to a projectable element it blocks labeling. All of these mechanisms are motivated by labeling requirements, and either induce or block labeling, helping to centre LA in a range of environments.

But what of labeling, the operation required for interpretation? Recently, movement and feature inheritance (forms of Merge and direct feature transfer, respectively) have been argued to license the labeling of unlabelable structures.<sup>3</sup> Typically, functional features such as tense,  $\varphi$ -features, Q, and others, are located in C, with T inheriting them from C in C-T configurations and the same process occurring in the  $v^* \text{-} R(=V)$  relation.<sup>4</sup> Let’s now consider (3) to examine a standard

---

2 These ideas are in line with long-supported notions such as the Inclusiveness Condition (Chomsky 1995: 228) since they purely involve the rearrangement of features already in the workspace.

3 We think it is reasonable to derive certain cross-linguistic differences from feature inheritance, i.e., in language X features  $\alpha + \beta$  are inherited by T from C, but in language Y only feature  $\alpha$  is inherited, etc. Carstens *et al.* (2016) note that assuming that feature inheritance is a form of copying (as in Chomsky 2013: 47) can cause crashes for [up] on phase heads. Hence, we will assume that feature inheritance purely involves non-phase heads taking features from phase heads, with no copy left behind at the phase head. This is equivalent to Ouali’s (2008) donate operation. While the topic of feature inheritance will not be central to our discussion here, see Abdelhady (2017) for a recent defence of donate via agreement and anti-agreement in Berber.

4 Like the Agr(eement) head in Chomsky (1995), T is a purely functional head with no lexical content. Indeed, it seems to us that T exists in current theories purely due to tense features and

labeling process assumed in Chomsky (2013, 2015b), where labels/projections such as NP are used only for expository purposes, assuming that phrases refer simply to a set headed by N.

(3) a. The boy bought a toy.  
 b.  $\{\alpha R(\text{=bought}), NP(\text{=a toy})\}$   
 c.  $\{\beta NP, \{\alpha R, \langle NP \rangle\}\}$   
 d.  $\{v^*, \{\beta NP, \{\alpha R, \langle NP \rangle\}\}\}$   
 e.  $\{v^*, \{\beta NP, \{\alpha R, \langle NP \rangle\}\}\}$       ( $\alpha = R, \beta = \langle \varphi, \varphi \rangle$ )

In Chomsky's system, LA (along with other syntactic operations) is hypothesized to apply at the phase level, i.e. LA can be executed once a phase head such as  $v^*$  has been introduced into the derivation.<sup>5</sup> Given this assumption, the first relevant derivational stage in terms of labeling is (3d), where the label of the two SOs,  $\alpha$  and  $\beta$ , cannot yet be determined due to the following additional assumptions in Chomsky (2015b):

(4) a. Category-free roots such as R are (universally) too 'weak' to function as a label.  
 b. R can serve as a label ('strengthen' in Chomsky's terms) only when its Spec is occupied by *an element exhibiting agreeing features with it*.

With (4) in mind, consider now (3e) where the  $\varphi$ -features on  $v^*$  are inherited by R. The two SOs,  $\alpha$  and  $\beta$ , can now be labeled since R has turned into a labelable head thanks to its newly acquired  $\varphi$ -features that agree with NP in its

---

could otherwise be eliminated - although it should be stressed that T could well survive as an independent head, and the fact that it is less featurally enriched than other heads does not automatically rule it out as a syntactic representation. A separate question concerns the locus of Tense; we refer the reader to the discussion in Chomsky (2007: 20), where it is suggested that one advantage of assuming Tense on T is that "T will then have at least some feature in the lexicon, and it is not clear what would be the status of an LI with no features".

5 It is not clear to us, however, whether execution of LA should be delayed until the external argument is introduced, though our discussion does not hinge on this question.

Spec (i.e. satisfying (4b)). Consequently,  $\alpha$  is labeled by the legitimate unique head, i.e. R, and  $\beta$  is labeled as a pair of the agreeing features between R and the head of NP, i.e.  $\langle\varphi,\varphi\rangle$ .<sup>6</sup> In later work, Chomsky (2015b) maintains that  $\langle Q, Q \rangle$  configurations can also serve as labels for the CP domain, in parallel with  $\langle\varphi,\varphi\rangle$  configurations in the TP domain. For example, in  $\{\text{wh-}DP, CP\}$  structures like 'Which dog does Mary like?', the *wh*-DP and C both host the Q-feature, permitting it to be labeled  $\langle Q, Q \rangle$ . We will argue below that neither  $\langle\varphi,\varphi\rangle$  nor  $\langle Q, Q \rangle$  configurations can serve as labels.

## 2. Copy invisibility and the weak/strong distinction

This section will outline and expand on Shim's (2018) proposal concerning the relationship between labeling and copies. We will ultimately develop a phase-level LA with the same general architecture as the LA in the current Minimalist Program (Chomsky *et al.* 2019) and also the revised 'Reading Program' (Chomsky 2019b), but with some modifications to the standard model with respect to its sensitivity to copies. Afterwards, we will contrast these ideas with other proposals in the literature.

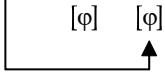
In Shim (2018), it is proposed that the labeling operation of searching for agreeing features, as in  $\langle\varphi,\varphi\rangle$  or  $\langle Q, Q \rangle$  structures (i.e. the Q-feature of a wh-word and the Q-feature an interrogative C), in fact adds an additional computational burden by forcing LA to perform a comparison search (i.e. compare two features,  $F_1$  and  $F_2$ ) alongside the standard Minimal Search (Larson 2015; see also Müller and Sternefeld 1996: 480-481).<sup>7</sup> Responding to this additional computational burden, Shim (2018) proposes that LA searches only for the categorial features of heads. As such, labeled objects such as  $\langle Q, Q \rangle$  or  $\langle\varphi,\varphi\rangle$  (which respectively arise from what has traditionally been termed  $\bar{A}$ -movement and A-movement) become superfluous.<sup>8</sup> For instance, although the  $\langle\varphi,\varphi\rangle$  label

6 Chomsky (2013, 2015b) postulates two ways for LA to identify the label of a given structure: one is to find the least embedded unique head, and the other is to locate agreeing 'prominent' features between two equally embedded heads.

7 By invoking these non-categorial  $\langle\varphi,\varphi\rangle$ -type structures, Chomsky (2013) in effect brought back the 'feature intersection' option he discarded in Chomsky (1995).

8 See Mizuguchi (2019a) for a clear definition of A/ $\bar{A}$ -positions: "The nP is in an A-position if it is

in (3e), repeated here as (5b), is a possibility, it is also possible that the derivation from (3b), repeated here as (5a), could generate (5c) with R raising to  $v^*$  (via 'Head Movement'), yielding the copy  $\langle R \rangle$  invisible to LA. As such, IM of NP to Spec-R in Chomsky's system is unnecessary, and the simplest derivation, (5c), produces no non-categorial labels.<sup>9</sup>

(5) a.  $\{\alpha R(=bought), NP(=a\ toy)\}$   
 b.  $\{v^*, \{\beta NP, \{\alpha R, \langle NP \rangle\}\}\}$       ( $\alpha = R, \beta = \langle \varphi, \varphi \rangle$ )  
  
 c.  $\{[R-v^*], \{\alpha \langle R \rangle, NP\}\}$       ( $\alpha = N$ )

Chomsky's (2015b) (copy-)invisibility (of R) to LA gives rise to another problem when we consider the derivation that follows (5c), where R raises to  $v^*$ :

(6)  $\{[R, v^*], \{\beta NP, \{\alpha \langle R \rangle, \langle NP \rangle\}\}\}$

This notion that lower copies are in some manner invisible to syntactic operations has its roots in 'trace invisibility' (Chomsky 2001), whereby lower copies would not induce intervention effects. Note that Chomsky's copy invisibility requires labeling of  $\alpha$  and  $\beta$  in (6) to precede raising of R to  $v^*$ ; otherwise, labeling of  $\alpha$  and  $\beta$  will both fail as the copy of R, which is invisible to LA in Chomsky's system, can no longer serve as a label. This leads Chomsky (2015b) to stipulate what we see as an unnecessary order between IM (of R-to- $v^*$ ) and labeling (of  $\alpha$  and  $\beta$ ) (see also Bošković 2016).<sup>10</sup>

Another problem with Chomsky's system is that unlike categorial labels like N and V, non-categorial labels such as  $\langle \varphi, \varphi \rangle$  can cause a problem with regard to CI interpretation provided that these labels host both an interpretable feature

merged with an SO headed by a head bearing  $\varphi$ -features; otherwise, it is in an  $\bar{A}$ -position."

9 In order for the derivation in (5c) to be fully feasible, one must also drop the idea tacitly assumed in Chomsky (2015b), i.e. the idea that despite its 'invisibility' to LA, the original copy of R 'somehow' blocks LA from seeking the head of NP instead.

10 A reviewer makes an additional observation that movement of NP to the specifier of the same head as R is banned by Chomsky (2001) and Grohmann (2003). Notice that we are not embracing this position over the moved NP, and will present our own account in the next section.

and an uninterpretable feature, e.g. the interpretable  $\varphi$ -features from N and the uninterpretable  $\varphi$ -features from R (via feature inheritance from  $v^*$ ). Given Chomsky's assumption that labels are necessary information for a syntactic object to be properly interpreted at the CI interface, it is not clear how a label containing a CI uninterpretable feature can serve the suggested function. This is also an observation made by Mizuguchi (2017), who ultimately concludes that the notion of labelability is reducible to interface interpretability (see also Epstein *et al.* 2014; Hiraiwa 2016).

As with the common deployment of  $\langle\varphi,\varphi\rangle$  structures in contemporary LA theory to avoid bare Rs, it has also been assumed that T (in English) is too 'weak' to serve as a label. However, Shim (2018: 31) notes that structures such as (7) pose a problem for this assumption.

(7) a. John seems to like Mary.  
 b.  $\{C \{_\delta \text{John} \{_\gamma T_2\text{-seems} \{_\beta \langle\text{John}\rangle \{_\alpha T_1\text{-to} \{v^*P \langle\text{John}\rangle \text{ like Mary}\}\}\}\}\}\}$

Labeling of  $\delta$  and  $\gamma$  proceeds via, respectively, the agreeing  $\varphi$ -features between  $T_2$  and *John*, and strengthened  $T_2$ , yet the labeling of  $\alpha$  and  $\beta$  becomes problematic since Spec- $T_1$  is only covertly filled and there are no agreeing features between  $T_1$  and *John*. Hence  $T_1$  remains 'weak' or 'unlabelable', and standard LA theory has no way of labeling  $\alpha$  or  $\beta$ .

The core of all the problems discussed above seems to us to lie in the inconsistent treatment of the notion of copy in the standard LA theory. Chomsky (2007 *et seqq.*) contends that IM of X as in  $\{X_1, \{Y, X_2\}\}$  yields '*two copies*' of X, one external to  $\{Y, X\}$ , one within  $\{Y, X\}$ . In other words, it is not the case that  $X_2$  is indeed a copy, while  $X_1$  is something else other than a copy; the two occurrences are both copies of X. Yet copies are treated differently as far as LA is concerned, depending on the structural position they occupy, i.e. the structurally highest copy is visible to LA, while all other copies are not. Resolving these inconsistencies, Shim (2018) proposes the following principle.

(8) Copy Invisibility (to LA)  
*All* copies are invisible to LA.

Copy Invisibility assumes that *both* the ‘original’ element and the ‘moved’ element(s) are equally invisible to LA, such that once an element undergoes IM and is merged elsewhere, LA is no longer sensitive to it.<sup>11</sup> This produces a subtle but clear revision to the syntax-semantics interface: LA is sensitive only to newly-merged, i.e. ‘externally’ merged elements, and transformations on elements are used to provide new ways for detecting these unmoved elements, ‘clearing away’ objects which interfere in minimal search of labels.<sup>12</sup> In conjunction, Shim proposes that LA seeks only categorial features (presumably part of Full Interpretation), attempting to eliminate non-categorial labels from CI and certainly minimise their role in the grammar. In effect, Copy Invisibility leads to categorial-sensitive labeling - in direct contrast to standard LA theory assumptions about copy visibility leading to weak Rs and Ts (e.g. Chomsky 2015b). We should also consider that the weak/strong distinction plainly violates Chomsky’s (2001: 2) Uniformity Principle, which “assume[s] languages to be uniform, with variety restricted to easily detectable properties of utterances”. In related work, Saito (2016) argues that rather than  $\varphi$ -feature agreement it is rather Case valuation that creates the feature-sharing configuration for labeling (see also Mizuguchi 2017 for an intriguing hypothesis that ‘weakness’ is to be cashed out in terms of remaining unvalued features - typically inherited from  $v^*$  to R and from C to T - on heads). Our present proposal leads to the elimination of all such feature-sharing labeled configurations. We are also effectively extending the proposals in Abe (2016), according to whom copying an object leaves the original inactive to further syntactic (hence labeling) computation.

In addition, there is another fundamental problem with the feature-sharing configurations. The original example given by Chomsky (2013) to motivate a

11 A more basic defence of standard copy invisibility can be found in intervention effects.

(i) Which car did the man buy?

The *wh*-phrase moves to the outer Spec- $v^*P$  for transfer of VP before the CP phase formation and  $\varphi$ -features are inherited by T. Yet unless the partially moved *wh*-phrase is considered invisible, it will intervene in the agreement of T and the subject.

12 Notice that Copy Invisibility also forces a reinterpretation of the function of standard transformations/movement operations and centring on their function in avoiding labeling failures. Put differently, IM renders syntactic objects invisible to LA. It may be, therefore, that the scope-discourse properties which are specific to IM emerge as an indirect effect of this design feature of syntax.

feature-sharing LA is an [NP, TP] structure, which Chomsky claims is labelled as T and not N. Labelling this structure as  $\langle\varphi,\varphi\rangle$  does not solve the fundamental issue of interpretation, since this suggests that after the entire derivational process of phase-level LA operations, [NP TP] is nevertheless interpreted as a nominal, with interpretable  $\varphi$  being a property of NPs/DPs. T, meanwhile, hosts uninterpretable  $\varphi$ , and cannot therefore reach the CI interface to serve as a label as part of any given feature-sharing configuration. This renders the elegant simplicity of  $\langle\varphi,\varphi\rangle$  to the less elegant  $\langle\text{up},\text{ip}\rangle$  structure, not suitable for a TP.<sup>13</sup>

What are the consequences of these proposals? Rather than stipulating that R is too weak to label, as in standard LA theory, the claim that LA seeks only categorial labels allows us to assume that R is unlabelable not because it is weak and requires feature inheritance to become part of a  $\langle\varphi,\varphi\rangle$  structure, but simply because it lacks a categorial feature (and will only be assigned one when merging with a category-defining functional head like *n*, *a* and *v\**; see Embick 2012).<sup>14</sup> In addition, T can now invariably serve as a label irrespective of its finiteness. Furthermore, there is no need to stipulate an ill-defined, vague notion of 'prominent' features for LA, such as Q or  $\varphi$ , and traditional categorial features in combination with Copy Invisibility suffice to generate LA (see also Abe 2016 and Mizuguchi 2019c for further discussion of unlabelable R).<sup>15</sup>

Reevaluating the landscape, consider again (3b=9b), which as mentioned can only yield the derivation in (3c=9c) in the standard LA theory. Given what we have outlined, (9b) can now generate either (9c) or (9d). We use outline font to uniformly indicate invisibility of (all) copies:

(9) a. The boy bought a toy.

13 A reviewer asks why we are choosing to distinguish (un)interpretable features the way we have. We acknowledge that the status of uninterpretable features and their heads is an issue yet to be resolved, and we have assumed that CI reads out only interpretable features.

14 It is standardly assumed that in [R/T, XP] configurations, with R being a verbal root, neither R nor T act as phase heads and so either XP or a category within XP needs to move to Spec-R/T, with featural commonality between these elements (e.g.  $\varphi$ ) implementing feature valuation. The result is a  $\langle\varphi,\varphi\rangle$  label, under the assumption (that we also maintain here) that labeling follows valuation.

15 'Prominence' has been invoked as a semantic and also syntactic notion yet a method of adjudicating between these (and distinguishing between them in principle) has not been forthcoming.

- b.  $\{\alpha \text{ R} (= \text{bought}), \text{a toy}\}$
- c.  $\{v^*, \{\beta \text{ a toy}, \{\alpha \text{ R, a toy}\}\} \quad (\alpha, \beta = \text{R})$
- d.  $\{v^*, \{\alpha \text{ R, a toy}\}\} \quad (\alpha = \text{R})$

Under this analysis, it only takes the category-defining  $v^*$  to merge in order for R to serve as the label.<sup>16</sup> Consequently,  $\alpha$  can be labeled R irrespective of IM of NP; both options are permitted here, with both IM of NP or no IM of NP licensing the ultimate labeling of R. Nevertheless, we also acknowledge that movement of 'a toy' in (9c) may not even be needed if we assume a ban on domain-internal movement (*à la* Chomsky 2000; Grohmann 2003); in which case, feature inheritance alone would suffice.

More generally, our Copy Invisibility proposal also ensures that lower copies from YP in [XP,YP] structures do not interfere with the labeling of XP, for the same reasons as in Chomsky's (2013) model. To our knowledge, the proposed differences between lower and higher (and intermediate) copies typically suggested in the literature violate assumptions of efficient computation and are purely stipulative, not able to be derived from any component of the syntax. For instance, Miyagawa *et al.* (2019: 3), following Fox (2000, 2002), claim that the lower copy "would retain just enough information for being interpreted as a copy/trace". What precisely the content of this copy-identifying information would be is not explored further, and the mechanism which would be responsible for retaining "just enough" such information is unclear.<sup>17</sup>

Reconsidering now (7b), repeated here as (10b), the present analysis can also provide a label for  $\alpha$  and  $\beta$ , with  $T_1$  providing a clear categorial label for both

16 We assume that some form of feature inheritance (from  $v^*$  to R) provides R with a category feature.

17 Interestingly, a reviewer points us towards the work of Citko *et al.* (2018) who discuss non-agreeing subjects in languages such as Polish. Non-agreeing subjects for us (but not Chomsky 2013) can move to Spec-TP, where, under our assumptions, they are invisible to LA. In which case, we would assume they are labeled due to T. Citko *et al.* (2018) claim that non-agreeing subjects could also move directly to a higher position and never stop at Spec-TP. For instance, with respect to locative and predicate inversion, the position would be Spec-TopP, where the moved element shares a Topic feature with TopP. This would naturally permit a  $\langle F, F \rangle$  labeled configuration. In response, we would argue that the Topic feature can remain on the subject and T can serve as the appropriate label, and no shared featural configurations are needed. Since the moved element is invisible to LA for us, the Topic features can be read out at CI but not interfere with the simplest labeling solution, i.e. T.

in the absence of *John* in its Spec (considered invisible to LA).

(10) a. John seems to like Mary.  
 b. {C { $\delta$  John { $\gamma$  T<sub>2</sub>-seems { $\beta$  John { $\alpha$  T<sub>1</sub>-to { $\nu^P$  John like Mary}}}}}}  
 (α, β, γ, δ = T)  
 c. { C { $\delta$  John { $\gamma$  T<sub>2</sub>-seems { $\alpha$  T<sub>1</sub>-to { $\nu^P$  John like Mary}}}}}}

What's more, Copy Invisibility permits two different scenarios; one in which *John* cyclically moves to every Spec-TP (i.e. (10b)), and the other in which *John* moves to matrix Spec-TP in a single move<sup>18</sup> (i.e. (10c)) - neither option appears to cause labeling problems.

In what follows, we will maintain that only the categorial features of heads are relevant for label-based interpretation at CI (something we regard as a tautology; Murphy 2015a, 2015b), but we will argue that a core distinction exists between, on the one hand, CI-specific categorial labels (which we claim are well-motivated and necessary), and, on the other hand, the postulated non-categorial labels such as  $\langle\varphi,\varphi\rangle$  and  $\langle Q,Q\rangle$  (which we will argue should be dispensed with).<sup>19</sup> For example,  $\langle\varphi,\varphi\rangle$  structures have been argued to play a role in marking ' criterial positions' (Rizzi 2016), such that the NP that moves is 'frozen' and can no longer move if it merges in a Spec position of an object labeled as  $\langle\varphi,\varphi\rangle$ . Reformulating these ideas within LA theory, Chomsky (2019a) further states that "a criterial position is an XP-YP structure where the labels of XP and YP agree", such as *wh*-interrogatives and subject-predicate structures. Going against this account, we will explore here to what extent a Copy Invisibility analysis (tied closely to assumptions of category-based labeling, licensing R and T as labels) can provide an alternative perspective on the nature of criterial freezing and other syntactic phenomenon, without recourse to non-categorial labels.

18 See, among others, Chomsky (2001), who suggests a possibility where IM of *John* takes place in one fell-swoop to the matrix T in constructions such as (10a).

19 We therefore follow Hornstein (2009) and Cecchetto and Donati (2015) in assuming that labeling is necessary for both syntactic computation and CI interpretation. Chomsky (2013: 37, emphasis ours) also states that labeling is for "identification of the *category* of a phrase".

### 3. Routes to labeling

#### 3.1. Criterial positions as markers of feature valuation

In a framework that we will argue complements our own, Hosono (2018) proposes that IM is not free (*contra* Chomsky 2013) and that the long-debated Criterial Position (Rizzi 2006, 2015) is the position in which a raised category completes the valuation of unvalued features (e.g. [*lCase*], [*up*]), triggering Transfer to CI.<sup>20</sup> This *Halting Problem* can be demonstrated below, where the *wh*-object moves from its originally merged position to Spec-(embedded)CP in order for feature valuation to take place between [*lQ*] of *which dog* and [Q] of C<sub>Q</sub>. When it tries to move out of this criterial position (CriP), the sentence is ungrammatical.

(11) a. You wonder [CP [Q which dog] C<sub>Q</sub> John likes [<sub>e</sub>-which dog]].  
b. \*[<sub>a</sub> [Q which dog] do [<sub>B</sub>= $\varphi\varphi$ ] you wonder [<sub>y</sub>= $QQ$ ] [<sub>e</sub>-which dog] C<sub>Q</sub> [John likes [<sub>e</sub>-which dog]]]]?

It is argued in Hosono (2018) that in (11b) feature valuation occurs first between the verbal head *likes* and *which dog* assigning the latter accusative Case, but since *which dog* retains its [*lQ*], it moves to the embedded Spec-CP.<sup>21</sup> *Wonder* subcategorizes as *wh*-clause and so the embedded C hosts [Q], and *which dog* and C<sub>Q</sub> proceed to feature valuation, labeling  $\gamma <Q,Q>$ . *Which dog* completes valuation of its own unvalued features and so cannot move further. However, under the system we are developing here, (11a) is interpretable not just because of the *wh*-object valuing its features, but because C<sub>Q</sub> provides a clear label, as shown in (12), with Copy Invisibility leading both occurrences of *which dog* to be invisible to LA.

20 Hosono is effectively embedding some older claims within the LA framework: Consider Epstein's (1992) analysis that a *wh*-phrase cannot move out of the Spec of the embedded C hosting [+*wh*], or Bošković's (2011) analysis that after an uninterpretable *wh*-feature is checked in the intermediate Spec it cannot move up (here, we assume only valuation, not interpretability).

21 It is not clear in Hosono's system what prevents feature valuation from taking place in situ, especially between the Q feature of *which dog* and that of C.

(12) You wonder [ $\gamma = C$  [which dog]  $C_Q$  [John likes which dog]]?

Along with Spec-(embedded)CP, Spec-TP is claimed to be another CriP. What Hosono (2018) observes is that these two positions share something in common: They both are slots where raised elements have their unvalued features valued due to (some form of) probing.<sup>22</sup> This process then bars further movement in conjunction with Copy Invisibility forcing LA to encounter an unlabelable structure. What to make of this? A possible way to integrate Hosono's observation is to assume that elements seeking feature-valuation via raising ultimately play an interpretive role at CI but only as part of a labeled structure headed by a distinct element not subject to movement.<sup>23</sup>

In order to attempt to capture these effects in a more principled manner, we would here like to propose a (CI-)interface condition, *Equal Embedding (EE)*, which induces particular effects in narrow syntax.

(13) **Equal embedding**

Agreeing features must be equally embedded.

Syntactically speaking, what (13) requires is that once a feature is 'next to' another feature of the same type (e.g. [... [ $\iota\phi$ - $\iota\phi$ ] ...]), these features are required to remain together and cannot be separated, since this would interfere with an interpretation procedure with minimal computational effort. Essentially, this means that an operator can only be interpreted as such by CI if it occurs in a [ $\iota Q$ - $Q$ ] configuration. As such, under our account halting effects arise due to the CI-localised embedding restriction. Further, structures standardly analysed as  $\langle\varphi,\varphi\rangle$  configurations are, under our system, labeled as whatever categorial feature merges with it, with EE requiring that these  $\varphi$ -bearing objects be syntactically embedded at the same hierarchical level. More broadly, EE appears

22 One could say that Spec-Head agreement is 'back', then, in a different form. Relatedly, a moved category needs to host some unvalued feature for it to enter valuation with a head in its raised position (Hosono 2018: 52). Likewise, adverbials do not move out since they do not host unvalued features in the unmarked case.

23 See also Hosono (2016) and Richards (2016) for an argument that constraints on movement are imposed by PHON/Sensorimotor System (SM). This suggests that movement from/into the criterial position is allowed iff movement is required by phonology.

to fall out of minimal search requirements, and is hence a good candidate for a third factor requirement at CI.

Notice that our proposal also grounds principles of efficient interpretation into CI, rather than purely into the syntax. A standard position in generative grammar is that semantics is effectively a form of syntax since it involves symbol manipulation (e.g. Chomsky 1995). Indeed, Pietroski (2018) proposes some efficiency principles at CI, such that the ‘instructions’ to build concepts at CI are interpreted with notable simplicity and minimal computational effort, through Pietroski’s M-join and D-join (see also Al-Mutairi 2014; Collins 2020 and González Escribano 2005 for discussion of simplicity at the semantics interface; in particular, González Escribano 2005: 57 discusses “the role of Economy throughout” much of the core language architecture, most notably semantics, leading to a “semantocentric minimalist grammar”). Likewise, Hauser *et al.* (2002: 1574) already noted that ‘at least’ syntax involves computational efficiency, pointing to the possibility of other components of language being organized along these lines. Thus, we see our proposal as being faithful to this architectural standing.<sup>24</sup>

In combination with Hosono’s (2018) assumption that labeling results from feature valuation (except when LA takes a phase head), our Copy Invisibility framework provides a simple account of the phenomena explored in contemporary labeling theory: Merge remains a ‘free’ operation but once feature valuation occurs, invisible copies are formed and all categorial elements can serve as labels for interpretation.<sup>25</sup> CriP is ‘marked’ by feature valuation by

24 A reviewer notes that our positioning of uninterpretable features here seems counter to our claim that EE is a CI condition, since by definition CI does not interpret the uninterpretable. We readily acknowledge that the architectural positioning of EE may not be as directly assumed here and that it may more appropriately be characterised as a principle of efficient syntactic computation (possibly related to the memory buffer or workspace). One potential solution is to assume that  $uF$  as in  $\langle uF, F \rangle$  cannot help but be handed over to CI because  $uF$  is fixed with  $F$ ; it cannot be separated from  $F$ . In addition, CI does not ‘interpret’ (the value of)  $uF$ ; rather, for us it checks whether the agreeing features are equally embedded.

25 What’s more, under standard LA theory assumptions, it is unclear why only the moved copy must serve as the label, rather than the originally merged copy, in particular given that one of the original motivations for copies under minimalism was the idea that they serve crucial interpretative roles. Hence, one would assume that this factor would motivate LA to be sensitive to the original copy - and, if anything, more sensitive. Nevertheless, we concur with Chomsky *et al.* (2019) that those objects constructed in the syntax which are not appropriately interpreted at CI

default and  $\langle Q, Q \rangle$  structures do not in fact directly mark criterial positions (and even if they did, this would only be an epiphenomenon of the valuation procedure, not an inherent feature of non-categorial labeling). Yet, as we will now explore, while all categorial items can serve as labels, CI imposes its own demands on which ones get filtered out.<sup>26</sup>

### 3.2 Some alternatives to current solutions

Mizuguchi (2019b) argues that the labeling ambiguity in [XP,YP] structures - the cornerstone of LA theory - can be tolerated and result in no labeling failure. It is argued that the configuration can be labeled either X or Y, arising from economy, and that the well-formedness of labeling is due to CI, with syntax “not caring about the outcome of labeling” (Mizuguchi 2019b: 1). As such, the ambiguity will end up at CI, not in the syntax, and since it is common for semantics to exhibit ambiguity in interpretations it appears unproblematic for it to also be presented with ambiguous labels. Mizuguchi’s (2019b) elegant proposal results in an appreciation of categorial labeling, which we intend here to explore further.

Evidence for this position can be found in [XP,YP] cases in which neither XP nor YP moves out of [XP,YP] (yielding X or Y as the label) and agreeing heads also do not appear (yielding, say,  $\langle \varphi, \varphi \rangle$  as the label), and yet the structure is felicitous. To take one of the four cases Mizuguchi (2019b) presents (with the others being partial *wh*-movement, object shift, and in-situ subject constructions), consider sentences with non-nominal or non-*nP* subjects. Subjects can sometimes be non-*nPs* such as *that*-clause subjects in (14a) or prepositional subjects in (14b), which do not agree in  $\varphi$ -features with T (Emonds 1976; Stowell 1981; although see McCloskey 1991 for evidence for agreement).<sup>27</sup>

---

will be filtered out.

26 For Gallego (2018a), it is suggested the freezing takes place not just because of feature valuation, but because of CI demands: Movement of a *dXP* (=dislocated XP) out of a phase edge can cause an absence of relevant discourse-interpretation of the edge at CI (e.g. topic, focus) since interpretations of theta-roles and criterial-roles (and so on) cannot accumulate.

27 They receive unmarked third person, singular interpretation at CI and do not host  $\varphi$  (see Preminger 2014).

(14) a. [That the world is not flat] was demonstrated by Columbus.  
 b. [After four] would be a good time to meet.

Subjects in (14) are unable to agree with the matrix T. The embedded CP in (14a), for example, does bear  $\varphi$ -features (of NP *the world*) but these features agree with the embedded T, not with the matrix T. Further, it could also be the case that the NP *the world* with  $\varphi$ -features would have been transferred (hence inaccessible to the matrix T) by the time the matrix T is introduced.<sup>28</sup>

Mizuguchi notes that since non-nominal subjects are in pre-verbal positions and the EPP is forced in English, they should be seen as occupying Spec-TP, like standard *nP* subjects. Thus, they form [XP,YP] configurations, as below, but the subjects do not agree with T.

(15)  $[\gamma [\alpha X [YP]], [\lambda T [vP]]]$

The structure above is interpretable, yet poses a problem for standard LA theory. For Mizuguchi, this suggests that labeling can be ambiguous and that in fact it would require further stipulation to prevent labeling ambiguity, such that LA would be forced to choose either X or Y in [XP,YP], rather than settling for either which would lessen computational load. This also keeps to core minimalist assumptions via SMT-based syntax: As long as third factor principles are adhered to, we should not require any further stipulations about forced movement or non-categorial featural agreement. Well-formedness is thus a CI notion due to various interface conditions, irrelevant to the syntax (for reasons going back to Chomsky 1956). Consequently, LA can label (15) as either X (i.e. C, *p* or *n*, depending on the type of subject) or T, yet the outcome is evaluated at CI. When  $\gamma$  in (15) is merged with C, becoming its complement, CI will dictate that  $\gamma$  be labeled T, since C selects T but not any of the X objects. If  $\gamma$  is labeled X, then the C-T-v structure will be inferred and clausal interpretation

28 As far as we can understand it, Mizuguchi (2019b) attributes unavailability of  $\varphi$ -features for the matrix T in (14a) to unavailability of the  $\varphi$ -features on the embedded C. That is, Mizuguchi argues that no  $\varphi$ -features for the matrix T are left available on the embedded C because the features would have been inherited by the embedded T and eventually transferred by the time the matrix T is introduced. Notice, however, that the  $\varphi$ -features on C are 'uninterpretable' and therefore unable to agree with the matrix T even if they were to be located by the matrix T.

will not be established.<sup>29</sup> As such, LA remains a minimal search procedure but CI conditions impose their own demands for which an ambiguous label is selected; that is, those demands amount to our interface condition EE in (13). This strikes us as a coherent framing given that labels are ultimately established only for reading off at CI.

Focusing on EE, the insight we would like to present here is that this CI condition can also explain CriP phenomena: *nQ* must be next to or equally embedded as *Q* in order for *wh* and/or  $C_{[Q]}$  to be properly interpreted as an operator and/or  $C_{[Q]}$ . This in turn demands a particular syntactic (set-theoretic) configuration, which is violated by moving one of the featural elements beyond the equally embedded site.

Considering the case of in-situ subjects, Mizuguchi (2019b) argues that LA can label the in-situ structure as either *n* or *v\**. However, CI will impose its own selection such that only a *v\** label will satisfy the selectional relation of T (i.e. the C-T-*v\** sequence), which does not select an *nP*. This keeps to our proposals that categorial labels alone are sufficient for interpretation and that Copy Invisibility would predict that *v\** would be selected as the label given that it does not undergo movement.<sup>30</sup>

Turning to partial *wh*-movement, this can generate an [XP,YP] structure in which the embedded C and the *wh*-phrase (or *n* or {*whP*}) do not agree, and no movement takes place to yield an invisible copy.

(16) a. *Was* meinst du [γ **wen**; C [Peter Hans *t* vorgestellt hat]]?  
 'Who do you think Peter has introduced to Hans?'  
 (modified from Mizuguchi 2019b, (40a))

b. [γ [*whP*], [CP]]

As illustrated in (16b), LA will detect *n* (in *whP*) and C, and hence the configuration can be labeled either *n* or C. However, CI will dictate that it must

29 Mizuguchi (2019b) contends that 'clausal interpretation at CI requires C, T and *v\** to be merged as C-T-*v\**'.

30 A reviewer notes that, dropping the assumptions about C-T-*v\** configurations, both N/D of the subject and *v\** can serve as the label under our model, being categorial. We acknowledge that there may therefore be elements of unresolved labeling ambiguity in our model, which we intend to explore in future work.

be labeled C, since an *n* label would yield a “nominal” phrase rather than a “clausal” phrase in a configuration in which the verb selects CP, not *nP*. CI properties will only be satisfied with the C label, to satisfy selectional relations between sets.

We think this position is compatible with our assumption that CI demands categorial labels and that both T and R can serve as immediate labels in ambiguous labeling configurations, countering any claims of labeling failures. This does not, of course, mean that CI will always license any X or Y label, merely that it can do in the event of no structural prominence. In brief, T will label  $\lambda$  in (15) without requiring any movement to ‘strengthen’ it (contra Chomsky 2015b), and X will immediately label  $\alpha$ , leaving  $\gamma$  to be read by CI as T if  $X[YP]$  is a moved copy (hence invisible). If  $X[YP]$  is base-generated in Spec-TP then moves, it, too, becomes invisible to LA. Nevertheless, no agreement is needed, and in this case no  $\langle\varphi,\varphi\rangle$  configuration is even generated, certainly not read at CI. The study of labeling here turns more towards the study of CI conditions (such as EE) rather than the specifics of LA itself, which remains purely an implementation of minimal search.

### 3.3 Using copies to distinguish labeling from agreement

Where we differ from Mizuguchi (2019b) is in relation to Copy Invisibility. Mizuguchi explains the ill-formedness of (17a) in terms of the standard weak T analysis, noting that only in the event of a merged *nP* at  $\gamma$  will T be strengthened in English to license a well-formed interpretation. However, for us, no weak T stipulations are needed, and the invisibility of both the Spec- $\alpha$  copy of *the student* and the initial copy ensure that  $\alpha$  cannot be assigned an appropriate label to then merge with  $\gamma$ . It is only when  $\gamma$  is merged and *the student* undergoes successive-cyclic movement to generate (17b) that feature valuation can proceed and  $\delta$  is labeled TP.

(17) a. \*[ $_{\gamma}$  Seems to be likely [ $_{\alpha}$  the student [to [ $t_i$  understand the theory]]]].

b. [ $_{\delta}$  The student [ $_{\gamma}$  seems to be likely [ $_{\alpha}$   $t_i$  [to [ $t_i$  understand the theory]]]]].

Indeed, Mizuguchi (2019b) goes so far as to claim that ambiguous labeling motivates the elimination of any form of copy invisibility. He claims that copy invisibility is a stipulation requiring a principled explanation. Considerable empirical coverage is presented in Shim (2018) to defend the notion, but with respect to a more principled explanation, addressing Mizuguchi's concern, we suggest that here the common (but typically vague) notion of prominence in LA theory can support the principle of Copy Invisibility we are defending here. Since they have been manipulated in the workspace, it is likely that copies are afforded a distinct degree of prominence to certain elements in the syntax, such as LA. Workspace manipulation yields a degree of computational de-centering such that LA will detect non-copies (more accurately, non-moved elements, given that all syntactic objects are strictly speaking copies) over copies, in accord with minimal search (an issue we return to below). Of course, all copies (and repetitions) are vital for CI interpretation, but not for labeling.<sup>31</sup>

An alternative account for (17) is to assume that copies are invisible to LA, but not to Agree. EE also requires that these agreeing features be equally embedded; that is, the matrix T in (17a) will agree with 'the student' but the two agreeing  $\varphi$ -features are not equally embedded, in violation of EE. We believe this to be a core distinction between these otherwise similar operations: Even though both are implementations of minimal search, Agree is sensitive to copies whereas LA is not; indeed there are cases of Agree with lower copies (e.g. Holmberg and Hróarsdóttir 2003). Perhaps this is because Agree is solely an operation computing feature-detection and is characterised by notions such as Probe-Goal (which naturally involve close scrutiny of an SO's features), whereas LA is also centred on feature-detection but not the direct *relations* between multiple elements and is rather concerned with using the features of a single element in the service of economical interpretation at CI. Indeed, languages contain a rich range of functional elements that aid in either agreement or labeling (Miyagawa *et al.* 2019), and are separable along the dimensions carved

---

31 It may also be the case that a generic 'look ahead' procedure bars LA from selecting a copy as the label due to the regularity of successive-cyclic movement, i.e. effects of movement frequency would prime LA to hold off on using a moved element as a label given that the next phase might see it undergo further movement. Nevertheless, this is relatively weak explanation, and so we will present a more comprehensive account below.

out by these two processes. Furthermore, LA is also concerned with categorisation above and beyond the current workspace and set of linguistic features; LA is concerned with generating *sets of phrases*, whereas Agree is concerned with identifying *sets of features*. Our proposal also lessens the burden on 'first factor' requirements, since no additional Search-related design mechanism needs be uniquely attributed to LA.<sup>32</sup>

Gallego (2018b: 608) also notes that even in an [XP,YP] structure like (18), Agree would not be able to detect simultaneously the features on x and y due to a lack of c-command relationship.

(18)  $\{\{\mathbf{x}, \{\dots\}\}, \mathbf{ZP}\}, \{\mathbf{y}, \{\dots\}\}, \mathbf{WP}\}$

The only way x and y would communicate is if we invoke more complex notions than c-command, like percolation, which is not in keeping with the minimal search nature of Agree or LA.<sup>33</sup>

In this connection, we see an inherent symmetry between Agree and LA: the former involves minimal search (via Probe-Goal relations) and an interface process (feature-valuation; which, following Bobaljik 2008 and Epstein *et al.* 2019, is reassigned to the morpho-phonological component). The latter also involves minimal search (via categorial feature-detection) and an interface process (category-specific read-out at CI). We would therefore like to propose that Agree and Labeling should be separated not by their minimal search component - which are both realisations of the same fundamental computation - but should

32 A reviewer notes that there are also cases of copy-transparency for Agree, which we assume can be derived from mechanisms and features quite independent of the architecture we are discussing here; that is, the existence of Agree-sensitive copies and non-Agree-sensitive copies will be a separate issue centred on the broader role of agreement in the syntax. Nevertheless, we acknowledge that our account may need to provide a more direct encounter with this issue, which we reserve here for reasons of space for future research.

33 Interestingly, van Gelderen (2018a) reanalyses specifiers as heads, and heads as higher heads. In essence this extends the scope of what items can serve as heads. In this sense it is in accord with our present claim that T and R can act as labels in standard cases of labeling failure. This is argued to aid simple search for LA, such that phrases can act as heads: van Gelderen reanalyses demonstratives as C, D and T heads, *wh*-elements as C heads, Adverb Phrases as ASP heads, negative adverbs to Neg heads, and PPs as C heads. The motivation for this is that no stipulations will be needed to ensure that LA can search within lexical items and feature bundles, but can more simply detect a head.

rather be separated by their interface relations (Agree to SM, Labeling to CI) and sensitivity to copies (see Carstens 2016 for Agree and minimal search implementations). As Chomsky (2015b: 6, emphasis ours) notes, “since the same labeling is required at CI and for the processes of externalization (though not at SM, *which has no relevant structure*), it must take place at the phase level”.

Table 1. Asymmetry between Agree and Labeling Algorithm

	<b>Agree</b>	<b>Labeling Algorithm</b>
<b>MINIMAL SEARCH</b>	Probe-Goal	Categorial Feature
<b>INTERFACE</b>	SM (Feature-Valuation)	CI (Category Interpretation)
<b>COPIES</b>	Visible	Invisible

#### 4. Conclusion

As far as we can tell, there are at least three logically possible models of copy visibility.<sup>34</sup> We have here defended (iii):

- (i) All copies are visible (e.g. Mizuguchi 2019b).
- (ii) The highest copy is visible (e.g. Chomsky 2013, 2019a).
- (iii) No copies are visible.

In addition, Takita *et al.* (2016) point out that Chomsky’s basic idea behind the implementation of labeling through movement, which assumes that a copy left behind by movement is invisible to minimal search, is not only incompatible with the copy theory of movement but also violates the No Tampering Condition: It demands that after IM the properties of the lower copy need to somehow be altered to bar detection.

We believe our model fits well into some broader themes noted in the literature; for instance, syntactic derivations more generally seem to be a cyclic process of symmetry-breaking and symmetry-formation, with the ultimate state typically being one of symmetry (for clear motivation, see Narita *et al.* 2017). The

<sup>34</sup> We ignore other possibilities that are logically conceivable but we think are highly unlikely, such as ‘only the original copy is visible’ or ‘only (some of) the intermediate copies are visible’ (e.g. as in Kitahara 2018).

standard distinctions between External Merge (EM) and Internal Merge (IM), and the lexical relations (predicate-argument structure, selection etc.) and discourse relations (quantificational, topic-focus etc.) they each derive, can also be seen respectively as exhibiting asymmetric and symmetric relations.<sup>35</sup> We would like to suggest that there is also a layer of asymmetry *between* EM and IM alongside the standardly-discussed symmetries, such that only the former generates LA-sensitive objects. The standard formulation of copy invisibility assumes that only the original copy is invisible to LA, yielding a clear and empirically unjustified asymmetry, while the principle of Copy Invisibility we defend here exhibits an asymmetry between all copies (invisible) on the one hand, and non-copies and repetitions (visible) on the other.<sup>36</sup> Notice that we are not claiming that copies are unimportant for interpretation at CI, which for reasons of, for instance,  $\theta$ -interpretation, they are. Richards (2019) has provided additional motivations for assuming this form of EM-centric labeling algorithm.<sup>37</sup> In particular, Richards assumes a simpler model of exocentric labeling than Chomsky's (2013) in which all labels are uniformly added via EM of categorizing phase heads, turning unlabelled (and hence uninterpreted) non-phase syntactic objects into labelled phrases. He further suggests that all phase heads (e.g.  $\eta$ ) can

35 Narita *et al.* (2017) also extend the symmetry-forming tendencies found in Agree operations to formal syntactic features more generally.

36 We assume that invoking phase-level memory can yield the distinction between copies and repetitions, involving comparing distinct workspaces as the derivation proceeds; although see Collins and Groat (2018), who acknowledge this possibility but point to some apparent difficulties, i.e. it requires that the interfaces are sensitive to multiple workspaces rather than simply syntactic outputs. However, we merely need to assume that the interfaces *can* be sensitive to the structure of multiple workspaces in the event of potential crashes yielded by multiple elements of the copy/repetition variety - we do not need to assume a radical revision to what structures we assume are regularly transferred to the interfaces. Indeed, if CI can permit ambiguity of labels but imposes its own demands on which label is ultimately selected, then we see no considerable difficulty with the assumption that the interfaces can also sustain ambiguity in syntactic elements like copies vs. repetitions and that they use multiple configurations of workspaces to adjudicate between them.

37 In addition, as explored in Murphy (2015a), crossing dependencies of the kind generated by IM ( $A^n C^m B^p D^q$ ) can be interpreted with a mildly context-sensitive grammar and a linear bounded automaton, while EM produces the less memory-intensive nested dependencies ( $A^n B^p$ ) able to be formed by push-down memory. As such, there may be some LA-specific level of sensitivity exhibited by the products of EM relative to the products of Internal Merge. Since agreement relations are also most commonly present at non-local distances, formed by IM, the language system may have co-opted this domain-general featural comparison operation at the point of IM emergence (see also Murphy 2019).

categorize non-phasal syntactic objects and exclusively serve as labels, and so all non-phase heads and their syntactic objects are labelled by phase heads. As such, just as how a category-less root, R, receives its label externally (i.e. via merging with *v*, *n*, *p* etc.), so too do unlabelled non-phasal XPs receive their label externally through merging with a phase head.

Given that we have defended (iii) ('No copies are visible'), we also believe that future work should seek to explore the (in)compatibility of Copy Invisibility with the very reasonable assumptions about the workspace discussed in Chomsky (2019b) and Chomsky *et al.* (2019). These authors redefine Merge from a process of combining two lexical elements, to a process (termed (capital) MERGE) which rather merges elements *to a workspace* and as such MERGE involves mapping from a workspace, WS, to another, WS'. MERGE operations are "operations on the workspace", not on syntactic objects (Chomsky 2019b: 275). Chomsky (2019b: 280) considers the workspace in (19a) and defines MERGE as the operation in (19b), which replaces two elements in (19a) with the set in (19b):

(19) a.  $\Sigma = (X_1, X_2, \dots, X_n)$   
b.  $\text{MERGE}(\Sigma) = \{\{X_1, X_2\}, \dots, X_n\}$

The derivation is permitted (optionally) to terminate when there is only one syntactic object left in the workspace (but can proceed beyond this), i.e. when a discrete number of items have been merged into various configurations, forming a structured set.<sup>38</sup> We do not see this position as conflicting with ours, but as a reviewer also notes, it is not obvious whether there is any explanatory link between them. We leave this issue open to future research.<sup>39</sup> For instance, Chomsky (2019b) also notes that recursion for language is different to recursion in other cognitive domains, in that it adheres to a computational principle of

<sup>38</sup> Dobashi (2018) applies the notion of workspace to intonational phrasing, suggesting that the domain of a terminated derivation (i.e. a workspace with a single syntactic object) is interpreted as an intonational phrase. See also Komachi *et al.* (2019).

<sup>39</sup> We should stress that a proper formulation of the notion *workspace* has not been forthcoming in the literature, but its explanatory scope may be considerable. For instance, adjuncts could be seen as objects merged to terminated workspaces (i.e. a new object being merged to a single constructed object).

Resource Restriction, limiting the range of objects able to be operated on in a given workspace to maximise efficiency. This may have implications for the design of LA.

In addition, we would like to suggest that, by making copies invisible, this reduces the search burden on LA, extending the original intuition in Shim's (2018) critique of  $\langle F, F \rangle$  labeling. In effect, what we could call a *Reduce Search Space* imperative mirrors Resource Restriction (Chomsky 2019b), applied directly to interface search algorithms. In the same way that syntax reduces the number of elements accessible in the workspace (Chomsky *et al.* 2019), CI (via LA) may implement a search-minimization principle, motivating the elimination of copies from LA. As such, the elimination of copies from LA is not a stipulation, but arises from principles of efficient search, extending core minimalist principles beyond the syntax and into syntax-CI relations.

We have also assumed that LA is sensitive only to the categorial features of SOs. As argued in Shim (2018), this also allows us to avoid the additional layer of complexity involved in creating ordered pair labels (e.g.  $\langle Q, Q \rangle$ ) which require a more complex form of Comparison Search rather than Minimal Search. We have therefore ruled out labels of the type  $\langle F, F \rangle$ , and we can also eliminate the stipulative *strong* vs. *weak* distinction of T (see also Goto 2017 for a critique of this distinction). For us, T can always serve as a label. Our account additionally explains why R is universally weak, and we suspect that our account has no need to posit ad hoc rule ordering (*à la* Chomsky 2015a), although we will leave this issue open for future research.

## References

Abdelhady, Saleem. 2017. Agreement and anti-agreement in Berber: A multiple-feature inheritance account. *International Journal of Humanities and Cultural Studies* 4(2): 1-14.

Abe, Jun. 2016. Dynamic antisymmetry for labeling. *Lingua* 174: 1-15.

Al-Mutairi, Fahad Rashed. 2014. *The minimalist program: The nature and plausibility of Chomsky's bidialectics*. Cambridge: Cambridge University Press.

Bobaljik, Jonathan David. 2008. Where's phi? agreement as a postsyntactic operation. In Daniel Harbour, David Adger, and Susana Béjar (eds.), *Phi theory: Phi-features across*

*modules and interfaces*, 295-328. Oxford: Oxford University Press.

Boeckx, Cedric. 2015. *Elementary syntactic structures: Prospects of a feature-free syntax*. Cambridge: Cambridge University Press.

Borer, Hagit. 2014. Derived nominals and the domain of content. *Lingua* 141: 71-96.

Bošković, Željko. 2011. On unvalued uninterpretable features. In Suzi Lima, Kevin Mullin, and Brian Smith (eds.), *Proceedings of NELS* 39. GLSA. 109-120.

Bošković, Željko. 2016. On the timing of labeling: deducing comp-trace effects, the subject condition, the adjunct condition, and tucking in from labeling. *The Linguistic Review* 33: 17-66.

Carstens, Vicki. 2016. Delayed valuation: a reanalysis of goal features, "upward" complementizer agreement, and the mechanics of case. *Syntax* 19: 1-42.

Carstens, Vicki, Norbert Hornstein, and Daniel Seely. 2016. Head-head relations in problems of projection. *The Linguistic Review* 33: 67-86.

Cecchetto, Carlo and Caterina Donati. 2015. *(Re)labeling*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1956. *Logical structure of linguistic theory*. Ms. Plenum, 1975; Chicago, 1975.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, Juan Uriagereka, and Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* 89-155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland and Hans-Martin Gärtner (eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*, 1-30. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero, and María Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud* 133-166. Cambridge, MA: MIT Press.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130: 33-49.

Chomsky, Noam. 2015a. Preface to the 20th anniversary edition. *The minimalist program*, vii-xiv. Cambridge, MA: MIT Press.

Chomsky, Noam. 2015b. Problems of projection: extensions. In Elisa Di Domenico, Cornelia Hamann, and Simona Matteini (eds.), *Structures, strategies and beyond: Studies in honour of Adriana Belletti*, 1-16. Amsterdam: John Benjamins.

Chomsky, Noam. 2019a. Puzzles about phases. In Ludovico Franco and Giulia Bellucci (eds.), *Linguistic variation: Structure and interpretation - a festschrift in honour of M. Rita Manzini*, 163-168. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2019b. Some puzzling foundational issues: The Reading program. *Catalan*

*Journal of Linguistics Special Issue* 263–285.

Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics Special Issue* 226–261.

Citko, Barbara, Allison Germain, and Jacek Witkoś. 2018. If you cannot agree, move on! On labels and non-nominative subjects. *Glossa: A Journal of General Linguistics* 3(1), 28.

Collins, Collins and Erich Groat. 2018. Distinguishing copies and repetitions. Ms. ling.auf.net/lingbuzz/003809.

Collins, John. 2020. Conjoining meanings without losing our heads. *Mind & Language* 35: 224–236.

Dobashi, Yoshihito. 2018. Termination of derivation and intonational phrasing: A preliminary study. Hisao Tokizaki (ed.), *Phonological externalization*, 9–23. Sapporo University.

Embick, David. 2012. Roots and features (an acategorial postscript). *Theoretical Linguistics* 38(1–2): 73–89.

Emonds, Joseph. 1976. *A transformational approach to English syntax: Root, structure-preserving and local transformations*. New York: Academic Press.

Epstein, Samuel. 1992. Derivational constraints on Ā-chain formation. *Linguistic Inquiry* 23: 235–259.

Epstein, Samuel, Hisatsugu Kitahara, and Daniel Seely. 2014. Labeling by minimal search: Implications for successive-cyclic A-movement and the conception of the postulate 'phase'. *Linguistic Inquiry* 45(3): 463–481.

Fox, Daniel. 2000. *Economy and semantic interpretation*. Cambridge, MA: MIT Press.

Fox, Daniel. 2002. Antecedent contained deletion and the copy theory of movement. *Linguistic Inquiry* 33: 63–96.

Fukui, Naoki. 2017. *Merge in the mind-brain: Essays on theoretical linguistics and the neuroscience of language*. London: Routledge.

Gallego, Ángel J. 2018a. Freezing effects in a free-merge system. In Jutta Hartmann, Marion Jäger, Andreas Kehl, Andreas Konietzko, and Susanne Winkler. (eds.), *Freezing: Theoretical approaches and empirical domains*, 66–104. Berlin: De Gruyter Mouton.

Gallego, Ángel J. 2018b. Projection without agreement. *The Linguistic Review* 35(4): 601–623.

Gelderen, Elly. 2018a. Cyclical change and problems of projection. Ms. ling.auf.net/lingbuzz/003815.

Gelderen, Elly. 2018b. Problems of projection: the role of language change in labeling paradoxes. *Studia Linguistica* 72: 113–127.

González Escribano, José Luis. 2005. Semantocentric minimalist grammar. *Atlantis* 27(2): 57–74.

Goto, Nobu. 2017. Eliminating strong/weak parameter on T. *Proceedings of GLOW in Asia XI, Volume 2. MIT Working Papers in Linguistics* 85: 57–71.

Grohmann, Kleanthes K. 2003. *Prolific domains: On the anti-locality of movement dependencies*.

Amsterdam: John Benjamins.

Hauser, Marc, Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298(5598): 1569-1579.

Hicks, Glyn. 2009. *The derivation of anaphoric relations*. Amsterdam: John Benjamins.

Hiraiwa, Ken. 2016. Labeling roots and pronouns. Paper presented at North East Linguistic Society (NELS) 47, University of Massachusetts, Amherst, 14 October.

Holmberg, Anders and Thorbjörg Hróarsdóttir. 2003. Agreement and movement in Icelandic raising constructions. *Lingua* 113: 997-1019.

Hornstein, Norbert. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.

Hosono, Mayumi. 2016. Exceptional movement from/into the criterial position. *Working Papers in Scandinavian Syntax* 97: 23-39.

Hosono, Mayumi. 2018. Constraints on movement. *Working Papers in Scandinavian Syntax* 100: 41-63.

Kitahara, Hisatsugu. 2018. On Chomsky's READING lecture. Paper presented at Keio University.

Komachi, Masayuki, Hisatsugu Kitahara, Asako Uchibori, and Kensuke Takita. 2019. Generative procedure revisited. *Reports of the Keio Institute of Cultural and Linguistic Studies* 50, 269-283.

Larson, Bradley. 2015. Minimal search as a restriction on merge. *Lingua* 156: 57-69.

Marantz, Alec. 1997. No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2): 201-205.

McCloskey, James. 1991. There, it and agreement. *Linguistic Inquiry* 22: 563-567.

Miyagawa, Shigeru, Danfeng Wu, and Masatoshi Koizumi. 2019. Inducing and blocking labeling. *Glossa: A Journal of General Linguistics* 4(1): 141. 1-26.

Mizuguchi, Manabu. 2017. Labelability and interpretability. *Studies in Generative Grammar* 27(2): 327-365.

Mizuguchi, Manabu. 2019a. A-movement in phases and its (counter-)cyclicity. Ms. Department of Social Psychology. Toyo University.

Mizuguchi, Manabu. 2019b. Ambiguous labeling and full interpretation. *Studia Linguistica: A Journal of General Linguistics* 73: 563-603.

Mizuguchi, Manabu. 2019c. Optional raising in ECM and labeling of XP-YP. *Generative Grammar Research* 29(2): 373-411.

Müller, Gereon and Wolfgang Sternefeld. 1996. A'-chain formation and economy of derivation. *Linguistic Inquiry* 27: 480-511.

Murphy, Elliot. 2015a. Labels, cognomes and cyclic computation: an ethological perspective. *Frontiers in Psychology* 6: 715.

Murphy, Elliot. 2015b. Reference, phases and individuation: topics at the label-

ing-interpretive interface. *Opticon* 18(5): 1-13.

Murphy, Elliot. 2019. No country for Oldowan men: Emerging factors in language evolution. *Frontiers in Psychology* 10: 1448.

Narita, Hiroki. 2014. *Endocentric structuring of projection-free syntax*. Amsterdam: John Benjamins.

Narita, Hiroki and Naoki Fukui. 2020. *Symmetry-driven Syntax: An inquiry into endocentricity and symmetry in human language*. London: Routledge.

Narita, Hiroki., Hironobu Kasai, Takaomi Kato, and Mihoko Zushi. 2017. 0-search and 0-Merge. In Naoki Fukui (ed.), *Merge in the mind-brain*. New York: Routledge. 127-154.

Ouali, Hamid. 2008. On C-to-T  $\varphi$ -feature transfer: The nature of agreement and anti-agreement in Berber. In Roberta D'Alessandro, Susann Fischer, and Gunnar Hrafn Hrafnbjargarson (eds.), *Agreement restrictions*. 159-180. Berlin: Walter de Gruyter.

Pietroski, Paul M. 2018. *Conjoining meanings: Semantics without truth values*. Oxford: Oxford University Press.

Preminger, Omer. 2014. *Agreement and its failures*. Cambridge, MA: MIT Press.

Reuland, Eric. 2011. *Anaphora and language design*. Cambridge, MA: MIT Press.

Richards, Marc. 2019. Problems of 'problems of projection': breaking a conceptual tie. *Catalan Journal of Linguistics Special Issue* 139-152.

Richards, Norvin. 2016. *Contiguity theory*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2006. On the form of chains: criterial positions and ECP effects. In Lisa Lai-Shen Cheng and Norbert Cover (eds.), *Wh-movement: Moving on*. 97-133. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2015. Cartography, criteria, and labeling. In Ur Shlonsky (ed.), *Beyond functional sequence: The cartography of syntactic structures*, Volume 10, 314-338. Oxford: Oxford University Press.

Rizzi, Luigi. 2016. Labeling, maximality and the head-phrase distinction. *The Linguistic Review* 33(1): 103-127.

Saito, Mamoru. 2016. (A) Case for labeling: labeling in languages without phi-feature agreement. *The Linguistic Review* 33: 129-175.

Saito, Mamoru. 2018. Kase as a weak head. Ms. Nanzan University.

Shim, Jae-Young. 2014. Phase-head initiated structure building: Its implications for feature-inheritance, transfer and internal Merge. *Studies in Generative Grammar* 24(2): 455-474.

Shim, Jae-Young. 2018.  $\langle\varphi, \varphi\rangle$ -less labeling. *Language Research* 54(1): 23-39.

Stowell, Tim. 1981. Origins of phrase structure. PhD dissertation. MIT.

Takita, Kensuke, Nobu Goto, and Yoshiyuki Shibata. 2016. Labeling through Spell-Out. *The Linguistic Review* 33(1): 177-198.

**Elliot Murphy**

Researcher

Vivian L. Smith Department of Neurosurgery,  
McGovern Medical School,  
University of Texas Health Science Center,  
6431 Fannin St, Houston, TX 77030, USA  
E-mail: elliot.murphy@uth.tmc.edu

**Jae-Young Shim**

Lecturer

Department of English Language and Literature  
Gyeongsang National University  
501, Jinjudaero, Jinju,  
Gyeongsangnamdo 52828, South Korea  
E-mail: biolinguistics@gmail.com

Received: 2020. 04. 06.

Revised: 2020. 05. 26.

Accepted: 2020. 06. 19.