Linguistic Research 37(3): 439–475 DOI: 10.17250/khisli.37.3.202012.003

A-movement: Its successive cyclicity revisited*

Manabu Mizuguchi (Toyo University)

Mizuguchi, Manabu. 2020. A-movement: Its successive cyclicity revisited. *Linguistic Research* 37(3): 439-475. This paper reconsiders the successive cyclicity of A-movement, which has been controversial in the literature. I argue that contrary to what has been argued, A-movement can be both successive cyclic and non-successive cyclic, showing that this proposal follows as one consequence of simplest Merge, which applies freely. I claim that whether A-movement proceeds successive cyclically or not depends on how Merge applies to C and T (as well as to v and R) in the derivation. I show that the discussion in the paper is cross-linguistically endorsed. It is also shown that the proposal has favorable implications for clausal construction, the labelability of T and *wager*-class sentences. The present paper is one illustration of Merge playing a key role in syntactic derivation, supporting the hypothesis that the operation is the core of the Faculty of Language. **(Toyo University)**

Keywords recursive Merge, adjunction, composite heads, cyclic Transfer, clausal architecture, labeling

1. Introduction

The successive cyclicity of A-movement has been controversial and there are arguments for and against successive cyclic A-movement in the literature. For instance, Bošković (2002a) gives examples such as (1) and (2), and argues that the intended interpretations are possible only when the surface subject, as shown in (3), moves successive cyclically through the edge of the embedded clause (conveniently marked as "S" in (3)) and is interpreted in intermediate positions, which are marked as " $_$ " in (1) and (2):

^{*} I would like to thank two anonymous reviewers for helpful questions and suggestions. I also thank the audience at Arizona Linguistics Circle (ALC) 13, where an earlier version of this paper was presented, for comments and discussions. The present paper is based on the project supported by Japan Society for the Promotion of Science (JSPS) under Grant-in-Aid for Young Scientists (B) (#16K16835). Needless to add, all the remaining errors and inadequacies are solely my own.

- (1) a. Mary seems to John [____ to appear to herself to be in the room].
 b. *Mary seems to John [____ to appear to himself to be in the room]. (Bošković 2002a: 179-180)
- (2) [His_i mother's_j bread] seems to every man_i [______ to be known by her_j to be the best there is].
 (Bošković 2002a: 180)
- $(3) [NP [\dots [s [\dots [s [\dots t \dots]]]]]]$

In (1a), the copy in the intermediate position, which is the closest subject, can bind the reflexive in the embedded clause while in (1b), it blocks binding by *John*, causing a Specified Subject Condition effect.¹ In (2), the co-indexed reading will be possible if and only if the surface subject reconstructs to the underlined position, which indicates that its copy is created in that position through successive cyclic movement.

On the other hand, there are also arguments that A-movement is not successive cyclic; it proceeds in a single leap (e.g., Boeckx 2000; Castillo et al. 2009; Epstein and Seely 2002, 2006; see also Epstein et al. 2005 for relevant discussion):

 $(4) \quad [\underset{\bigstar}{\text{NP}} [\ldots [s \ldots [s \ldots t \ldots]]]]$

For instance, Boeckx (2000) and Castillo et al. (2009) claim that the EPP should be wiped out, with no EPP on intermediate T, which leads to the absence of successive cyclicity in A-movement. Epstein and Seely (2002, 2006) argue that A-chains should be eliminated, which is to say that there are no intermediate steps in the process of A-movement. Castillo et al. (2009) and Epstein and Seely (2002, 2006) demonstrate that examples such as (1) and (2) can be explained without successive cyclic A-movement (or without intermediate copies).

(i) a. Pictures of *himself* seem to *John* to be cheap.
 b. *It seems to *him* that *John* is in the room.

(Bošković 2002a: 179)

¹ The NP can c-command out of the experiencer phrase. Consider (i):

In (ia), the reflexive is bound by *John* in the reconstructed position of *[pictures of himself]* and in (ib), a Condition C violation is incurred due to *him* binding *John*. These examples show that the NP embedded in the experiencer phrase can c-command out.

The purpose of this paper is to reconsider in the recent Minimalist setting this long-standing debate on the derivation of A-movement. Assuming that Merge is the core of the Faculty of Language (FL) (Chomsky 2000; especially Chomsky 2010 and his subsequent papers), I claim that A-movement is not an "either-or" movement: it can be both successive cyclic and non-successive cyclic, providing a fresh perspective on the debate. I argue that whether A-movement proceeds successive cyclically or not depends on how Merge applies, showing that the operation plays a key role in determining the successive cyclicity.

The organization of this paper is as follows: section 2 first discusses theoretical background in this paper. Section 3 spells out my claim in the paper, arguing that the successive cyclicity is deduced from Merge. Section 4 shows that the proposed analysis is cross-linguistically endorsed. Section 5 discusses three implications of the proposal. Section 6 summarizes and concludes the paper.

2. Theoretical background

In this section, I discuss theoretical background and pave the way for my claim in the next section. It has been argued that language is a system that yields a digitally infinite array of hierarchically structured expressions or syntactic objects (SOs) with systematic interpretations at interfaces with the Conceptual-Intentional (CI) system and the Sensory-Motor (SM) system (e.g., Chomsky 2017 among others): it is a computational system with discrete infinity, having meaning and sound at the same time. The Minimalist Program attempts to achieve an explanatory theory of language through reduction or minimization, and the basic assumption is to seek the simplest account of FL or Universal Grammar (UG): UG must be quite simple at its core. This assumption is natural and reasonable for scientific and biological reasons, given that linguistics is a branch of science and that language is a biological endowment: scientifically, simpler theory suggests deeper explanation and leads to better understanding of the object of inquiry; in any scientific inquiry, less is better than more. Biologically, simplification of UG, which is an evolutionary product, will make it easier to attain an eventual account of the origin of language as far as this can

be attained (Chomsky 2015, 2017, 2019a). Considering the basic property of language, FL must at least have as its ingredients the operation that generates SOs, which is called "Merge," and CI and SM interfaces. This is summarized as (5):

(5) Language = Merge + interfaces

In search of the simplest account of FL, it can reasonably be assumed that the two ingredients are the elements of principled explanation: the properties of language are explained by these ingredients alone.

If UG is simply designed, its principles should work in a simple or efficient manner and it can reasonably be assumed that the Strong Minimalist Thesis (SMT) holds for language:

(6) The Strong Minimalist Thesis

Principles of UG operate in accord with language-independent conditions of computational efficiency (or "third factor" principles – Chomsky 2005). (Chomsky 2017: 296)

SMT says that language is perfectly designed, being a perfect solution to conditions imposed by the interfaces. Given SMT, the structure-building operation Merge is formulated as "simplest Merge":

(7) Merge: α , $\beta \rightarrow \{\alpha, \beta\}$

Simplest Merge is symmetric merge of any two elements in the lexicon or in the workspace, which forms a new set out of them.²

² Chomsky (2019a) and Chomsky et al. (2019) propose that Merge is an operation on the workspace (WS), which represents the stage of the derivation at any given point, not on lexical items or SOs as in (7), mapping the WS into a new WS (WS'); Merge in this sense is capital Merge or MERGE:

⁽i) MERGE: WS = $[\alpha, \beta] \rightarrow WS' = [\{\alpha, \beta\}]$

Under MERGE, square brackets represent WS and curly brackets represent syntactic objects ([...] \neq { ... }). Merge in this paper can also be understood as MERGE. Keeping this in mind, traditional representations are used for the purpose of simplicity in the paper.

The final assumption is that clauses are characterized by both C and T. Following Chomsky (1981), Epstein and Seely (2006), Ormazabal (1995), Pesetsky (1992, 2016, 2019), Tanaka (2002) among many others, I assume that both C and T are elements of clausal architecture: when clauses are formed, the two heads are selected and merged in the derivation.

With the above theoretical background in place, I now turn to section 3 and consider how A-movement proceeds in the derivation.

Simplest Merge and A-movement

3.1 Proposal

In this section, I claim that Merge explains the successive cyclicity of A-movement. Given simplest Merge in (7), unless otherwise stipulated, Merge will apply freely and can produce SOs or sets shown in (8):

(8) a. $\{\chi, \{\alpha, \beta\}\}$ b. $\{\alpha, \{\alpha, \beta\}\}$

In (8a), α and β are merged to form a set, with which another SO χ is merged to form another set; on the other hand, in (8b), Merge forms another set by taking α and recursively (or internally) merging it with the set formed out of α and β , embedding the set under α . Notice that the set (8b) is an adjunction structure, yielding the effect that β is adjoined to α : in (8b), β is put asymmetric to α by Merge due to α embedding the { α , β } set or β ; adjunction follows from self-embedding. Recursive Merge can produce adjunction (=(8b)) as well as substitution (=(8a)) in earlier frameworks.

As regards adjunction, Chomsky (2015) (also Chomsky 2004) proposes that it is the result of pair-Merge, which, as shown in (9), creates an ordered pair, not a set, and is considered a formally distinct operation from simplest Merge:

(9) <α, β>

The ordered pair $\langle \alpha, \beta \rangle$ is not on a par with the ordered pair $\langle \beta, \alpha \rangle$ ($\langle \alpha, \beta \rangle$ $\neq <\beta$, $\alpha >$), which suggests that α and β are asymmetric, with one adjoined to the other. As I have argued, given simplest Merge, Merge yields adjunction when applied in the way shown in (8b); recursive or internal merge of α with { α , β } makes β asymmetric to α as it embeds β under α . Adjunction follows as one consequence of simplest Merge and there is no need to assume pair-Merge in addition to simplest Merge (see also Chomsky 2019a and Chomsky et al. 2019 for arguments in favor of this conclusion). This argument is also supported by mathematical definitions of an ordered pair: Kuratowski (1921) proposes that it is equivalent to { $\{\alpha\}, \{\alpha, \beta\}$ } while Tourlakis (2003) argues that it is on a par with $\{\alpha, \{\alpha, \beta\}\}$. As regards this, Chomsky (2012) argues that a singleton set is syntactically identical to its member ($\{\alpha\} = \alpha$). In syntax, α can be a head as well as a phrase. Hence, α in { α { ... }} is " α " as argued in Tourlakis when it is a head; on the other hand, it is " $\{\alpha\}$ " as proposed in Kuratowski when it is a phrase. For the purpose of discussion, I use α instead of $\{\alpha\}$ in this paper. See also Fukui (2017) and Omune (2018) for relevant discussion. In this paper, only for ease of distinction and for illustration, I use the term "pair-Merge" for the merge that yields an adjunction structure or the set (8b), and the ordered pair (9) is employed for (8b) for ease of exposition.³

With (8) in place, now consider A-movement in (10):

(10) The student seems to be in the library.

In the derivation of the embedded clause, simplest Merge allows two pair-Merge options given that C and T are both merged to form a clause, and (11a) and (11b) can be derived as the embedded clause of (10):

(11) a. $[_{\delta} < C, T > [_{\alpha} \dots]]$ b. $[_{\delta} < T, C > [_{\alpha} \dots]]$

³ In { α , β }, α and β are symmetric ({ α , β } = { β , α }); hence, (8b) is no different from { α , { β , α }}. Likewise, { α , { α , β }} is no different from {{ α , β }, α }. Keeping this in mind, for the purpose of illustration, a recursively merged element (a target of adjunction) is put to the left of a set in this paper.

In (11a), T is pair-merged to C and the derived SO is merged with the verbal set α while in (11b), C is pair-merged to T and <T, C> is then merged with α .⁴ I argue that A-movement will be successive cyclic if (11a) is generated. The ordered pair <C, T> or {C, {C, T}} is syntactically on a par with C, with T being syntactically invisible and de-activated, since C embeds the {C, T} set, hence T; minimal search will locate C as it is closer. It follows from self-embedding by α that α and β are asymmetric and that { α , { α , β } has the properties of α , with β being de-activated and invisible (< α , β > = { α , { α , β } = α). <C, T> is a composite head with the properties of C, not T:

 $(12) < C, T > (\{C, \{C, T\}\}) = C$

Given that C is a phase head, having phasehood as one of its properties, <C, T> has this property and works as a phase head, with its complement cyclically transferred at the phase level. Then in order to move out, the NP must go through the Spec of <C, T>; otherwise, it will be unavailable to the computation in the higher phase for phase impenetrability due to cyclic Transfer:

(13) [NP [... [
$$\mu$$
 t [δ ([α ... t ...]]]]]
Transfer

A-movement will be necessarily successive cyclic for phase impenetrability if Merge produces (11a).

The phase status of a raising complement, hence successive cyclic A-movement, is endorsed by the fact that quantifier raising (QR) out of *seem*-type raising infinitives is impossible. To see this, consider (14):

(14) a. Mary seems to two women to be expected to dance with every senator.

⁴ A reviewer wonders if there is any morphological evidence for the distinction between (11a) and (11b) in English or in other languages. As far as I am aware, the two cannot be morphologically distinguished in language but only theoretically. For relevant discussion, see Mizuguchi (2018b, 2019c), who argues that <C, T> and C can be morphologically distinguished.

- b. #This soldier seems to someone to be likely to die in every battle.
- c. #The ball seems to a boy to be under every shell.
 - cf. Every shell seems to a (different) boy to be over the ball.

(Wurmbrand 2013: 620)

In (14), a universal quantifier in the embedded clause cannot take scope over an existential quantifier in the higher clause (* \forall » \exists). On the assumption that QR is a syntactic movement obeying the phase impenetrability condition (Takahashi 2010), Wurmbrand (2013) argues that if the raising infinitive forms a phase, a quantifier will have to undergo successive cyclic QR via the phase edge in order to move out into the higher clause; however, QR to the phase edge in the raising infinitive will be semantically vacuous and given Scope Economy (Fox 2000), such QR cannot be operated. Phase impenetrability blocks a universal quantifier from moving in a single leap from its externally merged position into the next higher phase, even though the movement may not be semantically vacuous. Consequently, the quantifier cannot take scope over an existential quantifier in the higher clause. Wurmbrand's argument provides evidence for the phase status of a raising complement and hence for the claim that raising infinitives are phasal when Merge produces (11a).

Notice that A-movement will be successive cyclic when C and T are merged as in (15):

(15) [NP [... [
$$_{\sigma}$$
 t [$_{\kappa}$ C ($_{\lambda}$ T [$_{\alpha}$... t ...]]]]]]

In (15), λ , which is a complement of a phase head, is cyclically transferred at the phase level and in order to move out, the NP has to pass through Spec,CP for phase impenetrability, just as in (13). This movement process, however, is banned as it is improper movement (Chomsky 1973): A'-movement of a constituent X cannot be followed by movement of X to an A-position. Consider (16):⁵

⁵ It can be considered that improper movement is illegitimate as it yields a non-uniform chain and causes interpretive ill-formedness at the CI interface (Chomsky and Lasnik 1993; Fukui 1993). See also May (1979) for a Condition C analysis.

(16) Improper Movement

$$*[\underline{A} [\dots [\underline{A'} [\dots [\dots t \dots]]]]]$$

It has standardly been assumed that Spec, TP is an A-position while Spec, CP is an A'-position. However, A- and A'-positions should not be tied to certain structural positions given that not all A-positions are "argument" positions in generative syntax after the 1980s, when the VP-internal subject hypothesis was introduced. Instead, A- and A'-positions should be attributed to featural properties of a given head. It has been argued in the literature that φ -features play a key role in determining A-positions (Chomsky 2007; Kratzer 2009; Obata and Epstein 2011; van Urk 2015 among others). To endorse this argument, van Urk (2015) reports that A'-movement in Dinka is movement to Spec,CP but shows the typical properties of A-movement in that it suppresses Weak Crossover, feeds anaphor binding and lacks reconstruction for Binding Principle C. He argues that C in Dinka is a composite probe, having both φ -features and those features that drive A'-movement: movement to Spec, CP in Dinka involves an Agree relation in φ -features as well (van Urk 2015: 109). What is important for the present paper is that movement to a specifier shows the properties of A-movement to the extent that the movement involves φ -feature agreement between the head and a moving XP, no matter what head it is. Mizuguchi (2019a,c), building on these arguments, argues that a sister relation with a φ-bearing SO, which is the basic relation established by simplest Merge (=(17)), yields an A-position for the NP, and proposes (18):6



(18) The NP is in an A-position if it is merged with an SO headed by a head bearing φ-features; otherwise, it is in an A'-position.

(Mizuguchi 2019c: 335)

⁶ Note that it also follows that the complement or object position is an A-position since the NP is merged with V, which has φ -features for inheritance.

As a phase head, C has unvalued φ -features, which are inherited onto T (Chomsky 2008, 2019b): it derivationally loses φ -features via feature inheritance when merged into the derivation. Given (18), Spec,CP is an A'-position for lack of φ -features in C and Spec,TP is an A-position for T gaining the features through feature inheritance. Successive cyclic A-movement under (15) will be ruled out for improper movement.

On the other hand, improper movement is not a problem for (13). Unlike Spec,CP, the Spec of <C, T> constitutes an A-position. In <C, T>, T, to which C's φ -features are inherited, has become part of C by Merge; unlike in (15), <C, T> is a composite head, with T being inside. It has been assumed that feature inheritance from a phase head (P) to a non-phase head (\neg P) takes place in the configuration of [P [\neg P [...]]], one instance of which is [CP C [TP T [...]]]. Given that the configuration allows features on P to be inherited onto \neg P, φ -features will not be inherited from C in (13) in the absence of [CP C [TP T [...]]]. As a result, <C, T>, which is on a par with C (=(12)), has φ -features. Given (18), the Spec of <C, T> is an A-position. Movement in (13), unlike the one in (15), does not violate (16).⁷

Let us now consider A-movement under (11b). When (11b) is produced, A-movement will proceed in a single leap. In $\langle T, C \rangle$ or $\{T, \{T, C\}\}$, T embeds the $\{T, C\}$ set, hence C, with C being syntactically invisible and de-activated since minimal search locates T; $\langle T, C \rangle$ is on a par with T, being a composite head having the properties of T, not C:

 $(19) < T, C> ({T, {T, C}}) = T$

Unlike C, T is not a phase head and given (19), <T, C> is not a phase head, either. Then in (20), α , the complement of <T, C>, is not cyclically transferred in the course of the derivation and in the absence of phase impenetrability, the NP

⁷ One reviewer points out that (15) (or (43) in section 5.1) will always be prohibited due to the possibility of improper movement, asking if this prediction is correct. As far as I am aware, the prediction is correct: to the extent that the NP moves out to an A-position in a higher clause, clause extension like (15)/(43) will always be banned. For instance, Mizuguchi (2019a) argues that (15) explains the impossibility of French ECM; moreover, Mizuguchi (2018a,b) claims that in cases where improper movement is well-formed, (13), not (15), is generated, which he argues is morphologically marked by a distinct complementizer. See the references for details.

in α can move to Spec, TP from its externally merged position without undergoing successive cyclic movement by way of intermediate Specs:

(20) [NP [... [δ <T, C> [α ... t ...]]]]

One piece of evidence for the non-phase status of $\langle T, C \rangle$ is shown by long-distance agreement. Consider (21):

(21) There seem to be likely to be three men here. (Boeckx 2009: 2)

In (21), the φ -probe on the matrix T agrees with the NP *three men* in the embedded clause without movement, which suggests that there are no phase boundaries in between and hence that α is not rendered impenetrable by cyclic Transfer.

Another piece of evidence is provided by (22a):

(22) a. *[Which picture of himself] did Mary seem to John [to like t]?

cf.b. [Which picture of himself] did it seem to John [that Mary liked t]?

(Abels 2003: 30)

In (22a), unlike in (22b), the reflexive *hinself* in the moved *wh*-phrase cannot be bound by *John* in the higher clause. This shows that the *wh*-phrase moves without undergoing successive cyclic movement. If the *wh*-phrase moves successive cyclically in (22a), we would expect *hinself* to be bound by *John* in the edge of the embedded clause just as in (22b), where the embedded clause is CP, a phase, and successive cyclic movement via the phase edge is required for phase impenetrability:

(23) ... [seem to John [[which picture of himself] [Mary [to like t]]]]

(22a) argues for the non-phase status of <T, C>.

The argument here is also endorsed by Standard Arabic. In this language, post-verbal subjects trigger partial agreement (gender agreement) while preverbal subjects (i.e., subjects in Spec,TP) trigger full agreement (gender and number agreement). For instance, consider (24):

(24) a. ?akal-at (/*?akal-na) ţ-ţaalibaat-u.
eat.PAST-3.FEM.SG eat.PAST-3.FEM.PL the-students.FEM.PL-NOM
'The students ate.'
b. ţ-ţaalibaat-u ?akal-na (/*?akal-at). (Benmamoun 1992: 121)

(24a,b) show that full agreement is realized when Spec,TP is created while partial agreement appears when the Spec is not yielded (see also Mizuguchi 2017, 2019c and references cited therein for relevant discussion).

With this in mind, now consider (25):

(25) a	a.	?aw∫akna	(?an)	tanjaħ(u/a)	l-t ^s a:liba:t-u.
		were.about.to.3. F.PL	(C/to)	succeed.3.F.SG	the-students.F-NOM
1	b.	?aw∫akat	(?an)	tanjaħ(u/a)	l-t ^s a:liba:t-u.
		were.about.to.3. F.SG	(C/to)	succeed.3. F.SG	the-students.F-NOM
	'The female students were about to succeed.'				

(Alexiadou et al. 2014: 2-3)

In (25a), though the NP appears in the post-verbal position in the embedded clause just as in (25b), full agreement emerges in the matrix clause, which suggests that the NP moves to the matrix Spec,TP but that its copy is pronounced (cf. Alexiadou et al. 2014, who argue for covert movement in (25a)). Notice that in (25a), just as in (25b), where the subject does not move to Spec,TP in the matrix clause, partial agreement is observed in the embedded clause. Given that Spec is required for realizing full agreement, this suggests that the movement is not successive cyclic but proceeds in a single leap, skipping intermediate Spec positions. This is possible only in the absence of phase boundaries in the raising complement, which argues for the non-phase status of <T, C>.8

⁸ Given that φ -features are properties of phase heads and are inherited onto non-phase heads, I assume that the embedded <T, C> gains φ -features through multiple feature inheritance in Standard Arabic: φ -features on the matrix C are inherited to T and then copied onto <T, C> in the embedded clause, which is possible thanks to the absence of phase boundaries.

The examples considered above show that phase boundaries can be absent in the raising complement, and the non-phase status is explained by (11b).

It should be noticed that the absence of phase impenetrability, that is, (11b), does not warrant one fell-swoop movement; simplest Merge will allow the derivation (26), where the NP moves to the Spec of $\langle T, C \rangle$ in the first step (=①) and then it moves on to the matrix Spec,TP in the next step (=②):

$$(26) \left[\underbrace{NP}_{2} \left[\dots \left[\mu \right]_{\mu} t \left[\delta < T, C > \left[\alpha \ \dots t \right]_{\alpha} \right] \right] \right]$$

This derivation, however, is independently ruled out given efficiency principles. Bošković (2019) argues that asymmetric relations are maximized in language, proposing (27), which is traceable to labeling by minimal search (for details, see Dadan 2019):

(27) Maximize Asymmetric Relations (MAR) (Boškovič 2019: 2)

One consequence of (27), Bošković argues, is that Spec positions are avoided as much as possible, which suggests that successive cyclic movement, which yields a symmetric structure (i.e., XP-YP), occurs only when it is really necessary, namely, when it is forced by phase impenetrability; "free" successive cyclic movement is barred by MAR. In (26), phase impenetrability does not arise in the absence of a phase head in the embedded clause since <T, C> does not work as a phase head. The derivation (26) is precluded by MAR.

Moreover, (26) comes into conflict with Determinacy. Chomsky (2019a) and Chomsky et al. (2019) argue for Determinacy in light of efficient computation and propose (28):

(28) Simplest Merge maps WS = [X, Y] onto $WS' = [\{X, Y\}]$, reducing its complexity and avoiding indeterminate rule application.

(Chomsky et al. 2019: 246)

Derivation is deterministic to the extent that derived structures are unambiguous

to rule application in the subsequent derivation; any rule referencing X or Y would ambiguously refer to the individual objects X, Y or to the terms of $K = \{X, Y\}$, if Merge(X, Y) derives WS' = $[X, Y, \{X, Y\}]$. If the NP moves successive cyclically in (26), then it will produce (29) or the WS at one point of the derivation, in which a movement rule or Internal Merge (IM) will ambiguously refer to the boxed NPs in the following derivation:

(29) WS =
$$[_{\mu} NP [_{\delta} < T, C > [_{\alpha} ... NP ...]]]$$

(29) will be ruled out in violation of Determinacy and the movement process (2) cannot be executed.

Both MAR and Determinacy can be considered properties of language which are reducible to third factor principles. Given that UG is subject to SMT, (26) will be excluded by efficient computation, even though there may be no problem with the movement process in (26) under simplest Merge.⁹

Summarizing the discussion, I have claimed in this section that the successive cyclicity of A-movement follows from how C and T are merged in the derivation, with Merge determining the behavior of A-movement.

3.2 Phasehood in the verbal domain

Before we leave this section, one more thing needs to be addressed in considering the successive cyclicity of A-movement, from which I have abstracted away: that is, the phase status of vP. Given that CP and vP are parallel in the derivation, a reasonable assumption is that both v and R, which is categorized as V by v, are merged in as elements of the verbal domain, just as both C and T are merged in as elements of the clausal domain. If, as Legate (2003) argues, v is phasal whether it is transitive or not, then the NP will move in a single leap only when the verbal domain is not phasal. I argue that this is

⁹ Also, note that (26) will be improper movement. Recall that $\langle T, C \rangle$ is on a par with T. T does not have φ -features of its own but inherits them from C. When $\langle T, C \rangle$ is produced, there is no C in the derivation from which T or $\langle T, C \rangle$ can inherit φ -features and $\langle T, C \rangle$ lacks the features. Given (18), the Spec of $\langle T, C \rangle$ will be an A'-position for want of φ -features and movement to Spec,TP, an A-position, via this position will be improper movement (Mizuguchi 2019a). The NP must bypass the Spec of $\langle T, C \rangle$ to undergo "proper" movement.

achieved when *v* is pair-merged to R and $\langle R, v \rangle$ is produced in the verbal domain. The phase does not emerge and cyclic Transfer does not apply since $\langle R, v \rangle$ is on a par with R, a non-phase head, for *v* being embedded under R through recursive merge of R ($\langle R, v \rangle = \{R, \{R, v\}\} = R$) and does not work as a phase head.¹⁰ One fell swoop A-movement will be possible when Merge generates both $\langle T, C \rangle$ and $\langle R, v \rangle$:

(30) [NP [... [
$$\pi$$
 \nu \geq [δ [α ... t ...]]]]]

On the other hand, when R is pair-merged to *v*, with $\langle v, R \rangle$ created by Merge, then it works as a phase head since it is on a par with *v*, a phase head $(\langle v, R \rangle = \{v, \{v, R\}\} = v)$, and A-movement will be successive cyclic via the Spec of $\langle v, R \rangle$ for phase impenetrability. This argument is supported by examples like (31), where the surface subject can be interpreted under the scope of *not*, which comes between T and *v*P (Laka 1990; Pollock 1989), and at the same time, it can bind a pronoun in the higher clause:

(31) a. Every participant didn't seem to his coach [to be in bad shape].b. All linguists didn't seem to their employer [to work hard].

(Sauerland 2003: 311)

Under the assumption that negation does not move (see Sauerland 2003 for this assumption), these interpretations will be possible only when successive cyclic A-movement occurs in the verbal domain and creates a copy or Spec in an intermediate position, which is the Spec of $\langle v, R \rangle$ under the proposed analysis.¹¹ When either $\langle C, T \rangle$ or $\langle v, R \rangle$, or both are generated in the derivation, the NP will move successive cyclically. Notice that under the proposed analysis, as shown in (32), A-movement can be successive cyclic without the EPP effect at the embedded "Spec,TP" when $\langle v, R \rangle$ is produced while $\langle T, C \rangle$ is generated in the embedded clause:

¹⁰ See also Epstein et al. (2016) for the argument that <R, *v*≥ does not work as a phase head.

¹¹ Given (18), movement via the Spec of <v, R> is not improper movement since the composite head bears φ-features for pair-merge of R to v; φ-features in v are not inherited in the absence of [P [¬¬P [...]]] or [_{vP} v [_{RP} R ...]]. Recall that <C, T> has the relevant features for pair-merge of T to C.

454 Manabu Mizuguchi

(32) [NP [... [
$$\sigma$$
 t [$\pi < v_i$, R> [$\delta <$ T, C> [α ... t ...]]]]]]]

Unless constrained by selection or verbal properties, which prefer certain pair-Merge options (see sections 4 and 5.3 for relevant discussion), the successive cyclicity of A-movement in clausal and verbal domains will follow from freely applying simplest Merge.¹²

4. Movement vs. Agree languages

In this section, I argue that the analysis of A-movement proposed in this paper is cross-linguistically supported. Alexiadou et al. (2014) argue on the basis of raising constructions that there are two types of languages: movement languages and Agree languages. In movement languages, the NP must move to Spec,TP because there are phase boundaries between T and the original position of the NP, and long-distance agreement is blocked due to phase impenetrability (=(33a)); on the other hand, in Agree languages, the NP can agree with T in the in-situ position in the absence of phase boundaries (=(33b)). Consider (33):

(33) a. [TP NP [T [... [... [... t ...]]]]] b. [TP T [... [... [... NP ...]]]]

Alexiadou et al. argue that Greek is an Agree language. Consider the following Greek examples:

(i) {what, {R, what}}

¹² One reviewer notes that if the object moves to Spec, RP as Chomsky (2015) argues, then (i) will be created in the course of object *wh*-movement, which is nothing other than (8b):

The reviewer points out that given the conventional assumption of adjunction, when X is adjoined to Y (or vice versa), the two elements behave as if they are a single element. If so, neither *what* nor R can be extracted and he/she wonders how *what* can move up to Spec,CP. This may be true under the conventional assumption of adjunction but is not the case under the proposed deduction of adjunction. Notice that (i) or (8b) is a set, having the effect of adjunction due to α embedding { α , β }, hence β and being asymmetric to β . Unless transferred, elements in a set are all visible and can be subject to computation. Even if (i) is generated, *what* can move out of it without any problems.

(34) a. Arhise {*pikni} [na skepazi i skoni {pikni} ta epipla]. started dense.FEM SUBJ cover.3SG the dust dense the furniture 'The dust started to cover the furniture densely.'

b. Stamatise na perni mono i Maria kakus vathmus. stopped SUBJ get only Mary grades weak 'It stopped being the case that only Maria got bad grades.'

(Alexiadou et al. 2014: 3-4)

According to Alexiadou et al, in (34), the NP in the embedded clause cannot license a matrix modifier (=(34a)); nor can it take scope over a matrix predicate in (34b) (*stop*»*anly*, **anly*»*stop*), which they say suggests that the NP does not move but agrees with T in situ for lack of phase boundaries.

I argue that movement and Agree languages are explained by (11): in movement languages, R (= V) selects a head with the properties of C, hence (11a), while in Agree languages, it selects the one with the properties of T, hence (11b). Recall that $\langle C, T \rangle$ works as a phase head while $\langle T, C \rangle$ does not:

(35) a. [R [$_{\delta}$ <C, T> [$_{\alpha}$...]]] \rightarrow movement languages b. [R [$_{\delta}$ <T, C> [$_{\alpha}$...]]] \rightarrow Agree languages

Notice that selection, which can be reduced to CI interface properties (see Fortuny 2008 and Pesetsky 1982), is required in some form or other in order to constrain head-complement relations. Given this, the resort to selection to explain whether a language is a movement or Agree language is not at all a stipulation.

If movement and Agree languages are explained by (11) and if whether a language is a movement or Agree language depends on whether R selects (11a) or (11b), then it is predicted that there are languages where R can select both (11a) and (11b). I argue that English is one such language. Recall from section 2 that in English, successive cyclic A-movement and long-distance agreement are both observed. Consider the following examples, which are repeated here for convenience:

- (1) a. Mary seems to John [____ to appear to herself to be in the room].
 - b. *Mary seems to John [__ to appear to himself to be in the room].

456 Manabu Mizuguchi

(

	(Bošković 2002a: 179-180)
21) There seem to be likely to be three men here.	(Boeckx 2009: 2)

This argues that both (35a) and (35b) are possible in the language.

On the other hand, Alexiadou et al. (2014) say that English is a movement language. They argue, following Hazout (2004a,b), that there is no Agree relation in (21) for the intervention of phases. Instead, the expletive is first merged in vP and then moves to Spec,TP for Case-theoretic reasons, with the associate licensed in a subject (*there*)-predicate (associate) configuration; apparent long-distance agreement between the verb and the associate is a manifestation of subject-predicate agreement, in this case holding between the moved expletive and the predicate nominal (i.e., the associate).

However, there are also persuasive arguments in the literature that the expletive does not move at all (Bošković 2002a,b). If the expletive is directly merged in its surface position without movement, then the associate, hence apparent long-distance agreement, will not be licensed in a subject-predicate configuration locally in *v*P; instead, a probe-goal Agree relation with T must be employed, which will not be possible if English allows only (11a).

In relation to (21), Alexiadou et al. also give (36), which shows that agreement with the *there*-associate is optional, and impossible when associates are conjoined, saying that this contrasts sharply with Agree languages like Greek (=(37)). Consider the following examples:

- (36) a. Essentially there seems/seem to be five compelling issues that...
 - b. There seems/*seem to be a pirate and a knight at the party.
 - c. A pirate and a knight seem/*seems to be at the party.
- (37) a. Arxis-an na kalipt-un i skoni ke to xioni ta epipla. started-3PL SUBJ cover-3PL the dust and the snow the furniture
 b. Arxis-an/*e na lin-onte/*ete polla simantika provlimata. started-3PL/*3SG SUBJ solve-3PL/*3SG many important problems (Alexiadou et al. 2014: 6)

This is not a problem, however. Schütze (1999) claims that English is similar

to German in that it has two expletive constructions, one uniformly singular, where T is 35G by default, and the other taking associate agreement. To the extent that his argument is correct, then examples such as (36a,b), where 35G agreement appears, will not refute the proposal that an Agree relation is involved in the *there*-construction, hence that English is also an Agree language.¹³

In English, *there* associates are restricted to non-specific NPs, which suggests that *there* and its associates are in a relation and that associates are licensed through this relation. This *there* associate relation will follow naturally when the two are put in a single phase, in which they are interpreted. If this is correct, it will also provide an argument for English as an Agree language: unless (11b) or (35b) is generated, there would be a phase boundary between *there* and its associates, and they cannot be in a single phase when the expletive is merged in its surface position without movement.

Finally, if English is uniformly a movement language for the intervention of phases, then binding would be possible between *John* and *himself* in (22a), just as in (22b), since the *wh*-phrase cannot move in a single leap but as shown in (23), must move successive cyclically for phase impenetrability via the edge of the embedded clause:

- (22) a. *[Which picture of himself] did Mary seem to John [to like t]?
- cf.b. [Which picture of himself] did it seem to John [that Mary liked t]?

(Abels 2003: 30)

(23) ... [seem to John [[which picture of himself] [Mary [to like t]]]]

Under the proposal in this paper, movement and Agree languages discussed in Alexiadou et al. (2014) follow from the way Merge applies in the derivation: in (35a) or (11a), phases emerge while in (35b) or (11b), they do not. As regards the emergence or disappearance of phases, Alexiadou et al. propose that it is derived by a particular type of selection of the raising complement. Assuming that the highest projection of a cyclic domain constitutes a phase, where cyclic domains are defined as the extended projections of VP and TP, with TP without

¹³ Schütze (1999) discusses cases like (36b), where only one of the agreement options is available, arguing that such cases can be independently explained. The reader is referred to Schütze (1999) for this discussion.

CP being one such case (see also Bošković 2014), they argue, following Wurmbrand (2013), that infinitives (and subjunctives) with a specific selected tense value (say, irrealis) involve an obligatory selectional valuation relation between the matrix V and the highest head in the embedded clause and that this voids the phasehood of the raising complement. This, however, raises a why question: why does such an obligatory selectional valuation relation cancel phasehood? The argument just redescribes what needs to be explained and states in different words that the raising complement is not a phase. The analysis proposed in this paper answers the question why a particular type of selection of the raising complement leads to phase cancellation: the raising predicate in the matrix clause selects T with C pair-merged to it (=(11b)), which does not work as a phase head (recall that <T, $C > = {T, {T, C}} = T$) and hence does not induce phase impenetrability.

As regards disappearing phases in the verbal domain, Alexiadou et al. (2014) argue that its phasehood is voided by phase extension due to v/V-movement to T (see den Dikken 2007 and Gallego 2005, 2010); a phase is de-phased and becomes transparent to operations if a phase head is head-moved. This is illustrated in (38):

(38)
$$[_{\text{TP}} \xrightarrow{\text{T+}\nu/\text{V}} [_{\nu/\text{VP}} \xrightarrow{\text{t}}_{\nu/\text{V}} [_{\text{XP}} \dots]]]$$

head movement
non-phase

However, relying on head movement to explain phase cancellation in Agree languages is problematic. First, there are languages like English where v/V does not raise to T but still long-distance agreement is observed (see (21)). This may not be much of a problem for Alexiadou et al., however; recall that they assume that long-distance agreement does not take place in English raising constructions such as (21). A more serious problem is that unlike what Alexiadou et al. have in mind, head movement does not cancel phasehood. Head movement is understood as "internal" pair-merge of heads under the current theoretical framework (Chomsky 2015). To illustrate the derivation of (38), consider (39). In (38), v/V is merged prior to the merge of T (=(39a)). Assuming Bobaljik and Brown's (1997) approach to head movement, in which a head X moves to

another head Y before Y is merged into the derivation, v/V moves and is internally merged with T before T is merged in, which yields {T, v/V} (=(39b)); T is then recursively merged with {T, v/V} to form an adjunction structure ({T, {T, v/V}} = <T, v/V>) (=(39c)). Finally, {T, {T, v/V} is merged with (39a), which generates (39d), which is (38):

(39) a. [v/V [x_P ...]]
b. {T, v/V}
c. {T, {T, v/V}}
d. [{T, {T, v/V}}, [t_{v/V} [x_P ...]]]

Head movement creates an occurrence of v/V in its original, externally merged position. Notice that this occurrence, though it is a copy, is syntactically visible, given the definition of visibility in (40) (Chomsky 2013: 44):

(40) α is taken to be in the domain D if and only if every occurrence of α is a term of D.

Recall that in $\langle \alpha, \beta \rangle$, where β is pair-merged to α and is put asymmetric to α , β is syntactically invisible, being de-activated, because it is embedded under α : minimal search locates α and $\langle \alpha, \beta \rangle$ is syntactically equivalent to α . If so, the occurrence of v/V in $\langle T, v/V \rangle$ is not syntactically recognized, and v/V left in its original position is identified as the only occurrence of v/V. Consequently, given (40), it works as a visible head; in other words, (38) is on a par with (41), where head movement or internal pair-Merge does not apply:

(41) [TP T [v/VP V/V [XP ...]]]

Epstein et al. (2016) also argue that occurrences created by head movement are syntactically visible and Mizuguchi (2018a) empirically argues for this with Icelandic. Hence in (38), the phasehood of the verbal domain does not disappear.

On the other hand, this problem does not arise under the analysis proposed in this paper: two heads are not pair-merged as in (39), where one of them is part of the existing SO, but they are both taken from the lexicon to be pair-merged, which does not produce a visible occurrence of a phase head in the original position.

5. Implications

In the discussion so far, I have argued that the successive cyclicity of A-movement is the result of how C and T (also, v and R) are pair-merged in the derivation (=(11a,b)) and is traceable to simplest Merge (=(7)). Assuming that the discussion is correct, three implications follow from the proposed analysis and in this section, I discuss them one by one, showing that they are theoretically and empirically favorable.

5.1 Clause reduction

The first implication is that Merge can not only expand the clause; it can also reduce the clause. Under the proposed analysis, a head can become part of another head thanks to recursive merge of the latter with the set formed out of the two heads (=(8b)). In other words, the merge yields a composite head. On the assumption that both C and T are ingredients of clausal architecture, if the two heads are merged in the way shown in (8b) or pair-merged, the clause will be reduced in the sense that there is only one set above the verbal set α in the clausal domain, and we get a smaller-than-full clause; on the other hand, if C and T are merged as in (8a), we get a full-sized clause in that there are two sets above α . Consider (42) and (43):



The discussion in this paper argues that both clause expansion and clause reduction are explained by Merge: the clause size follows from how heads are merged in the derivation.

This implication suggests that Merge can derive clause-type differentiation. It has been observed in the literature that non-finite complements are different from finite ones in that they allow subject extraction; likewise, the subject can move out of the raising complement while it cannot out of the control complement. Consider the following examples:

- (44) a. *The student seems [that is in the library].
 - b. The student seems [to be in the library].
- (45) a. *Which student does the professor believe [that is the most intelligent]?
 - b. Which student does the professor believe [to be the most intelligent]?
- (46) a. *The student was tried [to read the book].
 - b. The professor is likely [to come to the conference].

It has been argued in the literature that the contrasts in (44)-(46) are due to lack of CP in raising/ECM complements; lack of CP makes subject extraction possible (see, especially, Pesetsky 2016, 2019 for this argument). This is to say that the clause is smaller than full or it is reduced in raising/ECM infinitives. Proposals have been put forward to explain the reduction. For instance, Chomsky (1986) makes a lexical stipulation that the raising/ECM complement is TP while Chomsky (1981) and Pesetsky (2016, 2019) argue that CP is built but is syntactically erased by a deletion operation (see also Chomsky 2015 for this argument, which he argues explains (i) in footnote 14). Notice that these assumptions to explain clause-type differentiation are not necessary given the proposal in this paper. As shown in (42), Merge can explain clause reduction. This conclusion is also favorable in that deletion operations are unfavorable to third factor principles in that they violate the No Tampering Condition (NTC):

(47) Merge of X and Y leaves the two SOs unchanged. (Chomsky 2008: 138)

As with Chomsky (1981) and Pesetsky (2016, 2019), this paper argues for a derivational approach to clausal architecture but unlike in such works, the

structure is reduced by nature (before birth or by Merge), not afterwards (after birth or by way of deletion). To the extent that clause reduction is the key to the extraction contrasts in (44) through (46), clause-type differentiation, hence unique properties of raising/ECM complements, is explained by Merge.¹⁴

5.2 T as a label

The second implication is that T can work as a label on its own. Chomsky (2015) claims that T is subject to label weakness (=(48)). He argues that (48) explains subject raising or the creation of Spec,TP (hence, "the EPP" – Chomsky 1981, 1982) in examples such as (49):

(48) T by its nature is too weak to serve as a label on its own; in order to work as a label, T must have overt or visible Spec,TP.

(Chomsky 2015: 9-10)

(49) The book was written t by the professor.

As shown in (50), unless overt/visible Spec,TP is created, a T-headed set or λ will not be labeled due to label weakness of T, which violates Full Interpretation given that all SOs must reach the interfaces with labels (Chomsky 2015); on the other hand, T with Spec,TP strengthens as a label and can label λ :

(50) a. $[_{\lambda} T \text{ [be written the book by the professor]]} (\lambda = ?)$ b. $[_{\gamma} \text{ The book } [_{\lambda} T \text{ [be written t by the professor]]]} (\lambda = T)$

¹⁴ The argument here also extends to the contrast observed in (i):

⁽i) a. *Which student does the professor believe [that is the most intelligent]?b. Which student does the professor believe [is the most intelligent]?

Chomsky (2015) and Pesetsky (2016, 2019) argue that in (ib), the embedded clause is analyzed on a par with a raising/ECM complement: that is, it is not CP but TP, which is derived through syntactic deletion of CP. See also Bošković (1997), Doherty (2000) and Ishii (2004) for this argument. Given the analysis in this paper, (ib), hence the reduced complement, is explained by Merge (=(42)). See Mizuguchi (2018b, 2019c) for the discussion of (i) as well as (44)-(46) under this analysis and why clause reduction is required for subject extraction. See also Erlewine (2017) for relevant discussion.

With (48) in mind, recall that $\langle T, C \rangle$ is syntactically on a par with T and has the properties of T. Then in (20), which is repeated below as (51), δ will not be labeled in the absence of overt/visible Spec if T is subject to label weakness:

(51) [NP [... $[_{\delta} < T, \bigcirc [_{\alpha} ... t ...]]]]$ ($\delta = ?$) (=(20))

As far as the discussion in this paper is correct, it suggests that T can label on its own.

The implication that T is a labelable head is desirable in that it can eliminate the strong/weak distinction on the labelability of T. As has been pointed out in the literature, the distinction runs the risk of stipulation: it simply redescribes what needs to be explained (see, e.g., Gallego 2017; Goto 2017 and Mizuguchi 2017 for problems with such a distinction). For instance, Chomsky (2015) argues that null-subject languages like Italian and Spanish have "strong" T in that it can label without overt/visible Spec:

(52) a. Ha parlato.	Italian
has spoken	
'He spoke.'	(Burzio 1986: 85)
b. Hemos trabajado todo el día.	Spanish
have worked all the day	
'We have worked all day.'	(Perlmutter 1971: 103)

According to Chomsky, T with rich agreement is a labelable head. However, this does not explain at all, leaving unanswered the question why richness of agreement makes T labelable. Likewise, Epstein et al. (2014: 471), discussing (53), argue that non-finite T, unlike finite T, can label on its own: 'Minimal search finds the only visible head T as the label of α , and a completely labeled, hence Full-Interpretation-compliant, CI representation is generated':

(53) There is likely $[\alpha$ to be a man in the room].

But why is non-finite T, unlike finite T, not subject to (48)? All these arguments suggest that the distinction is just stipulated.

464 Manabu Mizuguchi

At the same time, however, the argument that T (or for that matter, any head) is not subject to label weakness will now raise the following question: why is T labelable? Simply saying that T can label is also a stipulation. The labelability of a head should not be a product of stipulation; instead, it should be explained in a principled manner. Mizuguchi (2017) claims that labelability is reducible to interpretability (i.e., Full Interpretation), arguing that label weakness or unlabelability is due to unvalued features on heads: labels must be interpretable and heads with unvalued features, which are uninterpretable, cannot serve as labels. He proposes (54):

(54) Heads can label only when they are without unvalued features. (Mizuguchi 2017: 331)

As discussed in section 2, the properties of language will receive the simplest, hence principled, account to the extent that they are explained by Merge and the interfaces (=(5)). Given (54), labelability is explained on the side of the interfaces. Recall that T does not have φ -features of its own but inherits them from C. In the derivation with <T, C>, there is no C from which T or <T, C> can inherit φ -features. It follows without any stipulation from (54) that T, hence <T, C>, is labelable on its own since it is free from any unvalued features including φ -features.¹⁵

5.3 Wager-class sentences

The final implication is that the proposed analysis can explain the peculiar behavior of *wager*-class verbs (Pesetsky 1992; Postal 1974, 1993). *Wager*-class infinitives, just like ECM infinitives, receive a propositional interpretation but unlike in ECM infinitives, the NP cannot halt in *wager*-class infinitives.¹⁶ Consider the following examples (examples in this section are cited from the references mentioned here):

¹⁵ Given (54), unlike Chomsky (2015), the EPP does not reduce to label weakness. See Mizuguchi (2017, 2019c), who argues that the EPP is explained by externalization. Likewise, under (54), R or V can be a label and R or <R, r> can label without overt/visible Spec.

¹⁶ See Lasnik (2001, 2004) for detailed arguments that the NP can stay in ECM infinitives and does not always undergo object raising into the higher clause.

- (55) a. *John wagered [Mary to have entered the room].
 - b. *He alleged [Melvin to be a pimp].
- (56) a. Mary was wagered [to have won the race].
 - b. Melvin was alleged [to be a pimp].
- (57) a. Who did John wager [to be crazy]?
 - b. Who did they allege [to be a pimp]?
- (58) a. The professor believes [the student to be the most intelligent].
 - b. Mary made out [John to be a fool].

I argue that the peculiar behavior of *wager*-class verbs follows from (11b): such verbs, unlike ECM verbs, select only $\langle T, C \rangle$ as their complements. The movement of the NP will yield an XP-YP set in *wager*-class infinitives. Arguing that labeling is executed by minimal search, Chomsky (2013) claims that a symmetric XP-YP set (κ in (59)), in which the heads X and Y are equally close to minimal search, can only be labeled either when X and Y agree (=(59a)) or when either XP or YP moves out to turn into a copy, which is invisible to minimal search given (40) (=(59b)):

As discussed, T with C pair-merged to it or $\langle T, C \rangle$ is on a par with T and does not bear φ -features, which are inherent properties of C. If (11b) is generated in *wager*-class infinitives, then the NP cannot agree with $\langle T, C \rangle$ and μ in (60), which is an XP-YP set created by the movement, cannot be labeled:

(60) $[_{\mu} \text{ NP } [_{\delta} < T, C > [_{\alpha} \dots t \dots]]] \quad (\mu = ?)$

Given that all SOs must be labeled at the interfaces, (60) will be ruled out in violation of Full Interpretation. It follows that the NP cannot halt in the embedded clause or in the Spec of <T, C>.

On the other hand, <C, T> is on a par with C and can agree with the NP for having φ -features (which are not realized morphologically) (see also section 3.1). If (11a) is produced in ECM, then μ can be labeled thanks to agreement

even when XP-YP is structured in the ECM infinitive. The NP can halt in the embedded clause:¹⁷

(61) $[_{\mu} \text{ NP } [_{\delta} < C, T > [_{\alpha} \dots t \dots]]]$ (μ = labeled)

The ill-formedness of (55) follows from labeling failure in the absence of agreement, which is due to pair-merge of C to T (=(11b)).¹⁸

The discussion thus far, however, raises another question: why does the NP move *all the way up* into the CP phase in the higher clause? In other words, why can't the NP halt in the higher *v*P phase, which is possible with the ECM subject. Consider the following examples:

- (62) a. *John wagered [Mary [to have entered the room]].
 - b. *He alleged [Melvin [to be a pimp]].
- (63) a. I believed [Nixon, incorrectly, [to be interested in ending the war]].b. Mary made [John out [to be a fool]].

I argue that *wager*-class verbs are inherently <R, v>, where v is pair-merged to R, which embeds v under R ({R, {R, v}}). This has the effect that φ -features are missing from <R, v> as it is on a par with R. Given that unvalued features are properties of phase heads, with R inheriting φ -features from v just as T inherits the features from C, <R, v>, like <T, C>, does not have φ -features; there is no v in the derivation from which R or <R, v> can inherit φ -features. Then the XP-YP set in (62) or σ in (64) cannot be labeled in the absence of agreement between <R, v> and the NP, and hence the NP cannot halt in the higher vP

¹⁷ See Mizuguchi (2019a) for details of this argument.

¹⁸ The proposed analysis predicts that the NP can halt in the Spec of $\langle C, T \rangle$ in raising infinitives as $\langle C, T \rangle$ can be created and the NP can agree with it, with μ labeled thanks to agreement. As shown in (ia,b), however, such examples are ill-formed regardless of the expletive *it*.

⁽i) a. *(It) seems to be likely $[\mu$ the student [to [t understand the theory]]].

b. *(It) seems [$_{\mu}$ the student [to [be likely [t [to [t understand the theory]]]]]].

c. The student seems to be likely to understand the theory.

This problem can be solved independently. The reader is referred to Mizuguchi (2019b) for solutions to this problem.

phase:19

(64)
$$[\sigma \text{ NP } [\pi < R, \nu \geq [\delta < T, C \geq [\alpha \dots t \dots]]]]$$
 $(\sigma = ?)$

In *wager*-class sentences, unlike in ECM, the NP cannot stop in intermediate positions in the absence of agreement.

The proposed analysis of (62) further suggests that Case valuation is not contingent on φ -feature agreement. As discussed, <R, i is free from φ and cannot agree with the NP. Still, since (57) is well-formed, the Case feature of who in Spec, CP, which is the subject of the embedded clause, is valued and it is valued without agreement. I argue that like other unvalued features, unvalued Case can function as a probe (Bošković 2007). In the case of Case valuation, the value assigned depends on the properties of a head, instead of inherently valued Case features on a head (Chomsky 2000). Following Epstein et al. (2012) and Pesetsky and Torrego (2001), I assume that such properties are tense for nominative and transitivity for accusative. I also assume, following Carstens (2016), that if an unvalued feature does not find any source of valuation in its c-command domain at its first Merge, upward probing will come into play, searching the structure upwardly up to the point of Transfer (i.e., up to the point where probing cannot continue for phase impenetrability). With these assumptions in place, in (57), the unvalued Case feature of who will start probing upwardly since it cannot find tense/transitivity at its first Merge, continuing to probe until it finds such properties. As shown in (65), in which the matrix CP phase as a whole is subject to Transfer since it is a root (Chomsky 2004), the Case feature can find <R, *v*≥ with a transitivity property in this domain through upward probing, being valued as accusative:

(i) $\begin{bmatrix} V & V & RP & NP & R_{\phi} & I & ... & I \end{bmatrix}$

The verb or R moves to v, which produces the VO order through externalization.

¹⁹ Also notice that (64) will be ruled out at the SM interface since the NP comes before a *waga*-class verb and (64) cannot be spelled out or externalized as the VO order. This problem does not arise with ECM: in ECM, unlike in *waga*-class sentences, (i) can be structured since ECM predicates are not inherently <R, *ν*>. In (i), the NP moves to the Spec of RP, the head of which agrees with the NP thanks to φ-feature inheritance, with RP labeled for the agreement (Mizuguchi 2019a):



With independently motivated arguments in the literature, which are theoretically and empirically well-founded, Case valuation in (57) is not a problem without φ -feature agreement.

6. Conclusion

This paper has discussed the successive cyclicity of A-movement, which has been controversial in the literature. Contrary to what has been argued, I have maintained that A-movement can be both successive cyclic and non-successive cyclic; it is not an "either-or" movement. The key to determining the successive cyclicity is Merge, which applies freely: whether A-movement proceeds successive cyclically or not depends on how Merge applies to C and T (as well as to *v* and R) in the derivation. I have demonstrated that the discussion in this paper is cross-linguistically supported by movement and Agree languages. As far as the proposed analysis is on track, it has implications for clausal construction, the labelability of T and *wager*-class sentences. In conclusion, this paper has demonstrated that Merge plays a key role in syntactic derivation, which supports the minimalist hypothesis that the operation is the core of language.

The present paper leaves one question that is yet to be answered. Recall that in languages like English, two pair-Merge options in (11) are freely available. If so, (11a) would be possible for examples like (21) and (11b) would be possible for examples such as (14). As discussed, both movement and Agree patterns are empirically possible in English and are explained by (11a,b); however, only a certain pair-Merge option will emerge in (14) and (21). As one of the reviewers correctly points out, it looks like the choice between (11a) or (11b) for each derivation can only be known after all the derivational steps, requiring look-ahead. I argue that look-ahead is not necessary. Under the assumption of optimally designed syntax, as far as syntax abides by SMT, it can freely generate derivations and does so without caring about their outcomes, which are to be ruled in or out independently at the interfaces when transferred: syntax is neither crash-proof nor yields only legitimate interpretations. Hence, there is no need to have look-ahead to avoid illegitimate or ill-interpreted derivations. Suppose that (11a) is yielded in (21). In this case, the derivation will fail at the interfaces since it contains features that are left unvalued at the interfaces in the absence of an Agree relation: T cannot agree with the NP for the intervention of phases between them. This argument also applies to (1) and (2): in these examples, if (11b) is produced, the outcomes will only be ill-interpreted. Notice that in these derivations, syntax is compliant to SMT but they happen to be ruled out at the interfaces.

The discussion above also answers the question of what it is that determines clause reduction (42) or clause extension (43). Under the framework in this paper, both (42) and (43) are possible for all complements. The choice between the two, I argue, independently follows from the interfaces. As discussed in 3.1, if (43) is produced for raising complements, the derivation produced by optimal computation will fail for improper movement; likewise, as Mizuguchi (2019a) discusses, if the clause is not reduced in the ECM complement, the ECM subject cannot agree with the higher *v*, with its Case feature transferred unvalued as a result. On the other hand, in the case of control complements, I suggest that control predicates select C without φ : if <C, T> is produced, φ will not be inherited and remain on C or <C, T>, which violates the selectional relation at the interfaces; likewise, the selection will not be satisfied if <T, C> is yielded.

These explanations, however, may not go through with (14) and (22a). I have to admit that these are left unsolved in the paper. Take (14). Suppose that (11b) is generated, which allows the embedded quantifier to move into the matrix clause without undergoing semantically vacuous movement. At the present stage of research, I do not have an answer to the question why (11b) is not possible for (14). As I have discussed with (21), empirical examples suggest that (11b) is possible with the *sæm*-type predicate. Moreover, the following example shows 470 Manabu Mizuguchi

that the embedded quantifier can take scope over the matrix quantifier:

(66) Someone seems to love everyone. (Aoun and Li 1993: 65)

Both $\forall *\exists$ and $\exists *\forall$ readings are possible for (66), which demonstrates that *everyone* can move into the matrix clause without undergoing semantically vacuous movement; in other words, it moves in one fell swoop into the matrix clause, suggesting that (11b) is yielded. Given that the same predicate is employed, selection is not the answer to the question. Many complications are involved in sentences with quantifiers and binding (see Epstein and Seely 2002, 2006 for relevant discussion), which makes the answer beyond the scope of this paper. For this reason, I have to leave further discussion for another occasion.

References

- Abels, Klaus. 2003. *Successive cyclicity, anti-locality, and adposition stranding.* PhD Dissertation. University of Connecticut, Storrs.
- Alexiadou, Artemis, Elena Anagnostopoulou, and Susi Wurmbrand. 2014. Movement vs. long distance Agree in raising: Disappearing phases and feature valuation. In Hsin-Lun Huang, Ethan Poole, and Amanda Rysling (eds.), *Proceedings of the Forty-Third Annual Meeting of the North East Linguistic Society*, 1-12. Amherst, MA: University of Massachusetts, Graduate Linguistics Student Association (GLSA).
- Aoun, Joseph E. and Yen-hui Audrey Li. 1993. *The syntax of scope* Cambridge, MA: MIT Press.
- Benmamoun, Elabbas. 1992. The feature structure of functional categories: A comparative study of Arabic dialects. Oxford: Oxford University Press.
- Bobaljik, Jonathan D. and Samuel Brown. 1997. Interarboreal operations: Head movement and the extension requirement. *Linguistic Inquiry* 28(2): 345-356.
- Boeckx, Cedric. 2000. EPP eliminated. Ms. University of Connecticut, Storrs.
- Boeckx, Cedric. 2009. On long-distance Agree. Iberia: An International Journal of Theoretical Linguistics 1(1): 1-32.
- Bošković, Željko. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, MA: MIT Press.
- Bošković, Željko. 2002a. A-movement and the EPP. Syntax 5(3): 167-218.
- Bošković, Željko. 2002b. Expletives don't move. In Masako Hirotani (ed.), Proceedings of the

Thirty-Second Annual Meeting of the North East Linguistic Society, 21-40. Amherst, MA: University of Massachusetts, Graduate Linguistics Student Association (GLSA).

- Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38(4): 589-644.
- Bošković, Żeljko. 2014. Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis. *Linguistic Inquiry* 45(1): 27-89.
- Bošković, Željko. 2019. Generalized asymmetry. Ms. University of Connecticut, Storrs.
- Burzio, Luigi. 1986. Italian syntax: A government-binding approach Dordrecht: Reidel.
- Carstens, Vicki. 2016. Delayed valuation: A reanalysis of goal features, "upward" complementizer agreement, and the mechanics of case. *Syntax* 19(1): 1-42.
- Castillo, Juan Carlos, John E. Drury, and Kleanthes K. Grohmann. 2009. Merge over Move and the Extended Projection Principle: MOM and the EPP revisited. *Iberia: An International Journal of Theoretical Linguistics* 1(1): 53-114.
- Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson and Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232-286. New York: Holt, Rinehart and Winston.
- Chomsky, Noam. 1981. Lectures on government and binding: The Pisa lectures. Dordrecht: Foris.
- Chomsky, Noam. 1982. Some concepts and consequences of the theory of government and binding. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1986. Knowledge of language: Its nature, origin, and use. New York: Praeger.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels, and Juan Uriagereka (eds.), Step by step: Essays on minimalist syntax in honor of Howard Lasnik 89-155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and beyond: The cartography of syntactic structures*, Volume 3, 104-131. Oxford: Oxford University Press.
- Chomsky, Noam. 2005. Three factors in language design. Linguistic Inquiry 36(1): 1-22.
- 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*, 1-29. Berlin: Mouton de Gruyter.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos. P. Otero, and Maria Luisa Zubizarreta (eds.), Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud, 133-166. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2010. Some simple evo devo theses: How true might they be for language? In Richard K. Larson, Viviane Déprez, and Hiroko Yamakido (eds.), *The evolution of human language: Biolinguistic perspectives*, 45-62. Cambridge: Cambridge University Press.
- Chomsky, Noam. 2012. Poverty of the stimulus: Willingness to be puzzled. In Massimo

Piattelli-Palmarini and Robert C. Berwick (eds.), *Rich languages from poor inputs*, 61-67. Oxford: Oxford University Press.

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

- Chomsky, Noam. 2015. 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. 2017. Language architecture and its import for evolution. Neuroscience and Biobehavioral Reviews 81: 295-300.
- Chomsky, Noam. 2019a. Some puzzling foundational issues: The Reading program. *Catalan Journal of Linguistics Special Issue, 2019*: 263-285.
- Chomsky, Noam. 2019b. Puzzles about phases. In Ludovico Franco and Paolo Lorusso (eds.), *Linguistic variation: Structure and interpretation*, 163-168. Berlin: Mouton de Gruyter.
- 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, 2019*: 229-261.
- Chomsky, Noam and Howard Lasnik. 1993. The theory of principles and parameters. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, Volume 1, 506-569. Berlin: Walter de Gruyter.
- Dadan, Marcin. 2019. Loss of movement and labeling: Grammatical pressure and diachronic change. In Richard Stockwell, Maura O'Leary, Zhongshi Xu, and Z.L. Zhou (eds.), Proceedings of the 36th West Coast Conference on Formal Linguistics, 109-114. Summerville, MA: Cascadilla Proceedings Project.
- den Dikken, Marcel. 2007. Phase extension: Contours of a theory of the role of head movement in phrasal extraction. *Theoretical Linguistics* 33(1): 1-41.
- Doherty, Cathal. 2000. *Clauses without 'that': The case for bare sentential complementation in English* New York: Garland.
- Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2012. Structure building that can't be. In Myriam Uribe-Etxebarria and Vidal Valmala (eds.), *Ways of structure building*, 253-270. Oxford: Oxford University Press.
- Epstein, Samuel D., Hisatsugu Kitahara, and T. 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.
- Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2016. Phase cancellation by external pair-merge of heads. *The Linguistic Review* 33(1): 87-102.
- Epstein, Samuel D., Acriscio Pires, and T. Daniel Seely. 2005. EPP in T: More controversial subjects. *Syntax* 8(1): 65-80.
- Epstein, Samuel D. and T. Daniel Seely. 2002. Rule applications as cycles in a level-free

⁴⁷² Manabu Mizuguchi

syntax. In Samuel D. Epstein and T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*, 65-89. Oxford: Blackwell.

- Epstein, Samuel D. and T. Daniel Seely. 2006. *Derivations in minimalism* Cambridge: Cambridge University Press.
- Erlewine, Michael Yoshitaka. 2017. The anti-locality signature of quirks of subject extraction. Paper presented at Workshop on Quirks of Subject Extraction. National University of Singapore.
- Fortuny, Jordi. 2008. The emergence of order in syntax. Amsterdam: John Benjamins.
- Fox, Danny. 2000. Economy and semantic interpretation. PhD Dissertation. MIT.
- Fukui, Naoki. 1993. A note on improper movement. The Linguistic Review 10(2): 111-126.
- Fukui, Naoki. 2017. Merge in the mind-brain: Essays on theoretical linguistics and the neuroscience of language. New York: Routledge.

Gallego, Ángel J. 2005. Phase sliding. Ms. University of Barcelona.

- Gallego, Ángel J. 2010. Phase theory. Amsterdam: John Benjamins.
- Gallego, Ångel J. 2017. Remark on the EPP in labeling theory: Evidence from Romance. *Syntax* 20(4): 384-399.
- Goto, Nobu. 2017. Eliminating the strong/weak parameter on T. In Michael Yoshitaka Erlewine (ed.), *MIT Working Papers in Linguistics* 85: *Proceedings of GLOW in Asia XI*, Volume 2, 57-71. Cambridge, MA: MIT Working Papers in Linguistics.
- Hazout, Ilan. 2004a. Long-distance agreement and the syntax of *for-to* infinitives. *Linguistic Inquiry* 35(2): 338-343.
- Hazout, Ilan. 2004b. The syntax of existential constructions. Linguistic Inquiry 35(3): 393-430.
- Ishii, Toru. 2004. The phase impenetrability condition, the vacuous movement hypothesis, and *that-t* effects. *Lingua* 114: 183-215.
- Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40(2): 187-237.
- Kuratowski, Casimir. 1921. Sur la notion de l'ordre dans la théorie des ensembles. *Fundamenta Mathematicae* 2: 161-171.
- Laka, Itziar. 1990. Negation in syntax: On the nature of functional categories and projections. PhD Dissertation. MIT.
- Lasnik, Howard. 2001. Subjects, objects, and the EPP. In William D. Davis and Stanley Dubinsky (eds.), Objects and other subjects: Grammatical functions, functional categories and configurationality, 103-121. Dordrecht: Kluwer.
- Lasnik, Howard 2004. The position of the accusative subject in the accusative-infinitive construction. In Peri Bhaskararao and Karumuri Venkata Subbarao (eds.), *Non-nominative subjects*, Volume 1, 269-281. Amsterdam: John Benjamins.
- Legate, Julie A. 2003. Some interface properties of the phase. *Linguistic Inquiry* 34(3): 506-516.
- May, Robert. 1979. Must COMP-to-COMP movement be stipulated? Linguistic Inquiry 10(4):

719-725.

- Mizuguchi, Manabu. 2017. Labelability and interpretability. Studies in Generative Grammar 27(2): 327-365.
- Mizuguchi, Manabu. 2018a. Interaction of Merge and labeling: Consequences for hyperraising. In Michal Temkin Martinez, Gail Shuck, and Tim Thornes (eds.), *Proceedings of Western Conference on Linguistics 2017*, 110-121. Fresno, CA: Department of Linguistics, California State University, Fresno.
- Mizuguchi, Manabu. 2018b. Merge, externalization and subject extraction. Paper presented at Arizona Linguistics Circle 12. University of Arizona.
- Mizuguchi, Manabu. 2019a. Optional raising in ECM and labeling of XP-YP. *Studies in Generative Grammar* 29(2): 373-411.
- Mizuguchi, Manabu. 2019b. Ambiguous labeling and full interpretation. *Studia Linguistica* 73(3): 563-603
- Mizuguchi, Manabu. 2019c. Ways of solving (counter-)cyclic A-movement in phase theory. *Linguistic Research* 36(3): 325-363.
- Obata, Miki and Samuel D. Epstein. 2011. Feature-splitting internal merge: Improper movement, intervention, and the A/A' distinction. *Syntax* 14(2): 122-147.
- Omune, Jun. 2018. Reformulating pair-merge of heads. English Linguistics 34(2): 266-301.
- Ormazabal, Javier. 1995. *The syntax of complementation: On the connection between syntactic structure and selection.* PhD Dissertation. University of Connecticut, Storrs.
- Perlmutter, David. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart and Winston.
- Pesetsky, David. 1982. Path and categories. PhD Dissertation. MIT.
- Pesetsky, David. 1992. Zero syntax, Volume 2: Infinitives. Ms. MIT.
- Pesetsky, David. 2016. Exfoliation: Towards a derivational theory of clause size. Paper presented at Workshop on Shrinking Trees. Universität Leipzig.
- Pesetsky, David. 2019. Exfoliation: Towards a derivational theory of clause size. Ms. MIT.
- Pesetsky, David and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In Michael Kenstowicz (ed.), Ken Hale: A life in language, 355-426. Cambridge, MA: MIT Press.
- Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. Linguistic Inquiry 20(3): 365-424.
- Postal, Paul. 1974. On raising: One rule of English grammar and its theoretical implications. Cambridge, MA: MIT Press.
- Postal, Paul. 1993. Some defective paradigms. Linguistic Inquiry 24(2): 357-364.
- Sauerland, Uli. 2003. Intermediate adjunction with A-movement. *Linguistic Inquiry* 34(2): 308-314.
- Schütze, Carson. 1999. English expletive constructions are not infected. *Linguistic Inquiry* 30(3): 467-484.

- Takahashi, Masahiko. 2010. Case, phases, and nominative/accusative conversion in Japanese. *Journal of East Asian Linguistics* 19: 319-355.
- Tanaka, Hidekazu. 2002. Raising to object out of CP. Linguistic Inquiry 33(4): 637-652.
- Tourlakis, George. 2003. *Lectures in logic and set theory*, Volume 2: *Set theory*. Cambridge: Cambridge University Press.
- van Urk, Coppe. 2015. *A uniform syntax for phrasal movement: A case study of Dinka Bo*r. PhD Dissertation. MIT.
- Wurmbrand, Susi. 2013. QR and selection: Covert evidence for phasehood. In Stefan Keine and Shayne Sloggett (eds.), Proceedings of the Forty-Second Annual Meeting of the North East Linguistic Society, 619-632. Amherst, MA: University of Massachusetts, Graduate Linguistics Student Association (GLSA).

Manabu Mizuguchi

Professor Department of Social Psychology Toyo University 5-28-20 Hakusan Bunkyo-ku Tokyo 112-8606, Japan E-mail: mizuguchi@toyo.jp

Received: 2020. 05. 25. Revised: 2020. 10. 13. Accepted: 2020. 12. 01.