# Copy Invisibility, (Non-)Categorial Labeling and Feature Embedding

**Elliot Murphy[1,2] & Jae-Young Shim[3]**

1. Department of Linguistics, University College London
2. Department of Psychology, University of Westminster
3. Department of English Language & Literature, Pukyong National University

**Abstract:** 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 reinterpreation of LA theory such that LA constructs only categorial labels, barring the construction of <Q,Q> and <φ,φ> configurations. We then propose an interface condition, *Equal Embedding* (*EE*), under which agreeing features must be equally as embedded in order for interpretion to be licensed. More broadly, *EE* appears to fall out of minimal search requirements, and is hence a good candidate for a third factor effect. Lastly, we discuss how this LA architecture can be situated within the syntactic workspace (WS). In particular, we propose an *Economy of Labeling* theorem, which states that if two WSs have distinct numbers of copies the WS with fewer copies is labeled.

**Keywords**: Copy invisibility, Criterial position, Labeling, Minimal search, Equal Embedding, Economy of Labeling, MERGE

## 1. Labeling Theory

The Labeling Algorithm (LA) in the tradition of Chomsky (2008, 2013, 2015b, 2019) 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 (2013) as the Single Label Condition on Interpretation:

(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 1995, Chomsky *et al*. 2018), with (1) claiming that every constructed set of independent syntactic objects must also have a specific label. LA is standardly assumed to search not just for categorial features of heads, but also for any agreeing features shared by two equally embedded heads in order to establish the label of an object.

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.

This notion is closely related to the "strong hypothesis" entertained in Chomsky *et al*. (2018: fn. 17) that "operations never extend [the workspace]".[1] 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-semantic and CI requirements.

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.[2] Typically, functional features such as tense,

[1] 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.

[2] We think it is reasonable to derive certain cross-linguistic differences from feature inheritance, i.e. in language X features α + β are inherited by T from C, but in language Y only feature α 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 [uφ] on phase heads. Hence, we will assume that feature inheritance purely involves non-phase heads taking

φ-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**-R(=V) relation.[3] Below is a standard labeling process (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(=bought), NP(=a toy)}

c. {$_{\beta}$ NP, {$_{\alpha}$ R, <NP>}}

d. {*v**, {$_{\beta}$ NP, {$_{\alpha}$ R, <NP>}}}

e. {*v**, {$_{\beta}$ NP, {$_{\alpha}$ R, <NP>}}} (α = R, β = <φ,φ>)

[φ] [φ]

In Chomsky's (2013, 2015b) 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. Given this assumption, the first relevant derivational stage in terms of labeling is (3d), where the label of the two SOs, α and β, 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 φ-features on *v** are inherited by R. The two SOs, α and β, can now be labeled since R has turned into a labelable head thanks to its newly acquired φ-features that agree with NP in its Spec (i.e. satisfying (4b)). Consequently, α is

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.

[3] 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 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".

labeled by the legitimate unique head, i.e. R, and β is labeled as a pair of the agreeing features between R and the head of NP, i.e. <φ,φ>.[4]

We must assume, then, that CI only reads off labels for interpretation (contra Takita 2018). As such, LA appears not simply to operate via Minimal Search[5], but also appears sensitive to syntactically bare elements.[6]

## 2. Copy Invisibility and the Weak vs. Strong Distinction

In Shim (2018), it is proposed that the labeling operation of searching for agreeing features, as in <φ,φ> or <Q,Q> 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).[7] 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 <Q,Q> or <φ,φ> (which respectively arise from what has traditionally been termed Ā-

[4] 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.

[5] 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:

(1) **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 Eptein et al. (2014).

[6] Presumably, this is simply a design feature of LA, such that it may not in fact be quite accurate to describe LA as being sensitive to bare elements, but rather that it is *only sensitive to syntactic features* (in particular, this includes categorial features like N). That is, it is not specifically the conceptual nature of R that prevents labeling, but rather its lack of syntactic features.

[7] By invoking these non-categorial <φ,φ>-type structures, Chomsky (2013) in effect brought back the 'feature intersection' option he discarded in Chomsky (1995).

movement and A-movement) become superfluous.[8] For instance, although the <φ,φ> label in (3e), repeated here as (5b), is a possiblity, it is also possible that the derivation from (3b), repeated heres as (5a), could generate (5c) with R raising to *v** (via 'Head Movement'), yielding the copy <R> 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.[9]

(5) a. {$_{\alpha}$ R(=bought), NP(=a toy)}

b. {*v**, {$_{\beta}$ NP, {$_{\alpha}$ R, <NP>}}} ($\alpha$ = R, $\beta$ = <φ,φ>)

[φ] [φ]

c. {[R-v*], {$_{\alpha}$ <R>, NP}} ($\alpha$ = N)

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

(6) {[R, *v**], {$_{\beta}$ NP, {$_{\alpha}$ <R>, <NP>}}}

Note that Chomsky's copy invisibility requires labeling of α and β in (6) to precede raising of R to *v**; otherwise, labeling of α and β 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 unecessary order between IM (of R-to-*v**) and labeling (of α and β).

Another problem with Chomsky's system is that unlike categorial labels like N and V, non-categorial labels such as <φ,φ> can cause a problem with regard to CI interpretation provided that these labels host both an interpretable feature and an uninterpretable feature, e.g. the interpretable φ-features from N and the uninterpretable φ-features from R (via feature inheritance from *v**).

As with the common deployment of <φ,φ> structures in contemporary LA theory to avoid bare Rs, it is also commonly 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.

[8] See Mizuguchi (2019a) for a clear definition of A/Ā-positions: "The *n*P is in an A-position if it is merged with an SO headed by a head bearing φ-features; otherwise, it is in an Ā-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.

(7) a. John seems to like Mary.

b. {C {$_{\delta}$ John {$_{\gamma}$ $T_2$-seems {$_{\beta}$ <John> {$_{\alpha}$ $T_1$-to {*v**P <John> like Mary}}}}}}

Labeling of δ and γ proceeds via, respectively, the agreeing φ-features between $T_2$ and *John*, and strengthened $T_2$, yet the labeling of α and β 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 α or β.

The core of all the problems discussed above seems to us to lie in inconsistent treatment of the notion of *copy* in the strandard LA theory. Chomsky (2007 *et seq*.) 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 structually highest copy is visible to LA, while all other copies are not. Resolving these inconsistencies, Shim (2018) proposes the following principle.

(8) **Copy Invisiblity** (to LA)

*All* copies are invisible to LA.

Copy Invisiblity 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.[10] This produces a subtle but clear revision to the syntax-semantic 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.[11] In conjunction, Shim proposes that LA seeks only categorial features (presumably part of Full Interpretation), attemping to eliminate non-categorial labels from CI and certainly minimise their role in the

[10] A more basic defence of standard copy invisibility can be found in intervention effects.

(1) 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 φ-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. This intervention issue also does not arise in Chomsky's (2015b) framework, since the object ('which car') does not move to the outer Spec-*v**P but to Spec-RP, lower than the subject.

[11] Notice that Copy Invisiblity also forces a reinterpretation of the function of standard transformations/movement operations and centring on their function in avoiding labeling failures. Put differently, the function of IM is to render syntactic objects invisible to LA.

grammar. In effect, Copy Invisibility leads the way to categorial-sensitive labeling – in direct contrast to standard LA theory assumptions about copy visiblity leading to weak Rs and Ts (e.g. Chomsky 2015b).

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 <φ,φ> 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).[12] 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 φ, and traditional categorial features in combination with Copy Invisiblity suffice to generate LA.

Reevaluating the landscape, consider again (3b), which as mentioned can only yield the derivation in (3c) in the standard LA theory. Given what we have outlined, (3b) can now generate either (9c) or (9d):

(9) a. The boy bought a toy.
b. $\{_{\alpha}$ R(=bought), a toy$\}$
c. $\{v^*, \{_{\alpha}$ R, a toy$\}\}$ ($\alpha$ = R)
d. $\{v^*, \{_{\beta}$ ~~a toy~~,$\{_{\alpha}$ R, ~~a toy~~$\}\}\}$ ($\alpha$, $\beta$ = R)

Under this analysis, it only takes the category-defining *v** to merge in order for R to serve as the label.[13] Consequently, α can be labeled R irrepective 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.

Reconsidering (7b), repeated here as (10b), the present analysis can also provide a label for α and β, with $T_1$ providing a clear categorial label for both in the absence of *John* in its Spec (considered invisible to LA).

(10) a. John seems to like Mary.

[12] 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. φ) implementing feature valuation. The result is a <φ,φ> label, under the assumption (that we also maintain here) that labeling follows valuation.

[13] We assume that some form of feature inheritance (from *v** to R) provides R with a category feature.

b. {C {$_{\delta}$ John {$_{\gamma}$ $T_2$-seems {$_{\beta}$ <John> {$_{\alpha}$ $T_1$-to {*v**P <John> like Mary}}}}}}

c. { C {$_{\delta}$ John {$_{\gamma}$ $T_2$-seems {$_{\alpha}$ $T_1$-to {*v**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 (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 <φ,φ> and <Q,Q> (which we will argue should be dispensed with).[14] For example, <φ,φ> 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 <φ,φ>. Reformulating these ideas within LA theory, Chomsky (2019) 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.

## 3. 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 completed the valuation of unvalued features (e.g. [uCase], [uφ]).[15] This *Halting Problem* can be demonstrated below, where the *wh*-object moves

[14] 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".

[15] 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).

from its originally merged position to Spec-(embedded)CP in order for feature valuation to take place between [uQ] of *which dog* and [Q] of $C_Q$. When it tries to move out of this criterial position (CriP), the sentence is ungrammatical.

(11) a. You wonder [$_{CP}$ [$_Q$ which dog] $C_Q$ John likes ~~[$_Q$ which dog]~~].
b. *[$_\alpha$ [$_Q$ which dog] do [$_{\beta<\varphi,\varphi>}$ you wonder [$_{\gamma<Q,Q>}$ ~~[$_Q$ which dog]~~ $C_Q$ [John likes ~~[$_Q$ 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 Acc, but since *which dog* retains its [uQ], it moves to Spec-γ.[16] *Wonder* subcategorizes as *wh*-clause and so the embedded C hosts [Q], and *which dog* in Spec-γ and $C_Q$ proceed to feature valuation, labeling γ <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 $C_Q$ provides a clear label, with Copy Invisibility leading both occurrences of *which dog* to be invisible to LA.

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.[17] This process then bars further movement in conjunction with Copy Invisiblity 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.[18]

We would here like to propose a (CI-)interface condition, *Equal Embedding* (*EE*), which induces what we call a 'Magnet Effect' in the narrow syntax.

[16] 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.

[17] 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.

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

(12) **Equal Embedding**

Agreeing features must be equally embedded.

Syntactically speaking, what (12) requires is that once a feature is next to another feature of the same type (e.g. [ … [*u*φ-*i*φ] …]), 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 [*u*Q-*i*Q] configuration. As such, under our account halting effects arise due to the CI-localised Magnet Effect (analogous to the case of magnetic poles). Further, structures standardly analysed as <φ,φ> configurations are, under our system, labeled as whatever categorial feature merges with it, with *EE* requiring that these φ-bearing objects be syntactically embedded at the same hierarchical level. More broadly, *EE* appears to fall out of minimal search requirements, and is hence a good candidate for a third factor requirement at CI.

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 to serve as labels for interpretation.[19] CriP is 'marked' by feature valuation by default and <Q,Q> structures do not in fact 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.[20]

[19] 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*. (2018) that those objects constructed in the syntax which are not appropriately interpreted at CI will be filtered out.

[20] For Gallego (2018a), it is suggested the freezing takes place not just because of feature valuation, but because of CI demands: Movement of a *d*XP (=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.

## 4. Ambiguous Labeling

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.

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, <φ,φ> 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-*n*P subjects.[21] Subjects can sometimes be non-*n*Ps such as *that*-clause subjects or prepositional subjects, which do not agree in φ-features with T (Emonds 1976, Stowell 1981).[22]

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

Subjects in (13) are unable to agree with the matrix T. The embedded CP in (13a), for example, does bear φ-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 φ-features would have been transferred (hence inaccessible to the matrix T) by the time the matrix T is introduced.[23]

---

[21] For example, Mizuguchi (2019b) reviews how in-situ subjects can be created without causing infelicitous structures at the interfaces:

(1) $[_{\gamma}\ [_{n\mathrm{P}}\ n,\ \mathrm{NP}],\ [_{v^*\mathrm{P}}\ v^*,\ \mathrm{VP}]]$

Mizuguchi also explores how labeling ambiguities can derive properties of free relatives exhibiting dual interpretations. This general framework provides a cognitively flexible picture of the core language system and is also in line with the label-driven semantic interpretation proposed in Murphy (2015b).

[22] They receive unmarked third person, singular interpretation at CI and do not host φ (see Preminger 2014).

[23] As far as we can understand it, Mizuguchi (2019b) attributes unavailability of φ-features for the matrix T in (13a) to unavailability of the φ-features on the embedded C. That is, Mizuguchi argues that no φ-features for the

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 *n*P subjects. Thus, they form [XP,YP] configurations, as below, but the subjects do not agree with T.

(14) [$_{\gamma}$ [$_{\alpha}$ X [YP]], [$_{\lambda}$ T [*v*P]]]

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 (14) 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 γ in (14) is merged with C, becoming its complement, CI will dictate that γ be labeled T, since C selects T but not any of the X objects. If γ is labeled X, then the C-T-*v* structure will be inferred and clausal interpretation will not be established.[24] As such, LA remains a minimal search procedure but CI conditions impose their own demands for which ambiguous label is selected; a coherent framing given that labels are ultimately established only for reading off at CI.

Returning to our interface condition, *EE* in (12), the insight we would like to present here is that this CI condition can also explain CriP phenomena: *u*Q must be next to or equally embedded as *i*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 'magnetised' 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 *n*P. This keeps to our proposals that categorial labels alone are sufficient for

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 φ-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.

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

interpretation and that Copy Invisibility would predict that *v** would be selected as the label given that it does not undergo movement.

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

(15) a. *Was* meinst du [$_{\gamma}$ **wen**$_{i}$ C [Peter Hans t$_{i}$ vorgestellt hat]]?
'Who do you think Peter has introduced to Hans?' (modified from Mizuguchi 2019b, (40a))

b. [$_{\gamma}$ [*Wh*P], [CP]]

As illustrated in (15b), LA will detect *n* (in *Wh*P) and C, and hence the configuration can be labeled either *n* or C. However, CI will dictate that it must 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 *n*P. 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 λ in (14) without requiring any movement to 'strengthen' it (contra Chomsky 2015b), and X will immediately label α, leaving γ 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 <φ,φ> 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.

Where we differ from Mizuguchi (2019b) is in relation to Copy Invisibility. Mizuguchi explains the ill-formedness of (16a) in terms of the standard weak T analysis, noting that only in the event of a merged *n*P at γ will T be strengthened in English to license a well-formed interpretation. However, for us, no weak T stipulations are needed, and the invisiblity of both the Spec-α copy of *the student* and the initial copy ensure that α cannot be assigned an appropriate label to then merge with γ. It is only when γ is merged and *the student* undergoes successive-cyclic movement to generate (16b) that feature valuation can proceed.

(16) 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.[25]

An alternative account for (16) is to assume that copies are invisible to LA, but not to Agree. 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. Perhaps this is because Agree is solely an operation geared towards 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 is less concerned with 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.

Lastly, *EE* also requires that these agreeing features be equally embedded; that is, the matrix T in (16a) will agree with 'the student' but the two agreeing φ-features are not equally embedded, in violation of *EE*.

## 5. LA Detection of $X^{min}$

[25] 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.

Just as we have argued, Gallego (2018b) proposes that all LA detects is lexical items (traditionally, $X^{min}$) and not their features. We take this as leading to the same predictions as our model. Drawing on similar concerns to the ones we have already raised, Gallego (2018b) brings into question the assumption in standard LA theory that non-categorial labels must be determined via the most 'prominent' features shared by X and Y in [XP,YP]. As well as demanding a form of *comparison search* (Shim 2018), Gallego notes that this assumption also demands that LA can determine the internal structure of a lexical item, a notion which seems at odds with the standard formulation of lexical items as atoms of computation (Chomsky 2013: 41). As such, Gallego questions the assumption that LA detects not just what have traditionally been termed formal features, but also semantic and phonological features. That is, he questions the idea that LA can see [+V] but also [+Abstract]-Subject or [+Animate]-Object (Chomsky 1965), which are difficult to disassociate, and indeed doing so via LA would no longer render them atoms. Gallego (2018b: 609) adds that non-categorial labels of the <φ,φ> type will require a feature-sharing process that is not compatible with Simplest Merge. As such, while non-categorial labels may not in fact exist, the process of non-categorial feature valuation does at least seem to halt movement even if this process does not produce an independent label used for interpretation.

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

(17) {{{**x**, {…}}, ZP}, {{**y**, {…}}, 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.

Theoretically, our claim that LA only constructs categorial labels and not non-categorial labels is analogous to Gallego's claim that LA is only sensitive to lexical items ($X^{min}$) and not lexical features.[26]

[26] 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,

## 6. Copies and Workspaces

In this section, we will discuss some broader implications for the LA model we have proposed here.

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 standard distinctions between External Merge and Internal Merge, 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.[27] We would like to suggest that there is also an additional layer of symmetry 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 a symmetry between all copies (invisible) on the one hand, and non-copies and repetitions (visible) on the other.[28] Following assumptions in Fukui (2017) and Narita and Fukui (2020), we can therefore assume that a derivation, D, can be represented as a sequence

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.

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

[28] 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.

$<S_0, S_1, ..., Sf>$ ($f > 0$), where each $S_i$ is a given stage of the workspace and is mapped to $S_i$+1 by Merge.

In addition, as explored in Murphy (2015a), crossing dependencies of the kind generated by IM ($A^n\ C^m\ B^n\ D^m$) require a mildly context-sensitive grammar and a linear bounded automaton, while EM produces the less memory-intensive nested dependencies ($A^n\ B^n$) 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. If LA emerged at the point (of language evolution) of the initial mutations (or series of macro-mutations) responsible for EM, then this would explain why it may not be sensitive to the outputs of IM, which emerged at a later stage (following the detailed timeline presented in Murphy 2018; see also references therein). Thus, we are able to present a syntactically, cognitively and evolutionarily well-motivated account of the format of LA. 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 introduction. As such, Agree would be sensitive to all objects manipulated by the syntax.

As far as we can tell, there are at least three logically possible models of copy visibility.[29] We have here defended (iii):

(i) All copies are visible (e.g. Mizuguchi 2019b).

(ii) The highest copy is visible (e.g. Chomsky 2019).

(iii) No copies are visible.

Alongside the above motivations for (iii), we also believe that Copy Invisibility is compatible with the very reasonable assumptions about the workspace discussed in Chomsky *et al*. (2018). 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′. 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.[30] Chomsky *et al*. (2018) discuss a fundamental

[29] 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).

[30] 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.

question: When MERGE takes X and Y from the workspace (WS) [X,Y] and forms the set {X,Y}, does it add this new object to WS, yielding WS′ = [X,Y,{X,Y}], or does it instead replace X and Y with the newly formed object, yielding WS′ = [{X,Y}]? Considerations of computational efficiency led Chomsky et al. to reasonably opt for the latter option (returning to proposals in Chomsky 1995). There are also more general reasons from the cognitive neuroscience of working memory supporting this 'cut (not copy) + paste' type of workspace (see Sauseng *et al*. 2018), but for our purposes it seems that this model is also compatible with Copy Invisibility, such that SOs not manipulated by IM will be reinforced in working memory (they will be 'doubly' present in the same format between the initial workspace and then again in the newly-formed one), while moved elements will be rehearsed in derivational memory as taking multiple set-theoretic positions.[31]

We believe this proposal is in line with moves already made in the literature to make the derivation more cognitively-oriented, taking into consideration memory and search processes. Complementing Adger's (2017) dual workspace model of syntax, Ke (2017) proposes the following:

(18) **Two-Phase Workspace Hypothesis** (Ke 2017: 11)
Narrow Syntax is able to keep two active phases, and no more, in the workspace at a time. A phase is active in the workspace if it is not Transferred.

There are a number of instances where two separate workspaces are needed[32], as when the lower phase contains unvalues features, such as in *v**P-internal subjects (where the external argument contains *u*F[-Case]) or long-distance *wh*-movement (the *wh*-phrase hosts *u*F[-Q]). If LA is only applicable at the workspace 'closest' to the interfaces (for Adger (2017), his Operating Space, as opposed to his Resource Space, would be the one responsible for interface transfer), then this would ensure that elements which have undergone IM will – as mentioned above – have a different status in the workspace, since they will have exploited a linear bounded automaton.

LA's minimal search will also be forced to track the derivational history of items which have undergone IM, which is a more complex search process than detecting items which have

[31] We admit that this approach risks violating the Inclusivity Condition, yet it also seems that the very notion of a workspace opens up a range of Inclusivity-violating implications which might need to be accommodated.

[32] See Komachi *et al*. (2019) where it is claimed that 'there is one and only one workspace at any stage of the derivation.'

undergone EM. While EM applies to a WS = [α, β, γ] and yields a newly updated WS′ = [{α, β}, γ], IM proceeds in a very different way, which we argue forces LA to make a clear distinction between it and EM. Consider WS = [{α, {β, γ}}], where β and γ have undergone EM and then α has externally merged with {β, γ}. MERGE takes γ and applies it to WS, creating WS′ = [{γ, {α, {β, γ}}}]. For EM, the three syntactic objects in WS = [α, β, γ] are reduced to two in WS′ = [{α, β}, γ]. But notice that for IM both WS and WS′ contain a single syntactic object. We argue that for LA's phase-based memory the only differentiating factor between WS and WS′ in the IM case are the copies of γ. LA detects that it is labeling a single transferred object, yet for minimal search α and β are more easily detectable due to their prominence in derivational memory as a result of a more stable set-theoretic position throughout rehearsal. This also allows us to adhere to one of Chomsky's (2017) *seven desiderata* to which MERGE must conform, *Determinacy*, which states that accessible terms can appear in a WS only once – if copies are invisible to LA, the status of γ is irrelevant after IM and minimal search can select either α and β as the functioning label for interpretation.[33]

What else can be said of *Determinacy*? We believe that removing LA's sensitivity to copies can help reduce ambiguity of rule application, in the following way: Consider three workspaces, where γ has been targeted by MERGE (IM) and then again to form $WS_3$.

(19) a. $WS_1$ = [{α, {β, γ}}, δ]
b. $WS_2$ = [{γ, {α, {β, γ}}, δ]
c. $WS_3$ = [{γ, {γ, {α, {β, γ}}, δ]

Putting aside issues of successive-cyclicity, consider only that $WS_3$ contains three copies of γ. While movement of this kind is indeed permitted (via assuming the Phase Impenetrability Condition), labeling any workspace output γP is not, and instead some additional element will serve as the head (say, δ, when it is introduced). That is to say, once we assume that the lower copy is invisible to LA, we gain no particular advantage in assuming that moved copies can be labeled. As a direct consequence of this, we propose the principle of *Economy of Labeling*, which we take to be a component of LA's minimal search procedure (following in the spirit of

[33] We should stress that a proper formulation of the notion *workspace* has, surprisingly, not been forthcoming in the literature, and syntacticians have only recently begun to explore the issue (but see Adger 2017 for excellent discussion). It appears, though, that its explanatory scope is 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).

the *Economy of Derivation*, in which derivations with fewer operations are preferred; Müller & Sternefeld 1996: 480-481) and hence assumes the status of a syntactic theorem:

(20) **Economy of Labeling**

If two WSs have distinct numbers of copies the WS with fewer copies is labeled.

This does not require that IM cannot occur; rather, it requires that LA (*and LA alone*) operates independently of it.[34] While *EE* is a clear CI condition, *Economy of Labeling* is aimed solely at LA. An immediate consequence of this theorem is that a WS with no objects targeted by IM will automatically be labeled over an immediate alternative WS with at least one instance of IM. [35] In conjunction with Ke (2017) and Adger's (2017) two-phase/dual workspace hypotheses, we can move beyond standard generative assumptions of single phase-based derivational memory and motivate a more pluratistic, cognitively motivated account of labeling whilst remaining grounded in traditional assumptions emanating from the memory architecture of formal language hierarchies. Dual workspaces will permit these competing workspaces to be constructed, and for *Economy of Labeling* to adjudicate between them. *Economy of Labeling* is, in effect, a workspace-centric explanation for Copy Invisibility; i.e. the reason copies are invisible is because of LA's economy principle of efficient search. Notice that we are not claiming that copies are unimportant for interpretation at CI, which for reasons of, for instance, θ-interpretation, they are. Rather, our theory concerns the architecture of LA, its scope and application.

Lastly, 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. <Q,Q>) which require a more complex form of Comparison Search rather than Minimal Search. We have therefore ruled out labels of the type <F,F>, and we can also eliminate the stipulative *strong* vs. *weak* distinction of T. For us, T can always serve as a label. Our account additionally explains why R is universally weak, and we

[34] Indeed, the very existence of structures formed only through EM at once suggests the potential independence LA has from IM.

[35] One might object that a consequence of *Economy of Labeling* is that only WSs with no IM will be labeled (e.g. $WS_1$ in (19)), yet these will often be ruled out on independent grounds (feature-valuation requirements, etc.) and as such LA will operate only on structures satisfying these additional, core CI requirements, which may require IM; after IM has applied, the economical procedure would be to label the WS which has undergone IM but to exclude the IM-manipulated object from search.

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

Adger, D. 2017. A memory architecture for merge. Ms. Queen Mary University of London.

Boeckx, C. 2015. *Elementary Syntactic Structures: Prospects of a Feature-Free Syntax*. Cambridge: Cambridge University Press.

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

Bošković, Ž. 2011. On unvalued uninterpretable features. Lima, S., Mullin, K., & Smith, B. (Eds). *Proceedings of NELS 39*. GLSA. 109–120.

Carstens, V., Hornstein, N., & Seely, T.D. 2016. Head-head relations in *Problems of projection*. *The Linguistic Review* 33: 67–86.

Cecchetto, C., & Donati, C. 2015. *(Re)labeling*. Cambridge, MA: MIT Press.

Chomsky, N. 1956. *Logical Structure of Linguistic Theory*. Ms. Plenum, 1975; Chicago, 1975.

——. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

——. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

——. 2001. Derivation by phase. In Kenstowicz, M. (ed.). *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press. 1–52.

——. 2007. Approaching UG from below. In Sauerland, M., & Gärtner, H-M. (Eds). *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Berlin: Mouton de Gruyter. 1-30.

——. 2008. On phases. In Freidin, R., Otero, C. P., & Zubizarreta, M. L. (eds.). *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press. 133–66.

——. 2013. Problems of projection. *Lingua* 130: 33–49.

——. 2015a. Preface to the 20th anniversary edition. *The Minimalist Program*. Cambridge, MA: MIT Press. vii–xiv.

——. 2015b. Problems of projection: extensions. Di Domenico, E., Hamann, C., & Matteini, S. (eds.). *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*. Amsterdam: John Benjamins. 1–16.

——. 2017. Generative linguistics in the 21st century: the evidence and the rhetoric. Talk presented at The University of Reading, May 11, 2017.

——. 2019. Puzzles about phases. Franco, L., & Belliuci, G. (Eds). *Linguistics Variation: Structure and Interpretation – A Feschrift in Honour of M. Rita Manzini*. Berlin: Mouton de Gruyter.

Chomsky, N., Gallego, Á.J., & Ott, D. 2018. Generative grammar and the faculty of language: insights, questions, and challenges. Gallego, Á.J., & Ott, D. (Eds). *Generative Syntax: Questions, Crossroads, and Challenges*. Special issue of *Catalan Journal of Linguistics*.

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

Dobashi, Y. 2018. Termination of derivation and intonational phrasing: a preliminary study. Tokizaki, H. (Ed). *Phonological Externalization*. Sapporo University. 9–23.

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

Emonds, J. 1976. *A Transformational Approach to English Syntax: Root, Structure-Preserving, and Local Transformations.* New York: Academic Press.

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

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

Fukui, N. 2017. *Merge in the Mind-Brain: Essays on Theoretical Linguistics and the Neuroscience of Language*. London: Routledge.

Gallego, Á.J. 2018a. Freezing effects in a free-merge system. Hartmann, J.M., Jäger, M., Kehl, A., Konietzko, A., & Winkler, S. (Eds). *Freezing: Theoretical Approaches and Empirical Domains*. Berlin: De Gruyter Mouton. 66–104.

——. 2018b. Projection without agreement. *The Linguistic Review* 35(4): 601–623.

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

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

Hicks, G. 2009. *The Derivation of Anaphoric Relations*. Amsterdam: John Benjamins.

Hornstein, N. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.

Hosono, M. 2016. Exceptional movement from/into the criterial position. *Working Papers in Scandivanian Syntax* 97: 23–39.

——. 2018. Constraints on movement. *Working Papers in Scandivanian Syntax* 100: 41–63.

Ke, A.H. 2017. Full phase transfer. Ms. University of Michigan.

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

Komachi, M., H. Kitahara, A. Uchibori, and K. Takita. 2019. Generative Procedure Revisited. *Reports of the Keio Institute of Cultural and Linguistic Studies* 50, 269-283.

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

Marantz, A. 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.

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

——. 2019b. Ambiguous labeling and full interpretation. *Studia Linguistica: A Journal of General Linguistics*. doi.org/10.1111/stul.12109.

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

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

——. 2015b. Reference, phases and individuation: topics at the labeling-interpretive interface. *Opticon1826* 17(5): 1–13.

——. 2018. No country for Oldowan men: self-domestication and cranial globularity as factors in language evolution. *UCL Working Papers in Linguistics* 30: 35-56.

Narita, H. 2014. *Endocentric Structuring of Projection-free Syntax*. Amsterdam: John Benjamins.

Narita, H., & Fukui, N. 2020. *Symmetry-Driven Syntax: An Inquiry Into Endocentricity and Symmetry in Human Language*. London: Routledge.

Narita, H., Kasai, H., Kato, T., Zushi, M., & Fukui, N. 2017. 0-Search and 0-Merge. In Fukui, N. (Ed.). *Merge in the Mind-Brain.* New York: Routledge. 127–154.

Ouali, H. 2008. On C-to-T φ-feature transfer: The nature of agreement and anti-agreement in Berber. In Roberta D'Alessandro, Fischer, O., & Hrafnbjargarson, G. H. (Eds.). *Agreement restrictions*. Berlin: Walter de Gruyter. 159-180.

Preminger, O. 2014. *Agreement and its Failures*. Cambridge, MA: MIT Press.

Reuland, E. 2011. *Anaphora and Language Design*. Cambridge, MA: MIT Press.

Richards, N. 2016. *Contiguity Theory*. Cambridge, MA: MIT Press.

Rizzi, L. 2006. On the form of chains: criterial positions and ECP effects. Cheng, L.L.-S., & Cover, N. (Eds). *Wh-Movement: Moving On*. Cambridge, MA: MIT Press. 97–133.

——. 2015. Cartography, criteria, and labeling. Shlonsky, U. (Ed). *Beyond Functional Sequence: The Cartography of Syntactic Structures, Volume 10*. Oxford: Oxford University Press. 314–338.

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

Sauseng, P., Peylo, C., Lena Biel, A., Friedrich, E.V.C., & Romberg-Taylor, C. 2018. Does cross-frequency phase coupling of oscillatory brain activity contribute to a better understanding of visual working memory? *British Journal of Psychology* doi:10.1111/bjop.12340.

Shim, J.-Y. 2013. External merge by phase: its implications for feature-inheritance, transfer and internal merge. Lee, I.-J., & Uujinbai, D. (Eds). *Proceedings of the 15th Seoul International Conference on Generative Grammar (SICOGG): Universals and Parameters*. Seoul: Hankuk Publishing Co. 359–387.

——. 2018. <φ, φ>-less labeling. *Language Research* 54(1): 23–39.

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

Takita, K. 2018. Labeling for linearization. Ms. Meikai University. Available at: ling.auf.net/lingbuzz/004152.