biolinguistics

PsychOpen GOLD

Articles

Check for updates

Constraining Free Merge

Jason Ginsburg [1]

[1] *Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan.*

Biolinguistics, 2024, Vol. 18, Article e14015, https://doi.org/10.5964/bioling.14015

Received: 2024-02-19 • **Accepted:** 2024-09-13 • **Published (VoR):** 2024-12-05

Handling Editor: Kleanthes K. Grohmann, University of Cyprus, Nicosia, Cyprus

Corresponding Author: Jason Ginsburg, Graduate School of Human and Environmental Studies, Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan. E-mail: ginsburg.jasonrobert.2h@kyoto-u.ac.jp

Abstract

Some recent influential work in the Minimalist Program takes the position that Merge, the core language ability to recursively combine two elements together, is free. However, if Merge were completely free, there would be an infinite number of possible derivations for every utterance. Thus, Merge must be constrained in some way. In this paper, I describe a computer model of language that implements a limited form of Merge that is free. I attempt to demonstrate that, within the confines of the language module, Labeling is generally sufficient to constrain Free Merge, and I discuss issues that arise regarding overgeneration of syntactic structures given Free Merge.

Keywords

Merge, Labeling, Box Theory, FormCopy, computer modeling

1 Introduction

If recent work in linguistics is correct, then Merge, the process of combining together linguistic objects, is a core property of language that is utilized by the language faculty to construct syntactic objects (SOs). Chomsky (2010, p. 52) writes that "unbounded Merge is the sole recursive operation within UG" and that it is "part of the genetic component of the language faculty." If this is correct, human language makes use of recursive Merge. Berwick (2011) suggests that non-human primates have lexical items but no Merge, whereas birds have something like Merge (used in songs) but no lexical items. Human language, crucially, makes use of lexical items and Merge.

Chomsky (2001, 2013, 2015) takes the position that Merge is free. Chomsky (2015, p. 14) writes that "[t]he simplest conclusion … would be that Merge applies freely" and

"[o]perations can be free, with the outcome evaluated at the phase level for transfer and interpretation at the interfaces." I take this to mean that both External Merge and Internal Merge are free. Crucially, Free (Internal) Merge would result in an infinite number of possible structures generated for every possible utterance. This is untenable. Thus, Free Merge must be constrained by the language faculty.

The question then arises of how Free Merge is constrained. In this paper, I demonstrate how, given a Merge-based model of language generation (based on recent work in linguistic theory), Merge can be constrained. Crucially, I argue that arguments can be subject to Free Merge, subject to the constraints of the language module, but that Labeling in general is sufficient to eliminate most impossible derivations.

In the following sections, I discuss my core assumptions regarding syntactic structure, which I implemented in a computer model that automatically generates sentences. Notably, in this model, I attempt to remove many of the basic problematic and overly complex assumptions in recent work in the Minimalist Program (Chomsky, 1995) with the goal of keeping language simple, in accord with the Strong Minimalist Thesis (Chomsky, 2000, 2001, 2010; Chomsky et al., 2023), the notion that "language keeps to the simplest recursive operation, Merge, and is perfectly designed to satisfy interface conditions (Chomsky, 2010, p. 52)." Then I explain how Labeling is generally sufficient to constrain Free Merge. This is followed by discussion of issues that arise with respect to overgeneration.

2 Computer Model of Language

For this work, I created a computer model that implements the theory that is presented in this paper. This model was created in the Python programming language, and the output is generated with HTML and JavaScript. This model is fed an input stream of lexical items, which it Merges together to form SOs. Selection and Merge of a lexical item from the input stream is External Merge. The model also implements Internal Merge (displacement) of elements from within an SO. This Internal Merge (IM) is the main focus of this paper. The model can compute multiple derivations for a single input stream, which is crucial for implementing a version of Free Merge. This model is a language generator (not a parser) because it generates phrases and sentences from a given input list of lexical items; it is not fed complete sentences as input.[1]

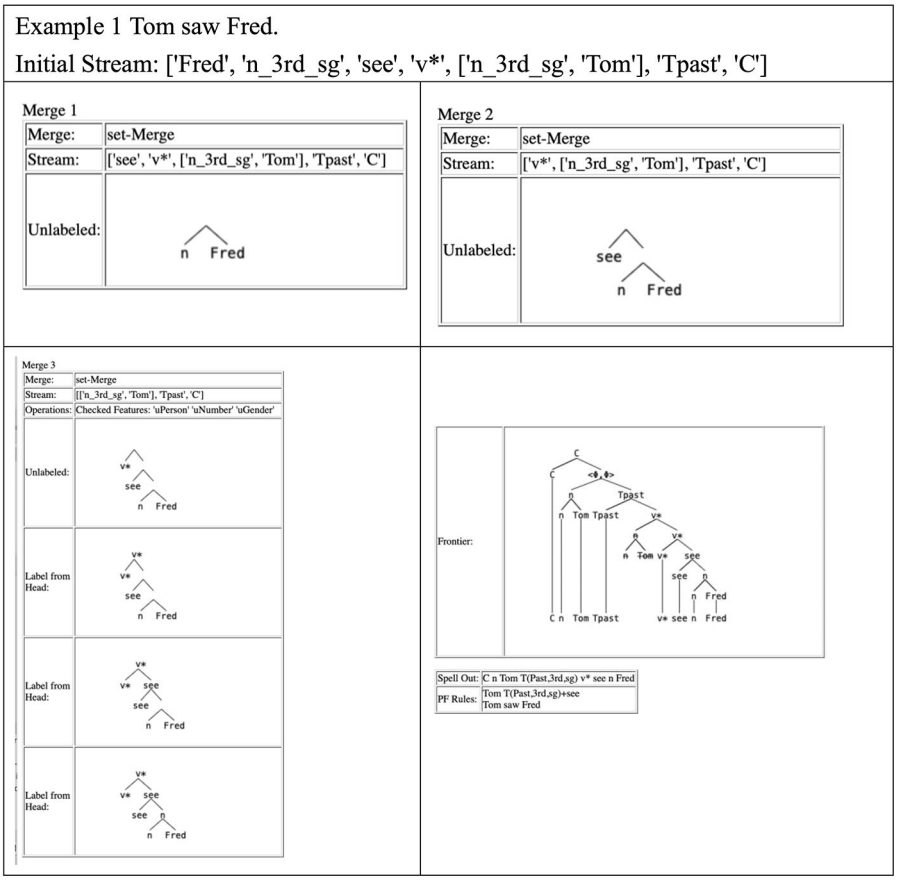
Portions of a derivation produced by the model are shown in Figure 1. An initial list of lexical items is fed into the model. The model consecutively selects and Merges together the lexical items, in accord with the theory that is developed in this paper. After

1) This computer model is simply an attempt to model syntactic theory in accord with recent work in the Minimalist Program. This model, as well as related models presented in Ginsburg (2016), Fong and Ginsburg (2019) and Ginsburg and Fong (2019), have no relation to the Minimalist Grammar formalisms of Stabler (1997, 2011) and related work.

each Merge step, the model checks for the possibility of agreement relations and for the possibility of Labeling (see Section 3). When a derivation is complete, it is transferred to Spell-Out, where a particular pronunciation is determined. For any particular example, there can be multiple successful derivations (that converge), as well as multiple crashed derivations.

Figure 1

Main Components of the Computer Model



In this paper, I utilized this computer model to test the theory that is developed in the following sections. The model produces complete step-by-step derivations for all target constructions, thus making it possible to find problems with, and verify the accuracy of, the target theory. The complete derivations for all target constructions presented in this

paper can be found in the Supplementary Appendix (see Ginsburg, 2024),[2] and these can be of use to researchers who are interested in verifying the proposals in this paper. The main focus of this paper is on linguistic theory. I used this model to test the accuracy of the theory that I develop in this paper.

3 Basic Assumptions About Language

In this section, I review the basic Labeling-based proposals of Chomsky (2013, 2015) and then I describe the basic properties of the language faculty that I assume to be at work in the language model that I created.

3.1 Labeling-Based Derivations According to Chomsky (2013, 2015)

Following Chomsky (2013, 2015), I assume that a form of Labeling is at work with respect to language generation. Labeling is necessary for interpreting phrases. Labeling refers to a process of finding a prominent feature of an SO via the search process involved in language, Minimal Search (Chomsky 2013, 2015).[3] Chomsky (2013, p. 43) writes that "[t]he simplest assumption is that LA [Labeling Algorithm] is just minimal search, presumably appropriating a third factor principle, as in Agree and other operations." In this way, Labeling is really just a form of Minimal Search, which finds prominent features that can function as labels.

Labeling via Minimal Search works as follows. Assume that a head X and a phrase YP Merge, forming {X, YP}. In this case, the label is X, assuming that X has prominent features that are capable of Labeling. If an XP and a YP, both phrases, Merge to form {XP, YP}, then shared features can label. For example, assume that XP (specifically the head X) has phi-features and YP (the head Y) has unchecked phi-features. Minimal Search results in the XP and YP forming an Agree relation so that the uPhi on YP are checked by the iPhi (interpretable phi-features) on X, and then the shared phi-features on XP and YP can label. Chomsky also takes the position that the English T is too weak to label on its own —this accounts for the requirement that a clause have a subject (the traditional EPP effect of Chomsky (1981)). Given the structure {T, YP}, T alone cannot label. However, given {XP, TP} where TP and XP Agree in terms of phi-features, the shared phi-features label. This is accounted for as follows. T inherits uPhi from C. Given an {XP, TP} structure, the uPhi on T Agree with the phi-features on X and the shared phi-features label. Crucially, in an {XP, YP} structure in which XP and YP do not Agree, Labeling is not possible.

2) More details about how the model works are given in the Appendix (see Ginsburg, 2024). An in-depth description about how the model works is beyond the scope of this paper.

3) An anonymous reviewer writes that in Chomsky's approach, "there is no labelling notion at all, and all there is the necessity for Minimal Search to univocally find a feature for each SO."

Also, given a {T, YP} structure, Labeling is not possible, assuming that T is too weak to label. These are summarized in (1). Furthermore, consider a root that has Merged with a functional head (e.g., a categorizer) to form what is essentially a head-head structure; for example, the root *walk* Merges with a categorizer n. In this case, the root is too weak to label by itself, but the functional head can label. This is the position that Chomsky (2013, p. 47) takes, following Marantz (1997), Embick and Marantz (2008), and Borer (2005a, 2005b, 2013). I assume that a root can label after it Merges with a categorizer.

(1) Labeling Failure

a. {XP, YP} – X and Y do not Agree
b. {T, YP} – T is too weak to label

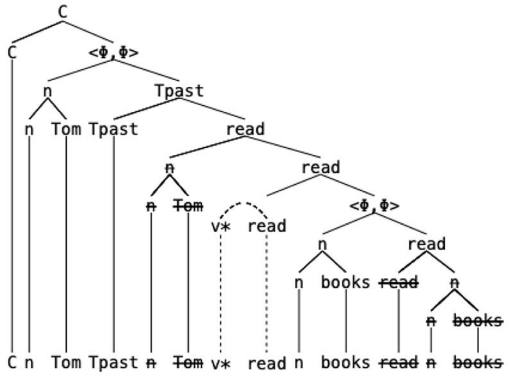
The convergent derivation of *Tom read books*,[4] following Chomsky (2013, 2015), proceeds as shown in Figure 2. The lines below each terminal node represent the frontier of the derivation—the portion of the derivation that is sent to Spell-Out to be pronounced. The root *books* set-Merges with the functional categorizer n, and n labels, as the root *books* is not capable of Labeling. Chomsky (2015) claims that a verbal root undergoes internal pair-Merge (head movement) with v to form <v*, read>, resulting in v* being dephased (also see Epstein et al., 2016). The <v*, read> pair-Merged structure is represented with a dotted arc. Dephasing is a process in which an element that would typically function as a phase head, thus being a point of transfer, no longer functions as phase head.[5] In this case, Chomsky proposes that phasehood is passed onto the complement of v*. Thus, the complement of v* will function as a phase and be transferred. A phase head passes uFs to its complement, so the uPhi (uninterpretable Phi) of v* are passed onto the verbal root *read*, which being a root, is unable to label by itself. The object *a book* undergoes IM (Internal Merge) with *read* to form an {XP, YP} structure. In the matrix clause, the subject is initially-Merged with vP, and then it internally-Merges with the TP. The uPhi of C are inherited by T. Minimal Search results in phi-feature agreement in the {NP, {Tpast…}} and {NP, {read…}} structures, and these shared phi-features are able to label, where the label is indicated as <ϕ, ϕ>.

4) Chomsky's original example is *Tom read a book*.

5) Dephasing is a complex process that involves somehow passing phasehood from one head to another. I will not utilize phasehood in my model.

Figure 2

Structure of "Tom Read Books"



Note. Adapted from Chomsky (2015, p. 10).

In the following subsections, I explain my assumptions about Labeling Theory. Note that, for reasons discussed in the following sections, I do away with some of the operations utilized in the type of derivation shown in Figure 2.

3.2 Phases

Labeling Theory follows the view that the structures of sentences are constructed hierarchically in a bottom-up fashion, and sentences consist of phases, which are portions of sentences that essentially become inaccessible after construction. The core phases are generally assumed to be a transitive Verb Phrase (v*P) and a Complementizer Phrase (CP), following Chomsky (2000, 2001). Both (2)a–b are well-formed and both crucially are formed from the same set of lexical items. These examples differ, however, with respect to the ordering of lexical items. The embedded CP in (2)a is a phase that is constructed from a lexical array that does not contain *there*. As a result, *a man* raises to subject position of the CP. The expletive *there* is associated with the higher phase of the matrix clause. In (2)b, on the other hand, the expletive *there* is available in the embedded CP phase. As a result, *there* is inserted in subject position of the CP and *a man* does not need to move.

(2) a. **There** is a possibility [$_{CP}$ that **a man** will be t in the room].
b. A possibility is [$_{CP}$ that **there** will be **a man** in the room]. (Epstein, Kitahara, & Seely, 2014, p. 469)

Once a phase is complete, the complement of the phase head becomes inaccessible to further operations, which is proposed to reduce memory burden—the mind can essentially put a completed phase to the side and compute the next phase. Note that when a phase head is Merged, there are differing views about which portions of the phase become inaccessible in accord with the Phase Impenetrability Condition (Chomsky, 2000, 2001; Müller, 2004; Richards, 2011). Under one version of the Phase Impenetrability Condition, the complement of the phase head becomes inaccessible and is transferred, but in another version, the complement (if present) of the lower phase head becomes inaccessible and is transferred. As noted by Boeckx and Grohmann (2007, p. 206),[6] referring to Chomsky (2000), "[c]omputation cost reduction is the prime conceptual advantage and motivation for phases." This means that all feature checking operations within the phase must be complete, and any elements that need to move out of the phase must have moved to the edge of the phase before completion. Since phases are thought to be complete (in some sense), they ideally should be of some advantage when accounting for island effects, although whether or not this is the case is open to debate (e.g., see Chomsky, 2008; Gallego, 2010). I incorporate the notion of phases into this model, since they are utilized in Labeling theory. I assume that the phases, following Chomsky (2001), are transitive VP (v*P) and CP. Note that, essentially following Chomsky (2021), I will assume that when v* or C is Merged, the v*P/CP is transferred. Thus, the head of the phase is transferred together with its complement but the specifier, if present, remains outside of the transferred phase.

3.3 Feature Inheritance and Agreement

Feature inheritance is an operation in which a phase head passes features onto a complement. The notion of feature inheritance was proposed by Chomsky (2008), based on work (to the best of my knowledge) by Carstens (2003) and Miyagawa (2005), among others (also see references in Carstens, 2003, and Miyagawa, 2005). Chomsky (2008, pp. 143–144) writes:

> ….for T, ϕ-features and Tense appear to be derivative, not inherent: basic tense and also tenselike properties (e.g., irrealis) are determined by C … or by the selecting V (also inherent)…In the lexicon, T lacks these features. T manifests the basic tense features if and only if it is selected by C…if not, it is a raising (or ECM) infinitival, lacking ϕ-features and basic tense. So it makes sense to assume that Agree and Tense features are inherited from C, the phase head.

6) See Boeckx and Grohmann (2007) for discussion of problems with the notion of phases.

Feature inheritance can be useful for accounting for Exceptional Case Marking (ECM) constructions. In the ECM (3)a, the embedded T, pronounced as *to*, occurs without C. In this case, T lacks agreement features and *him* Agrees with the matrix verb *expect*. In (3)b, on the other hand, T, pronounced as past tense on the verb *win* (resulting in *won*), occurs with C. T has agreement features and Agrees with the subject, resulting in the nominative pronoun *he*. These types of simple examples demonstrate how T, in the presence of C, has agreement features, which it lacks in the absence of C. While feature inheritance is useful for accounting for the ECM data, it isn't necessarily clear if it is required. If non-finite T simply lacks a full set of unchecked/uninterpretable phi-features (uPhi), and tensed T has a full set of uPhi, the same facts can be accounted for, without recourse to feature inheritance.

(3) a. I expect [$_{T}$ him to win].
b. I think [$_{C}$ that he won].

Feature inheritance notably is a complex operation that involves copying agreement features from C onto T, or the passing of features from C to T. Chomsky notes that this violates the No-Tampering Condition (Chomsky, 2000, pp. 136–137; Chomsky, 2008, p. 138), as it requires altering an already formed syntactic structure. The question then arises of whether or not it is conceptually necessary.

Complementizer agreement is found in a variety of languages such as Frisian, some Dutch and Germanic dialects, and Bantu languages (Koppen, 2017). Note that both C and a verb can show agreement with a subject, as in (4)a–b, in which the complementizer and the verb show agreement with the subject. Assuming that verbal agreement indicates agreement on T, then both C and the verb Agree with the subject in these examples.

(4) a. **datt-e** wiej noar 't park loop-t (Dutch, Hellendoorn dialect)
that-pl we to the park walk-pl
'that we are walking to the park' (Ackema & Neeleman, 2001, p. 34; Carstens, 2003, p. 397)
b. **dan ik** werken (West Flemish)
that-1sg I work-1sg (Ackema & Neeleman, 2001, p. 29)

Although the existence of complementizer agreement as in (4) has been given as evidence for feature inheritance (Chomsky, 2008; Miyagawa, 2005), complementizer agreement tends to be less common and less complete than agreement with T. Matasović (2018, p. 9) writes "that the most common agreement pattern within the domain of the clause is verbal agreement." Koppen (2017, p. 7) writes "The CA [Complementizer Agreement] paradigm is usually defective, however, in the sense that not all person/number combinations of the subject lead to an overt agreement reflex on the complementizer." Koppen (2005, p. 35) points out this defectivity in a variety of Germanic/Dutch languag-

es/dialects. In Frisian, a complementizer shows agreement only with a second person singular embedded subject, whereas a verb shows agreement with all types of subjects, as shown in Table 1. Koppen (2005) discusses similar paradigms in Tegelen Dutch, Bavarian, and Lapscheure Dutch. In all of these languages/dialects, there are variations with respect to the extent of complementizer agreement, but there are fewer complementizer agreement suffixes than verbal suffixes, thus providing further evidence for the notion that complementizer agreement tends to be defective.

Table 1

Agreement in Frisian

Person.Number	Comp. agreement	Verbal agreement
1 Per.Sg	-0	-n
2 Per.Sg	**-st**	-st
3 Per.Sg	-0	-t
1 Per.Pl	-0	-e
2 Per.Pl	-0	-e
3 Per.Pl	-0	-e

Note. Koppen (2005, p. 35).

If it truly is the case that verbal agreement is more common than complementizer agreement and that verbal agreement tends to be more complete than complementizer agreement, then this may be an indication that a complementizer is not always the origin of agreement features. If C were the locus of agreement features, then one might expect agreement with C to be more common than it is, and for agreement with C to tend to be more, not less, complete than agreement with T.

Feature inheritance also raises technical problems. If uPhi are inherited by T, one possibility is that all of the uPhi of C are passed from C onto T and no longer remain on C. This would be the case when agreement only shows up on T (usually visible on the verb). This does not appear to be the case in the examples in (4) in which there is agreement between a subject and both C and the verb (assuming the verbal agreement is the result of agreement on T). Another possibility is that the uPhi of C are copied onto T, so that they appear on both C and T. This could account for the data in (4). Again, copying of features from one element onto another seriously alters an already formed SO, again violating the No-Tampering Condition. It would be simpler if C and T come with their necessary agreement features.

Richards (2007) provides arguments for feature-inheritance, proposing that feature transmission is a "conceptual necessity" in order to avoid transfer of uninterpretable features. Uninterpretable features, by definition, cannot be processed by the semantic component of a derivation. If uninterpretable features are checked but are not transferred

immediately, then they should stay around and cause a derivation to crash, according to Richards, since they can't be interpreted. Thus, uninterpretable features must be transferred as soon as they are checked. The assumption seems to be that when checked, uninterpretable features are transferred with the phase. They are "deleted" so that they are no longer visible to the semantic component. Assume that T has uPhi that are checked via Agree with a subject, before the phase head C is Merged. When these features are checked, they cannot be transferred until after the phase head C is Merged. Thus, these uPhi cannot be deleted as soon as they are checked. These checked uPhi, according to Richards, then become indistinguishable from interpretable features and interpretable phi-features on T, presumably, will cause a derivation to crash. The idea seems to be that since T is not an argument, phi-features (which are associated with arguments) cannot be interpreted on T. On the other hand, if uPhi are inherited from C by T, then as soon as they are inherited, they are checked, and since the phase level has been reached, the checked uPhi are instantly transferred, so that they no longer remain for the semantic component. Assuming that uninterpretable features originate on a phase head predicts phase-level operations of inheritance of features, Agree (e.g., checking of uPhi on T by phi-features on a subject), and transfer of the relevant portion of the phase.

However, the idea that uninterpretable features need to be deleted as soon as they are checked is not necessarily a given. Since these features are uninterpretable, by definition, they could cause a derivation to crash if they are transferred, but as long as they are deleted before transfer, it isn't clear why they need to be deleted immediately – this seems to be a stipulation. Furthermore, some recent work takes the position that the complement of a phase head is not transferred immediately. Chomsky (2015) argues that phasehood can be transferred to the complement of a phase head, based on ECM constructions and the *that*-trace effect.[7] As noted by Goto (2017), the motivation for feature-inheritance based on the need to delete uninterpretable features as soon as they are checked may not necessarily hold.

Another issue with feature inheritance involves probe-goal agreement. Since Chomsky (2001), agreement has typically been assumed to involve a probe-goal relation. For example, assume that T in (5) has uPhi that probe for and Agree with the phi-features on a subject. Similarly, v* has uPhi that probe for and Agree with phi-features on an object. The relations Agree($T_{[\cancel{uPhi}]}$, $Mary_{[iPhi]}$) and Agree($v^*_{[\cancel{uPhi}]}$, $books_{[iPhi]}$) check the uPhi on T and Mary via probe-goal agreement.

(5) $[T_{[\cancel{uPhi}]}\ Mary_{[iPhi]}\ v^*_{[\cancel{uPhi}]}\ \text{bought}\ books_{[iPhi]}]$

7) Note that the complex operation of phasehood transfer is also suspect from the perspective of Minimalism. I do not adopt this operation in my model.

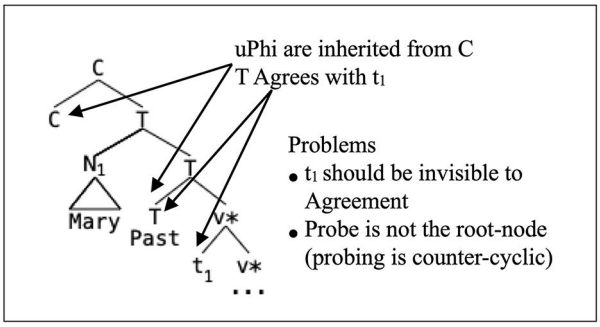
Now consider how probe-goal agreement of this sort works given feature-inheritance. If T must inherit its uPhi from C, then the uPhi on T cannot probe until after C is Merged. Thus, probing is counter-cyclic, not from a root node, which is contrary to the original notion of probe-goal in which probing occurs from the root node (Richards, 2006).

Another problem, pointed out by Epstein, Kitahara, and Seely (2022), hereafter EKS (2022), is that given feature inheritance, there are cases in which agreement must occur with a goal that is no longer visible. Assuming feature inheritance, in (6), the uPhi features on T are inherited from C. Thus, T does not obtain its uPhi features until after C is Merged, and also after the subject has internally Merged with the TP (assuming that a subject raises to the specifier of TP). Then, following Chomsky's (2013) view that only the highest copy of a syntactic object (SO) is visible to probing, the lower copy of the subject is not visible to probing. This means that the probe cannot find the subject in its base position. The higher copy of the subject is in the specifier of the TP, so that the past tense T does not c-command it. See Figure 3. EKS (2022) propose a solution based on Minimal Search (Agreement occurs between T and the subject in the TP). However, none of this is necessary if there is no feature inheritance. If T simply comes with its relevant set of uPhi features, then it can probe as soon as it is Merged. There is no need for counter-cyclic Agree relations, and the problem of agreement with an invisible copy of an SO does not arise.

(6) $[C\ Mary_{[iPhi]}\ [_T\ T_{[\sout{uPhi}]}\ Mary_{[iPhi]}\ v^*_{[\sout{uPhi}]}\ bought\ books_{[iPhi]}]]$

Figure 3

Agreement Given Feature Inheritance



The facts regarding feature inheritance are far from settled, but I will assume that from the perspective of the Strong Minimalist Thesis, it is best to do without it.[8] Feature inheritance is best eliminated from the current theory from the perspective of simplicity;

it is a complex operation, and a complex operation such as feature inheritance requires extraordinary justification.

3.4 Agreement and Case

Case is subject to a great deal of cross-linguistic and language-internal variation. As is well known, Case morphology in English is basically "phonologically zero" (Pesetsky & Torrego, 2011, p. 55), except for pronouns. Case shows up on nouns in Latin, as in (7). In Russian, there are declinable and indeclinable nouns (Pesetsky & Torrego, 2011, p. 55), so that whether or not Case appears overtly on a noun can depend on the particular noun, as in (8). In Icelandic, the verb *luku* 'finished' occurs with a dative object and the verb *vitjuðum* 'visited' occurs with a genitive object, as in (9)a–b. Icelandic is also well-known for constructions in which a subject appears with dative Case and an object with nominative Case, as in (9)c. Furthermore, Bobaljik (2008) points out that nominative-accusative case systems and ergative case systems (which typically mark the subject of an intransitive verb and the object of a transitive verb with ergative case) assign Case differently, but arguments seems to be treated syntactically in the same way in both types of systems, suggesting that Case is not truly a syntactic relation.

(7) libr-um (Latin)
book-Acc (Pesetsky & Torrego, 2011, p. 53)

(8) a. mašin-u b. mašin-y c. mašin-oj (Russian)
car-Acc car-Gen car-Instr
d. kenguru
kangaroo-Acc/Gen/Instr (Pesetsky & Torrego, 2011, p. 55)

(9) a. Đeir luku kirkjunni (Icelandic)
They finished the.church.Dat
b. Við vitjuðum Olafs.
We visited Olaf.Gen (Pesetsky & Torrego, 2011, p. 61)
c. Jóni líkuðu þessir sokkar (Icelandic)
Jon.Dat like.pl these socks.Nom
'Jon likes these socks.' (Jónsson, 1996, p. 143; per Bobaljik, 2008, p. 298)

8) Goto (2017) and Gallego (2017) both attempt to do away with feature inheritance. These works, however, make use of head movement, an operation that is potentially problematic. If there is no head movement of the sort that they propose, then these approaches potentially do not work. See Section 3.5 for discussion of head movement.

These Case facts can be accounted for if Case is primarily a Spell-Out phenomenon. Marantz (2000, p. 20) argues that "case and agreement morphemes are inserted only after SS [Sentence Structure] at a level we could call "MS" or morphological structure." Bobaljik (2008), following Marantz, writes that "the proper place of the rules of m-case [morphological-case] assignment is thus the Morphological complement, a part of the PF interpretation of syntactic structure (Bobaljik, 2008, p. 300)." Chomsky (2021, p. 23) suggests that "Case is part of externalization" further writing that "there seems to be no general semantic reason" for Case systems and "[p]erhaps establishing relations among elements facilitates perception/parsing."

In my model, I take a Spell-Out-based approach to Case. I assume that Case appears at Spell-Out, following Chomsky's view that Case is a reflex of phi-feature agreement (Chomsky, 2000, 2001). This approach can account, at least to a certain extent, for some of the language-internal and cross-linguistic idiosyncrasies that occur with Case. I assume that unchecked phi-features, uPhi, must be checked for a derivation to converge. The result of phi-feature agreement can lead to an argument being pronounced with overt Case morphology. Case, however, is a Spell-Out phenomenon. The exact form of Case can be subject to language internal and cross-linguistic variation, but the actual form of Case on an argument does not have an influence on syntax. Note that if an argument is unable to be pronounced with Case, a derivation can crash at Spell-Out (see Section 3.6).

3.5 Head Movement

Head movement is a controversial topic. It appears to be ubiquitous. However, it isn't clear how exactly it works. Consider the basic examples in (10) and (11) which show typical head-movement of T to C. Assume that C is selected and Merged with TP. Then assume that T raises and undergoes IM (undergoes head-movement) with C, as shown in (10)b and in (11)b (assume that *will* is in T). These head-movement operations appear to violate the No Tampering Condition because an already formed CP is altered. Head movement also violates the Extension Condition (Chomsky, 1993, 1995) which is the requirement that operations extend the SO. Movement of a head to a position within the clause does not extend the size of the SO.

(10) a. C Mary T bought the book.
b. C+T Mary ~~T~~ buy the book? → Did Mary buy the book?

(11) a. C Mary will buy the book.
b. C+will Mary ~~will~~ buy the book? → Will Mary buy the book?

Chomsky has proposed differing views on head movement. Chomsky (1993, p. 23) argues that the Extension Condition does not apply to adjunction operations, which means that it does not apply to head movement, assuming that head movement is adjunction (also

see Dékány, 2018, p. 5). However, making head-movement an exception to the Extension Condition is not exactly ideal. From the perspective of Minimalism, the notion that all Merge operations target the root of an SO is optimal. Thus, head-movement seems to violate this requirement, and this alone is enough to make head-movement suspect.

In recent work in Labeling Theory, a version which I assume here, Chomsky (2015) makes use of head movement. Chomsky (2015, p. 12) proposes that a verbal root R raises to v*, resulting in dephasing of v* (see Figure 2 above). This raising operation forms "the amalgam [R-v*]", which Epstein, Kitahara, and Seely (2016) (hereafter EKS, 2016) specifically describe as being a case of internal pair-Merge. EKS (2016, p. 90) write "pair-Merge internally forms <R, v*> (=R with v* affixed)." Internal pair-Merge, unless it is implemented via sidewards movement, requires R to internally pair-Merge with v* after v* has Merged with the SO. This head-movement again violates the No Tampering Condition and the Extension Condition.

Despite making use of head movement in some works such as those discussed above, a variety of issues, including the violation of the Extension Condition/counter-cyclicity, lead Chomsky (2001) to propose that head-movement is a phonological (PF) operation. Chomsky (2001, p. 37) writes "[t]here are some reasons to suspect that a substantial core of head-raising processes, excluding incorporation in the sense of Baker (1988), may fall within the phonological component." Some other reasons from Chomsky (2001, pp. 37–38) why head movement is problematic are as follows (see Roberts, 2011, for a summary of these arguments). There are no clear interpretation differences in languages such as French and Icelandic, in which the verb appears in a position generally considered to be T (possibly a result of head movement), compared with languages such as English, in which the verb remains below T. If head movement were responsible for verb movement, and if head movement influences interpretation, then the expectation is that semantic differences would arise (Roberts, 2011, p. 99). Another issue is that if a head raises and adjoins to a higher head within an SO, then "the raised head does not c-command its trace (Chomsky, 2001, p. 38)." Also, a phrase can undergo successive-cyclic movement whereby it moves from the specifier of one phrase to the specifier of another phrase, but this doesn't appear to occur with a head. Rather, "it always involves 'roll-up' (i.e. movement of the entire derived constituent … iterated head movement always forms a successively more complex head (Roberts, 2011, p. 201)." For example, in an English interrogative, an auxiliary (Aux) moves to T (assuming that the auxiliary is not base generated in T), and then Aux-T raises to C. An auxiliary cannot move to C and leave T behind.

There are a variety of proposals related to head movement in the literature which make use of syntactic movement and/or post-syntactic PF operations. Embick and Noyer (2001) propose that there are postsyntactic lowering operations in which a head can lower and combine with another head. Matushansky (2006) argues that typical cases of head movement can result from a syntactic operation of movement of a head to a

specifier position followed by a morphological (non-syntactic) operation that creates adjoined heads. Harizanov and Gribanova (2019) argue that some cases of head movement occur in the syntax via IM and other cases involve a morphological process that occurs outside of the syntax via an operation of "postsyntactic amalgamation" which can involve postsyntactic lowering or raising of a head to form a head-head adjunction structure. Harley (2004) argues for a non-syntactic operation of conflation (following Hale & Keyser, 2002) in which the defective phonological features of one head combine with the phonological features of a complement. Platzack (2013) also develops a purely phonological approach to head movement which permits pronunciation of heads to occur in a position that is different from in the syntax. See Roberts (2011) and Dékány (2018) for in-depth discussion of problems with head movement, as well as discussion of potential ways to account for head-movement, both in narrow syntax and at PF. Also see Roberts (2010) for an attempt to account for some components of head-movement in the syntax.

Notably, there are two types of proposals in the literature involving head-movement that occurs in the syntax without violating the Extension Condition.[9]

In some accounts (e.g., Harizanov & Gribanova, 2019; Matushansky, 2006; among others), a head can undergo IM to a specifier position. Thus, a head essentially functions as a specifier. The head moves to the root of the SO, so this movement is not counter-cyclic and does not violate the Extension Condition. However, the result of movement is an $\{_{Y}$ X, {YP…X}} structure in which the X does not label, so X functions as a specifier. This is problematic given the core notions of Labeling Theory, which assume that given an {X, YP} structure, the head X labels.

Another approach is consistent with Labeling Theory. In this approach, a head undergoes IM with the core SO, and then the head relabels. Thus, IM of X with YP, forms an $\{_{X}$ X, {YP…X}} structure in which X functions as the label. Presumably, X also functions as a label in its base position, and thus, this is sometimes referred to as "relabeling" as well as "reprojection". For example, this type of head movement might be plausible in relative clauses, if one assumes that a nominal head raises and relabels, and if relabeling by a simplex head does not go against the core assumptions of Labeling Theory. This type of relabeling approach can be found in Georgi and Müller (2010), Donati and Cecchetto (2011), Cecchetto and Donati (2015), and Fong and Ginsburg (2023), among others. Note that this type of head movement might be compatible with my computer model, although it is not utilized for the derivations discussed in this paper (as it is not necessary).[10]

9) See Harizanov and Gribanova (2019, p. 493) for a more extensive list of literature that discusses these two possible types of head movement.

10) In my model, Free Merge is limited to arguments (see Section 4). Thus, the notion that a head is able to undergo re-Merge, even in limited cases, might be problematic.

Although the issues regarding head-movement are far from settled, I adopt a model in which there is no counter-cyclic head movement in the narrow syntax. I take an affix hopping approach (Chomsky, 1957) in which a set of Spell-Out rules are applied to the output of the syntax.[11] Crucially, if X (an affix) and Y are adjacent at PF, then it is possible for Y to be linearized before X.

The basic rules that I implemented are given in (12) below. Note that only a small set of rules is sufficient to account for the basic English constructions produced by my model. The computer model generates a syntactic structure. If a derivation does not crash, the nodes of the tree are sent to Spell-Out. Then the basic PF (Phonological Form) rules apply when necessary. Examples of rule applications for particular Spell-Out forms are shown in Table 2. Note that for these rules to apply, the PF component of the derivation has to have access to some syntactic category information. Thus, if the model finds T adjacent to v* which is adjacent to a root R, or an auxiliary, then T attaches onto the root or auxiliary, and v* is eliminated from PF, since it has no pronunciation. As shown in Table 2, for *Tom saw Fred*, the adjacent SOs *T(Past,3rd,sg) v* see* are converted into *T(Past,3rd,sg)+see* which is pronounced as *saw*, where T ends up suffixing onto the adjacent *see*; with a regular verb, past tense is generally pronounced as *-ed*. The verbal head v* is not pronounced. For *Will Tom read a book*, the interrogative C and T combine to form C_Q+T(Pres,3rd,sg). Note that this requires T to attach onto C_Q by moving over the subject at Spell-Out. Furthermore, the auxiliary *will* must move over the subject and combine with T to form *T(Pres,3rd,sg)+will*, which ends up being pronounced as *will*. In cases in which T combines with C, but there is no overt element in T, then the appropriate form of *do* (depending on Tense and agreement) is pronounced. The appropriate forms of *do* as well as irregular verb forms are listed in a lexicon, which the model consults.[12]

(12) Basic PF (Phonological Form) Rules (R = Root, Aux = Auxiliary):
Tense: T v* R/Aux⟶ R/Aux+T
Modal: T Modal⟶ T+Modal
Passive: be -en v R ⟶ be R+-en
Progressive: be -ing v* R ⟶ be R+-ing
Perfective: have -en v* R⟶ have R+-en
Interrogatives: C_Q N T ⟶ C_Q+T N
Irregular verb forms are stored in lexicon.

11) This type of approach might be compatible with more complex approaches such as that of Harley (2004).

12) The lexicon contains a list of irregular verbs. When the model encounters a verb at Spell-Out, it checks the lexicon. If the verb is not listed in the lexicon, then regular tense rules apply. For example, the past tense is formed by adding *-ed*. If the verb is in the lexicon, then the appropriate form of the verb is selected from the lexicon.

Table 2

Spell-Out for Basic Sentences

Spell-Out	PF Rules
C n Tom T(Past,3rd,sg) v* see n Fred	Tom T(Past,3rd,sg)+see
	Tom saw Fred
C n Tom T(Pres,3rd,sg) will v* read a n book	Tom T(Pres,3rd,sg)+will
	Tom will read a book
C the n book T(Past,3rd,sg) be -en v read	the book T(Past,3rd,sg)+be
	the book was -en+read
	the book was read
C n she T(Past,3rd,sg) be -ing v* read a n book	she T(Past,3rd,sg)+be
	she was -ing+read
	she was reading a book
C n she T(Past,3rd,sg) have -en v* read a n book	she T(Past,3rd,sg)+have
	she had -en+read
	she had read a book
C_{φ} n Tom T(Pres,3rd,sg) will v* read a n book	C_{φ}+T(Pres,3rd,sg)
	T(Pres,3rd,sg)+will
	will Tom read a book

3.6 FormCopy

Chomsky (2021) proposes a FormCopy operation which accounts for how nominals are interpreted as Copies. Chomsky (2021) describes FormCopy as a rule that assigns "the relation *Copy* to certain identical inscriptions" (p. 17) that are in a c-command relation, and that presumably are in the same phase. Identical inscriptions refers to arguments that are identical in form. Consider how this accounts for the Control construction (Chomsky, 1981, 1986) in (13)a, as shown in (13)b. Here, I assume that the Control construction is a TP.[13] The NP *many people*$_1$ is externally-Merged in the vP theta-position and it undergoes IM to the non-finite TP specifier position. The NP *many people*$_2$ is separately externally-Merged in the matrix vP theta-position, and internally Merged with the matrix T. FormCopy applies and all inscriptions of *many people*, except for the highest *many people*$_2$ in the matrix TP, are interpreted as Copies, which are not pronounced.

13) It is crucial that this Control clause (the embedded non-finite clause) be treated as a TP. If it were a CP, then when the C phase head is Merged, it would be transferred, thus making it inaccessible to FormCopy applications from the matrix clause. Chomsky (2021) treats a Control clause as a TP. However, the status of a Control clause is not entirely straightforward. It has been analyzed as a CP as well as a TP. See Radford (2016, Chapter 4), and references therein, as well as Landau (2024) for arguments that a Control clause is a CP with a null infinitival complementizer. Note that if a non-finite clause contains a C that for some reason is not treated as being a phase head, then the current FormCopy account could be maintained.

If FormCopy were not to apply in (13), the lower *many people*$_1$ would be treated as a separate NP (a repetition, not a Copy) from the higher *many people*$_2$. This is fine, as it has a separate theta-role from the matrix *many people*. However, the construction would then crash. Why exactly it would crash is then an issue. I assume that it would crash at Spell-Out due to Case reasons, not due to issues with the syntax. The relevant argument *many people*$_1$, not being a Copy, would need to be pronounced, but it would not be in a position in which it could obtain Case, in the edge of a non-finite TP, and the matrix *try* is not the kind of verb that assigns accusative Case.

(13) a. Many people tried [many people to win].
b. [Many people]$_2$ T ~~[many people]$_2$~~ tried [~~many people~~]$_1$ to [~~many people~~]$_1$ win. (Adapted from Chomsky, 2021, p. 22)

Note that FormCopy has the advantage of eliminating the need for memory of movement operations. In (14), assuming the standard view of the VP-internal subject hypothesis, the subject *John* is externally-Merged in the v*P. Then it internally-Merges with T in subject position. With FormCopy, there is no need to retain memory of the movement of *John*. When the construction is generated, *John* undergoes IM. Then, at the phase level, FormCopy applies and the lower inscription of *John* is interpreted as a Copy of the higher *John*.[14]

(14) [$_T$ John T [$_{v^*}$ John v* walked to the store]]

A number of issues arise regarding the FormCopy operation and its formulation in the literature. First, the question arises of whether or not FormCopy can apply freely. Chomsky (2021, p. 25) writes:

> Let's return to simple transitive sentences, such as *John saw X*. Suppose *X* = *John*. With the subject inserted by EM [External Merge] in the predicate-internal position, they are in an IM-configuration [Internal-Merge-configuration]. If FC [FormCopy] applies, the expression will crash at CI [Conceptual-Intentional interface] with a θ-Theory violation. We conclude, then, that like other operations, FC is optional, not applying in this case so there is no deletion, just two repetitions of *John*.[15]

14) This example raises some questions if FormCopy does not apply. If FormCopy does not apply, then the lower *John* would have to be pronounced. This can potentially be ruled out by Theta Theory. If the lower John is not a Copy of the higher *John*, then this construction would contain two instances of *John* that each require separate theta-roles, but there is only one theta-role that is available in this clause.

15) In the case of *John*$_1$ *saw John*$_2$, Chomsky suggests that if FormCopy were to apply so that the lower *John*$_2$ were treated as a Copy of the higher *John*$_1$, the same argument *John* would be interpreted as

Chomsky indicates that FormCopy is optional, but I don't take this to mean that it applies freely. Rather, it applies in some, but not all, cases in which there are multiple inscriptions with the same form.[16] Specifically, it applies at the phase level. Chomsky et al. (2023, p. 25) write that "[i]n technical terms, the point at which FC [FormCopy] applies is referred to as the *phase level*." Limiting FormCopy to the phase level accounts clearly for why FormCopy cannot convert the lower *John*$_1$ into a Copy in *John*$_2$ *saw John*$_1$. As shown in (15), FormCopy should apply when the lower phase head v* is Merged, before *John*$_2$ is Merged. After *John*$_2$ is Merged, the lower v* and its complement are no longer accessible to FormCopy. Thus, FormCopy cannot apply between *John*$_2$ and *John*$_1$.

(15) [T John$_2$ [v* see John$_1$]]

In my model, I adopt this phase-level approach to FormCopy. Once a phase-head is Merged, FormCopy applies if possible. If X and Y are identical inscriptions of arguments in the same phase, then FormCopy converts Y into a Copy of X. Furthermore, FormCopy is beneficial as it enables us to do away with the need for the language faculty to memorize movement operations.[17]

3.7 Strength and Labeling

Chomsky (2013, 2015) utilizes a notion of strength, combined with the need for Labeling, to account for the traditional EPP effects in languages such as English. T is weak so that it must be strengthened. In order to be strengthened it must form an {XP, YP} structure with an argument. Chomsky (2015, p. 7) suggests that a lack of "rich agreement" may be a reason for the weakness of T in English, unlike in null subject languages like Italian that have rich agreement. Chomsky suggests that in a null subject language, T may be able to label by itself.

The Labeling-based analysis of TP in English, in one sense, enables us to do away with the EPP. But on the other hand, the notion of weakness is not entirely clear. If

having two separate theta-roles at CI, which is problematic. This argument isn't clear to me, as FormCopy can apply in Control constructions, like in (13), to convert arguments with separate theta-roles into Copies. However, *John*$_1$ *saw John*$_2$ could be ruled out with respect to Case at SpellOut. The lower *John*$_2$ is marked with accusative Case, and a Case marked argument in English appears to need to be pronounced, but if it is a Copy it is not pronounced. Also, as I note below, FormCopy shouldn't be able to apply between *John*$_1$ and *John*$_2$ anyway because they are in separate phases. The higher *John*$_1$ is outside of the v*P phase.

16) Chomsky (2024) suggests that FormCopy can be used "for convenience" but that "it need not be listed among the admissible operations." How exactly FormCopy can be done away with, but still be adopted for convenience is not clear to me. In this paper, I assume that FormCopy is an operation utilized by the Faculty of Language.

17) Strictly speaking, FormCopy should apply only if X c-commands Y. Note that I haven't formally implemented the c-command component in my model, as it is not crucial to the examples that I implemented.

richness of agreement is at play, then questions arise around languages (e.g., Japanese) which do not appear to show any agreement, but allow null subjects.

If certain heads like the English T can be weak, then the question also arises of what to do with non-finite T. Assume that finite T is weak and requires agreement to be strengthened for Labeling. But non-finite T does not show agreement. In (16), non-finite T, pronounced as *to*, and the embedded subject *Mary* are adjacent. One possibility is that the embedded TP has the structure in (16)b with *Mary* and the non-finite T forming an {XP, YP} structure. In (17), there is no overt argument in the embedded clause, but there should be a Copy of *John* (or PRO) in the embedded clause, as in (17)b. In this case, since the lower Copy of *John* is not pronounced, it should be invisible to Labeling, and thus non-finite T cannot be strengthened. Presumably, non-finite T can label, and thus, it does not require strengthening.

(16) a. John expects Mary to arrive.
b. John expects [$_T$ Mary to arrive]

(17) a. John tried to finish the work.
b. John tried [$_T$ ~~John~~ to finish the work].

Chomsky (2015) assumes that in an ECM construction such as (16), the embedded subject raises to the matrix object position. This follows work by Postal (1974) and Lasnik and Saito (1991), among others. For example, the embedded subject in an ECM construction can be passivized and it behaves like it is in the matrix clause with respect to binding effects. Chomsky argues that the Root *expect* is weak and thus must be labeled by an {XP, YP} structure. On this account, the embedded subject raises and forms an {XP, YP} structure with the matrix verbal root. For example, in (16), the structure {Mary, {expect ~~Mary~~ to arrive}} is formed. Problems with this approach are that the root *expect* must inherit uPhi from the higher v* (if feature inheritance is assumed) and that *expect* has to undergo some form of head movement over *Mary*. If head movement is a Spell-Out phenomenon, then this would happen at Spell-Out.

Due to the complexities involved, I take a simpler approach—*Mary* is in the embedded non-finite TP in (16), where it Agrees with the higher v*, resulting in accusative Case appearing on Mary at Spell-Out. The evidence that the subject of a non-finite clause can behave like the object of the matrix clause is strong, but whether or not this requires the embedded subject to actually undergo IM with the matrix verbal root is not clear. I will simply assume that, due to the lack of an intervening phase boundary, an ECM subject can behave like it is a matrix object even if it is in the embedded clause.[18]

18) As pointed out by an anonymous reviewer, there is evidence from a variety of languages, including English, that an ECM subject behaves like a matrix object. Chen (2018) distinguishes two types of analyses of ECM constructions. In one type of analysis, the subject of an embedded clause raises to matrix object position, as has been argued for

There is evidence that an overt subject can appear in a non-finite clause. In examples such as (18), the subject *him* appears to be in the embedded clause. It seems to be a fairly standard assumption that the embedded subject is in the non-finite T (e.g., see Chomsky & Lasnik, 1977). If Labeling Theory is correct, then there must be some type of agreement relation between non-finite T *to* and *him*.

(18) I want for **him** to go.

If the embedded subject of a non-finite clause remains in the non-finite T, as in (16) and (18), there is a potential Labeling issue. If non-finite T and the subject do not Agree, then there should be a Labeling failure. To get around this problem, I assume that non-finite T contains uPerson that Agrees with an argument. In (16) and (18), the uPerson of non-finite T is checked by the Person feature of the subject. Since the subject is internally-Merged with the non-finite T, the result is an {XP, YP} structure that is labeled with shared Person features. In (17), uPerson is checked by the Person feature of *John*. Since *John* is a Copy here, there is no visible {XP, YP} structure, and *to* labels by itself. Thus, non-finite T can either label by itself, or in an {XP, YP} structure with an argument that it shares a Person feature with.

The question then arises of whether or not there is evidence for agreement between a subject and non-finite T. Notably, agreement in infinitives can be found in a variety of languages. For example, standard Brazilian Portuguese has infinitives that can be inflected for person and number as in (19).

(19) a. (eu/você/ele/ela) fala-**r** (Brazilian Portuguese)
I/you/he/she speak-Inf-∅
b. (nós) fala-r-**mos**
(we) speak-Inf-1Pl
c. (vocês) fala-r-**em**
(you-Pl) speak-Inf-3Pl

with respect to English as well as other languages such as Icelandic (e.g., Sigurðsson, 2006; Sigurðsson & Holmberg, 2008, etc.), and in languages such as Japanese, Korean, Romanian, and Zulu. In another type of analysis, an object is base generated in a matrix clause and is co-indexed with an overt subject pronoun in an embedded clause, as in languages such as Madurese, Tagalog, and Sundanese. See Chen (2018) and references therein. There are also analyses in which an ECM subject remains in the ECM complement clause. For example, Chen (2018) argues that in Puyuma, an ECM phrase (which does not necessarily have to be a subject) remains within an embedded clause. In addition, there a variety of conflicting analyses on these types of constructions. For example, Kuno (1976), Tanaka (2002), and others (e.g., see Kishimoto, 2021) argue that an ECM subject in Japanese raises out of an embedded clause to a matrix object position. However, Kishimoto (2021) argues that an ECM subject in Japanese remains within an embedded clause. The facts and the various analyses regarding these constructions are complex, and thus I acknowledge that my simple assumption that an ECM subject remains within a complement clause is certainly worthy of further investigation.

d. (eles/elas) fala-r-**em**
(they) speak-Inf-3Pl (Pires, 2006, p. 92)

Inflected infinitives are also found in European Portuguese (Raposo, 1987), as well as other Romance languages such as Galician and Old Neapolitan (Groothuis, 2015; Scida, 2004). Hungarian also has inflected infinitivals, as shown in (20).

(20) a. Kellemetlen volt Jánosnak az igazságot bevalla-**ni-a**. (Hungarian)
unpleasant was John-Dat the truth-Acc admit-Inf-3Sg
'It was unpleasant for John to admit the truth.'

b. Péter nem hagyta megnéz-**ne-m** a filmet.
Peter not let-3Sg.Def watch-Inf-1Sg the film-Acc
'Peter did not let me watch the film.' (Tóth, 2000, p. 1)

An anonymous reviewer points out that the distribution of subjects with agreeing infinitives and with non-agreeing infinitives is different. According to Pires (2006, p. 93), in Brazilian Portuguese a non-inflected infinitival requires a PRO subject (which has a local antecedent), and an inflected infinitival has a pro subject (which does not require a local antecedent). Furthermore, a non-agreeing infinitive requires a sloppy reading under ellipses, whereas an agreeing infinitive permits a strict or sloppy reading, and a non-agreeing infinitive does not permit split-antecedence but an agreeing infinitive does. Although an in-depth analysis is beyond the scope of this work, I suggest that these differences boil down to whether or not the non-finite T Agrees partially or fully with an argument. In some Portuguese infinitivals, there may be full agreement with an argument (the infinitival property is due to the lack of tense, not phi-features), and pro is permitted. In non-agreeing infinitivals, there can only be partial agreement, which is not sufficient to check Case on an argument and only PRO is permitted.

Even though there is no clear overt indication of agreement in modern English infinitives, it is possible that there is partial agreement, as found in languages such as Portuguese, Hungarian, etc. Thus, I assume that T can label either by itself or via shared Person features.

Furthermore, I assume that heads can generally label. Mizuguchi (2017, p. 331) suggests that "[h]eads can label only when they are without unvalued features." If a head has unvalued/unchecked features, it is incomplete, and thus it is reasonable to assume that Labeling isn't possible. I adopt this view in my model; heads can label by themselves as long as they lack unchecked features. A root, however, needs to be categorized. Thus, a root cannot label by itself. For example, the root *walk* can be labeled only after it combines with a categorizer N or v.

3.8 Box Theory

Chomsky (2024) proposes Box Theory, in which the traditional A-bar elements (wh-phrases, topicalized phrases) are essentially placed into a box structure that can be accessed by C (and possibly other functional heads associated with topic and focus). The derivation of (21) proceeds along the following lines. When v* is Merged, the lower v*P phase is complete. At this point, the elements within the v*P are no longer accessible. However, the wh-phrase *what* is in a Box, where the Box could be thought of as a structure that contains focused (A-bar) elements.[19] When the matrix interrogative C_Q is Merged, it looks into the Box and finds *what*. The wh-phrase *what* ends up being pronounced at the position of C, but it still remains in the Box. Crucially, *what* in its base position must be converted into a Copy, and so FormCopy must apply.

(21) C_Q John T John [v* buy what]

Timing of insertion into the Box is an issue that arises. Chomsky (2024) suggests that boxing is contingent upon IM. Chomsky writes segregation of a boxed element is "established by IM [Internal Merge], which carries the derivation from the propositional to the clausal domain." Chomsky further writes that "we can think of the element E that is IM-ed to the phase edge as being put in a box, separate from the ongoing derivation D." Chomsky appears to be proposing that boxing results from IM of a particular SO to the phase edge. Note that it is simpler to put an SO into the Box, without boxing being contingent on IM, rather than to do IM of the SO followed by boxing. Furthermore, I also assume that IM of arguments is free (see Section 4). If IM to a phase edge results in boxing, then there could be overgeneration of boxed SOs.

Assuming that boxing happens without IM, it could be that as soon as a wh-phrase is externally Merged with an SO, it goes into the Box, or it could be that it goes into the Box at the phase-level. Also, when a phrase is accessed from the Box, its base position should be treated as a Copy. This means that FormCopy must apply. FormCopy could apply as soon as an SO goes into the Box, or it could apply as soon as the SO is accessed from the Box. Also, consider (22). In this case, C_Q and the subject *who* are within the same phase. So whether or not *who* has to go into the Box isn't clear as C_Q should be able to access *who* without looking into the Box.

(22) C_Q who T who v* buy a house

19) Note that if the Box can store multiple elements and the last element in is the first element that can be accessed, then the Box is similar to a Stack structure that is commonly used in computer science, and that has been used in some linguistics work (e.g., see Fong & Ginsburg, 2014, 2019).

In my implementation, I had to make decisions about the timing of the Box operation. From an implementational perspective, it is easier to put a wh-phrase into the Box as soon as possible, rather than to wait until a phase is complete, since waiting requires checking an already formed structure for wh-phrases (or other phrases that need to go into the Box). Thus, my model places a wh-phrase into the Box as soon as possible. Since the Box is assumed to exist, this applies to wh-subjects too. The model places wh-phrases in the Box and C_Q can only see into the Box. FormCopy applies when C_Q accesses a wh-phrase from the Box.[20]

Assuming the existence of the Box, successive-cyclic wh-movement is potentially an issue. If an argument is in the Box, there is no reason for it to undergo IM to an intermediary position. However, there is evidence for successive-cyclic wh-movement. Some well-known evidence is the existence of partial wh-movement (McDaniel, 1989). For example, in German and Albanian, a wh-phrase can appear in an intermediary position and a wh-phrasal scope marker (or some type of question element) can appear in the relevant scope position, as shown in (23) and (24). In Malay, as shown in (25), a wh-phrase can move to the edge of a clause in which it does not have scope, and be interpreted with scope in a clause in which there is no overt wh-marker.

(23) [**Was**$_i$ glaubst du [**was**$_i$ Hans meint [[**mit wem**]$_i$ Jakob **t**$_i$ gesprochen hat]] (German)
Wh believe you Wh Hans thinks with whom Jakob talked has
'With whom do you believe that Hans thinks that Jakob talked?' (Cheng, 2000, pp. 78–79)

(24) **A** mendon se [**kë**$_i$ ka takuar Maria t$_i$] (Albanian)
Q think-2s that who-ACC has met Mary
'Who do you think that Mary met?' (Turano, 1998, p. 163)

(25) Ali memberitahu kamu tadi [**apa**$_i$ (yang) Fatimah baca **t**$_i$] (Malay)
Ali told you just.now what that Fatimah read
'What did Ali tell you just now that Fatimah was reading?' (Cole & Hermon, 2000, p. 105)

The presence of a wh-phrase in an intermediary position has generally been taken to indicate that a wh-phrase undergoes movement through intermediary positions. Further well-known evidence for successive-cyclic movement is the existence of complementizers that agree with wh-phrases. For example, Irish has a particular complementizer that appears in a non-interrogative embedded clause when a particular wh-phrase undergoes long-distance movement (McCloskey, 1979, 2001).

Box Theory can deal with successive-cyclic wh-movement as follows. An intermediary complementizer can access the Box at externalization, but not in the syntax. This

20) The Box can also be used for non-wh-phrasal focused elements. For example, Chomsky (to appear) gives the example "Bill, John met yesterday" with the topicalized phrase *Bill*.

means that in some languages and constructions, there is access by an intermediary C, of an element in the Box, and this has an influence on Spell-Out. For example, an element in the Box can be pronounced at an intermediary position without actually being in that position in the syntax. Chomsky (2024) writes that in partial wh-movement constructions, presumably such as (23)–(25) above, there is a Labeling violation in the embedded clause in which a wh-phrase appears, as the wh-phrase does not share features with the non-interrogative C.[21] To deal with this issue, Chomsky writes that "the boxed *wh-XP* is accessed" in the intermediary position "and under Externalization, spelled out, but with no labelling problem since the phrase does not appear in the derivation." Although a variety of issues regarding when and how a wh-phrase is accessed for externalization require further examination, this analysis can account for partial wh-movement facts. If this approach is correct, then a wh-phrase does not actually undergo successive cyclic movement, but it can be accessed successive-cyclically, subject to language-internal and cross-linguistic variation, at the point of externalization.

The notion of the Box is beneficial in the following ways. First, it becomes possible to transfer a phase as soon as a phase head is externally-Merged. There is no need for an escape hatch at the v*P phase edge, which was previously assumed to exist to account for A-bar movement of a wh-phrase (e.g., see Chomsky, 2001). When v* is Merged, the v* head and its complement can be transferred. Furthermore, when C is Merged, transfer of the CP can occur. Elements at the edge of a CP presumably are accessed via the Box. Under the traditional Phase Theory view, the edge of a phase is transferred separately from the rest of the phase. But the traditional view complicates transfer of a matrix CP. Under the traditional view, when a matrix CP is formed, first the complement of C is transferred and then the CP edge is transferred, thus requiring two transfer operations to occur at the edge of a CP. This is no longer necessary. When C is Merged, the complete CP can be transferred.

While issues remain regarding the exact definition and nature of the proposed box structure, from an implementational perspective, it has some advantages.

3.9 Arguments as NPs

Although the status of determiners is peripheral to this work, it is necessary to explain how they are implemented in this model. I assume that arguments are NPs and not DPs (contra the typical DP hypothesis of Abney (1987), and much following work). The view that arguments are NPs, and not necessarily DPs, has been suggested by Chomsky (2007), as well as by Van Eynde (2006), Bruening (2009), Oishi (2015), Bruening, Dinh, and Kim (2018), and Bruening (2020), among others.

21) Chomsky cites Riny Huijbregts for pointing this problem out.

If an argument is really a DP, then problems arise regarding phi-features. As Bruening (2009, p. 28) points out, transitive verbs select nominal arguments, whereas they do not select for determiners. The typical approach in the Minimalist Program incorporates Agree relations between functional heads and arguments. Assume that unchecked phi-features on a probe on T or v* form an Agree relation with phi-features on an argument, in which uPhi on T Agree with phi-features on the argument. If the phi-features are on a nominal head N that is contained within a DP, then there is a potential problem because Agree(T, DP) would require T to see inside of the DP to the NP (or Agree would require features of N to percolate up to the D head). It is simpler for T to simply Agree with the head of the NP.

Although determiners can show phi-feature agreement in many languages (e.g., in Romance languages, etc.), gender, person, and number are properties of nominals, not determiners. Number shows up on nouns (e.g., *cat* vs. *cats*). Person and gender show up on pronouns (e.g., *I* vs. *you*, *he* vs. *she*). Agreement between determiners and nominals can occur in languages, such as in (26)a–b from Spanish, but as Bruening (2009, p. 30) points out, "every element in the nominal phrase must agree with the head noun in gender and number," suggesting that the core element in these phrases is the N, rather than a determiner, quantifier, or adjective.[22]

(26)

a.	todos	esos	lobos	blancos	b.	todas	esas	jirafas	blancas
	all.Masc	those.Masc	wolves	white.Masc.Pl		all.Fem	those.Fem	giraffes	white.Fem.Pl
	all those white wolves					all those white giraffes			

(adapted from Bruening 2009, p. 30)

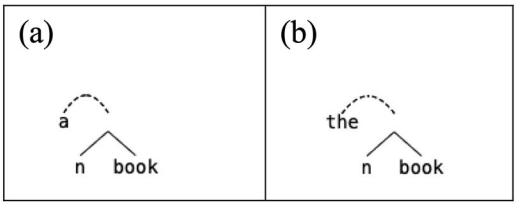
In my model, I indicate an argument as an NP as shown in Figure 4.[23] When there is a D, it is pair-Merged to the NP. Pair-Merge is indicated with a dotted arc. Given two SOs X and Y, when X is pair-Merged with Y, forming <X, Y>, X is less prominent than Y and generally not accessible to syntactic operations (see Chomsky, 2000, 2004). When the NP is Merged with another SO, the pair-Merged D is invisible to Agree relations.

22) I assume that phi-features are present on N only, but see Danon (2011) for arguments that a D can share the phi-features of an NP complement.

23) Much further examination of the categorial status of arguments is warranted, but this topic is beyond the scope of this work. For further recent discussion of the categorial status of arguments, see Blümel and Holler (2021), and references therein.

Figure 4

NPs Pair-Merged With D



4 Free Merge

Assume that IM is completely free, so that elements within an SO can be freely internally set-Merged to the root of the SO. This is untenable as an infinite number of possible SOs can be formed for every phrase and sentence.[24, 25]

If the discussions in the previous sections are correct, head movement is not a possible syntactic operation, and this greatly limits the Free Merge possibilities. For example, if head movement could apply freely, then illicit derivations such as those in Figure 5 would be possible; heads such as n, *book*, and *will* would be able to undergo IM. The resulting structures, however, would have to crash. Although structures of this sort could be ruled out as involving failures of Labeling and/or interpretation, generating them would involve a great deal of unnecessary and wasteful work. We can deal with this issue by simply assuming that head-movement (IM of heads) is not a possibility.

Free Merge, if it exists, must apply to IM of arguments at the phase level only. Assuming Box Theory, only topicalized/focused arguments such as wh-phrases can escape from a phase, and escape is via the Box. Not permitting non-focused/topicalized arguments to escape greatly constrains Free Merge. Thus, I assume that Free Merge is limited by phase boundaries.

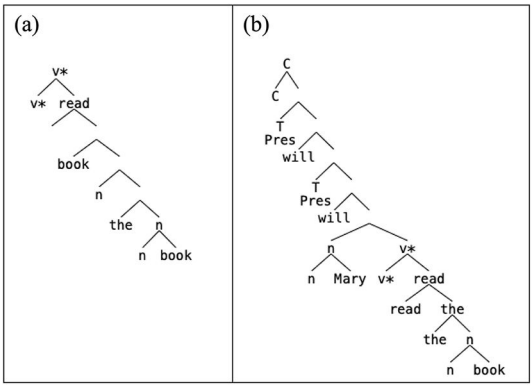
Given the constraints of the language module, as presented in this paper, it turns out that allowing Merge of arguments (NPs) to apply freely within a phase is not necessarily a problem. Ill-formed constructions can generally be ruled out as Labeling failures.

24) For example, starting with 4 lexical items, if internal and external Merge are allowed to apply freely, then given 8 possible Merge operations, there are more than 7 million SOs that can be generated (Ginsburg & Fong, 2018). If there is no limit on the number of Merge operations, then the number of SOs that can be generated is infinite.

25) Although external Merge could be free in some sense, in my model, external Merge is not free, since lexical items are selected from an input stream and externally Merged together. In language, there are clearly constraints on which lexical items can be externally Merged together.

Figure 5

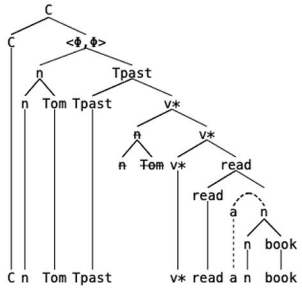
Illicit Derivations Involving Head Movement



Consider the derivation of the simple statement in Figure 6. When v* is Merged, the lower v*P phase is transferred. Then the subject *Tom* is externally set-Merged. After Tense (past tense Tpast) is Merged, *Tom* undergoes IM with Tpast. After C is Merged, at the phase-level, FormCopy applies and converts the lower inscription of *Tom* into a Copy. The Spell-Out is computed as shown, whereby the frontier of the tree structure is converted via pronunciation rules (PF rules) into the correct output. Tpast and *read* combine to form the past tense *read* and functional elements such as v* and C are not pronounced.

Figure 6

Tom Read a Book



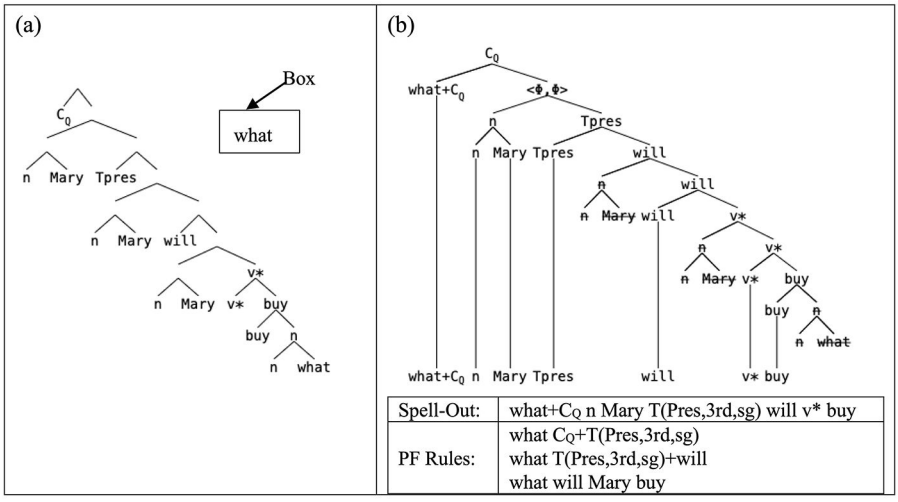
Spell-Out:	C n Tom T(Past,3rd,sg) v* read a n book
PF Rules:	Tom T(Past,3rd,sg)+read Tom read a book

Note. Chomsky (2015, p. 10).

Next, consider the derivation of the wh-construction in Figure 7. When the v*P phase is completed, *what* is inside of the Box. The interrogative C_Q is Merged, and then it looks into the Box and finds *what*. Importantly, *what* does not undergo IM to the CP. When *what* is accessed by C_Q, FormCopy applies to *what* in its base position. Assume that FormCopy can still access the lower phase via the Box. FormCopy also converts the lower inscription of the subject *Mary* into a Copy—this FormCopy operation happens at the phase level. The frontier of the derivation is show in Figure 7b. At Spell-Out, C_Q forces Tense to combine with it, forming C_Q+T, and also T forces the auxiliary *will* to combine with it, so the result is C-T-Aux. This is not movement in the syntax, but rather displacement in the pronunciation of lexical items, as discussed in Section 3.5 above.

Figure 7

What Will Mary Buy?

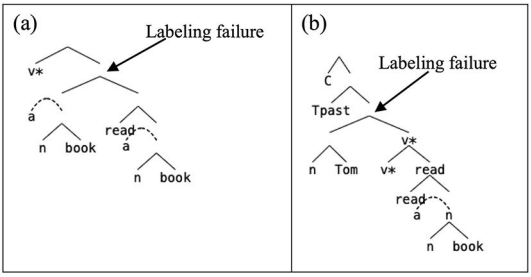


Note. Pesetsky and Torrego (2001, p. 369)

I next turn to crashed derivations that result from Free Merge. Two failed derivations (crashed derivations) of *Tom read a book* are shown in Figure 8. As Free Merge of nominals is permitted at the phase level, it is possible for the object *a book* to undergo IM with the SO headed by *read*, and it is also possible for the subject *Tom* to simply remain in its base position (*Tom* is free to not undergo IM). In each case, there are Labeling failures due to {XP, YP} structures that lack shared features.

Figure 8

Crashed Derivations of "Tom Read a Book"

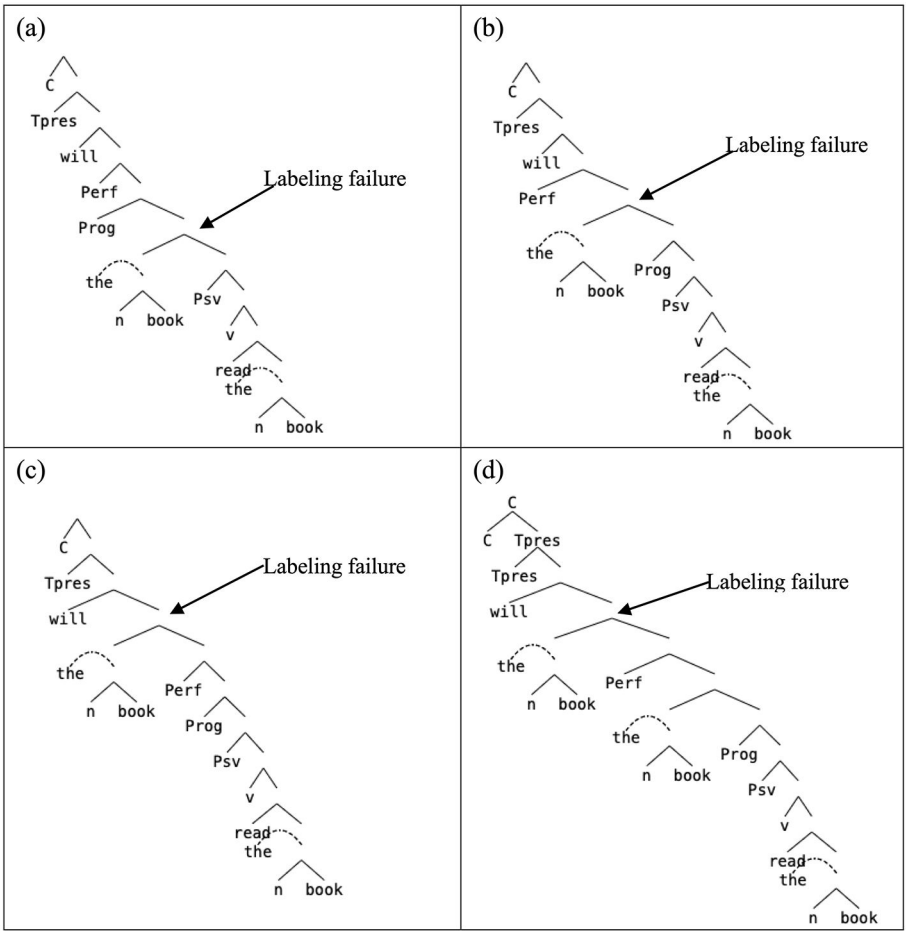


In some cases, there can be a large number of crashed derivations. Consider two crashed derivations of (27). These result from IM of an argument to a position in which Labeling cannot occur. In Figure 9a–d, the passivized object *the book* does not undergo IM to the TP. In each derivation, there is a Labeling failure at the position in which *the book* has undergone IM, due to a lack of shared features—the results are unlabelable {XP, YP} structures.

(27) The book will have been being read.

Figure 9

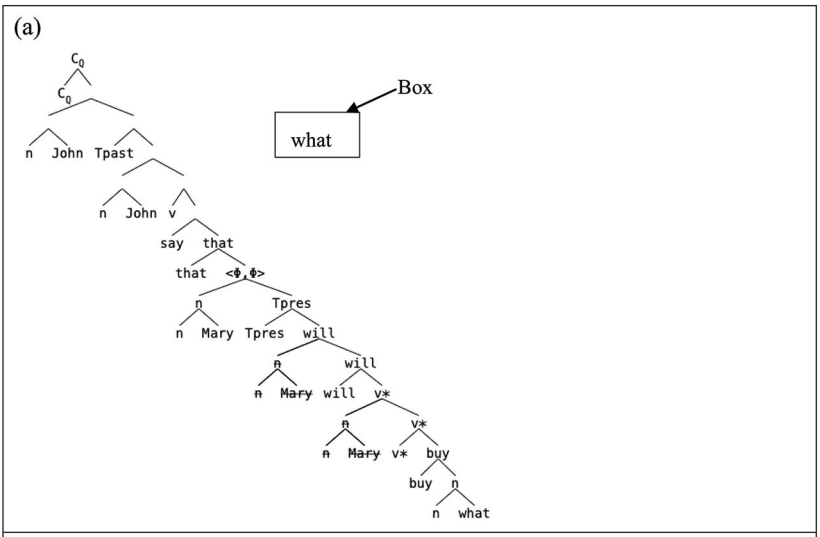
Labeling Failures for "The Book Will Have Been Being Read"



Next, consider *What did John say that Mary will buy?* which contains long-distance wh-movement. A successful derivation is shown in Figure 10. This construction contains 3 phases. The verb *say* takes a clausal complement, but it is not an ECM verb. So I assume that it occurs with the non-phasal v (Chomsky, 2001), which does not Agree with an argument. After *what* is initially Merged, it is inserted into the Box. At the embedded CP phase level, after *that* is Merged, FormCopy applies to the lower inscriptions of *Mary*. When the matrix C_Q is Merged, it looks into the Box and finds *what*. At Spell-Out, *what* is pronounced together with C_Q, and Tpast is pronounced adjacent to C_Q, resulting in pronunciation of *did*.

Figure 10

What Did John Say That Mary Will Buy?



Spell-Out:	what+C_Q n John T(Past,3rd,sg) v say that n Mary T(Pres,3rd,sg) will v* buy
PF Rules:	what C_Q+T(Past,3rd,sg) what did John say that Mary T(Pres,3rd,sg)+will what did John say that Mary will buy

Note. Pesetsky and Torrego (2001, p. 370).

Given Free Merge, there are five crashed derivations of *What did John say that Mary will buy?*. All of these are shown in Figure 11. These crash because of unlabelable {XP, YP} structures. In Figure 11a, *what* undergoes IM with the SO headed by the root *buy* to form an unlabelable {XP, YP} structure. In Figure 11b, *Mary* remains in-situ, resulting in an unlabelable {XP, YP} structure because *Mary* and v* do not share features. In Figure 11c, *Mary* undergoes IM with the SO headed by *will* and remains in this position, resulting in an unlabelable {XP, YP} structure because *Mary* and *will* do not share features. In Figure 11d-e, the derivations crash in the matrix clause because *John* remains in its base position forming an {XP, YP} structure with v, with which it does not share features. These two derivations are almost identical except that *Mary* has undergone IM from v* to Tpres in the embedded clause in Figure 11d, whereas in Figure 11e, *Mary* undergoes IM to the SO headed by *will* before it lands in the TP.

I next turn to a typical Control construction, such as *John tried to win*. In this case, there are crucially two separate arguments *John* in the same phase, assuming that the lower non-finite TP is not a phase. A convergent derivation is shown in Figure 12. $John_1$ is externally-Merged in theta-position in the embedded clause. $John_2$ is externally Merged with the matrix v in theta-position. Both $John_1$ and $John_2$ undergo IM to their respective TPs. In this case, FormCopy applies three times. FormCopy explains how $John_1$ and $John_2$ have the same referent, but separate theta-roles.

Given Free Merge, for *John tried to win*, a number of potentially problematic situations arise involving IM of the "wrong argument" as well as involving multiple instances of *John* (multiple specifiers) in the same phrase edge.

Figure 11

What Did John Say That Mary Will Buy (Crashes)

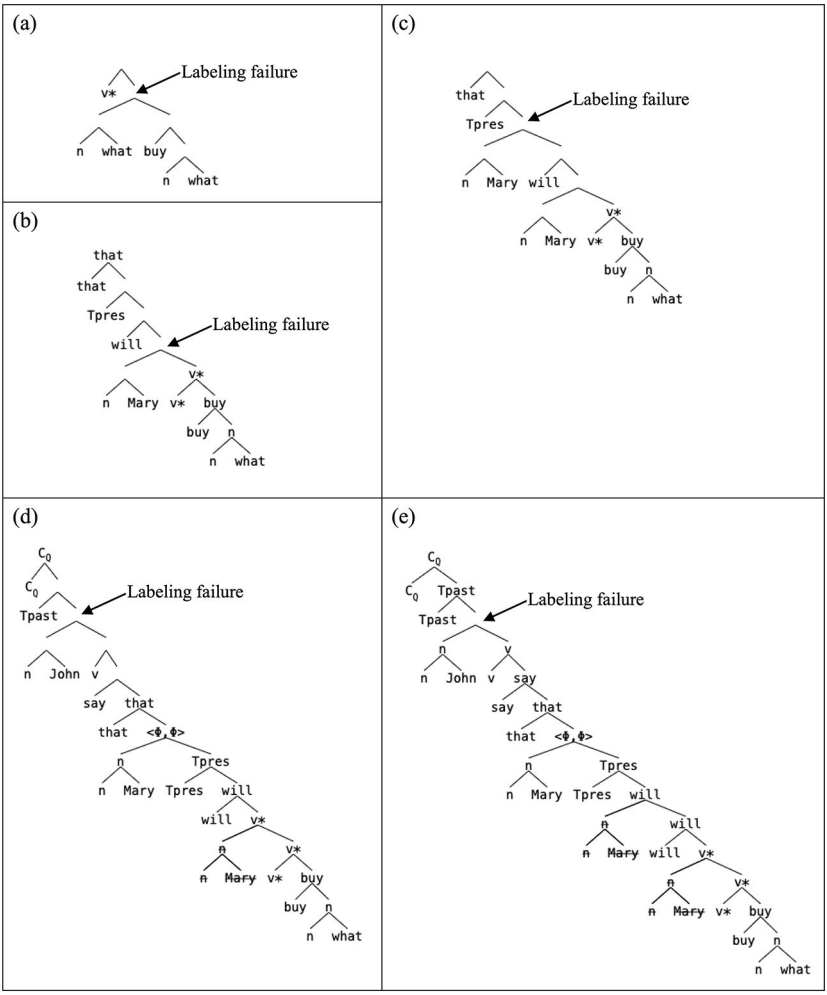
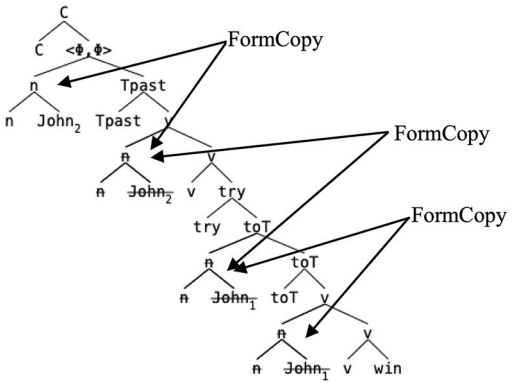


Figure 12

John Tried to Win

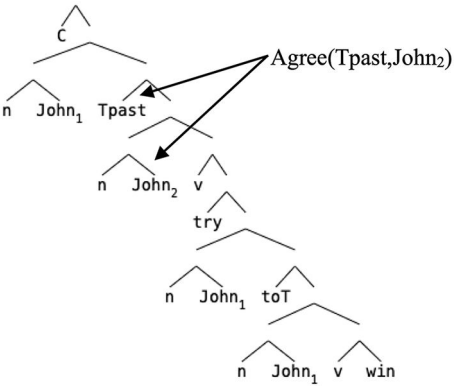


Note. Chomsky (2021, p. 21).

In Figure 13, Tpast undergoes an Agree relation with *John*$_2$ (not *John*$_1$). The uPhi on Tpast probe and Agree with the phi-features on *John*$_2$. Then *John*$_1$ (from the embedded clause) undergoes IM to the matrix TP. This derivation appears strange, since the wrong *John* moves to TP. However, this is permitted if Merge is truly free.[26]

Figure 13

John Tried to Win (John$_1$ *in TP and John*$_2$ *in vP)*



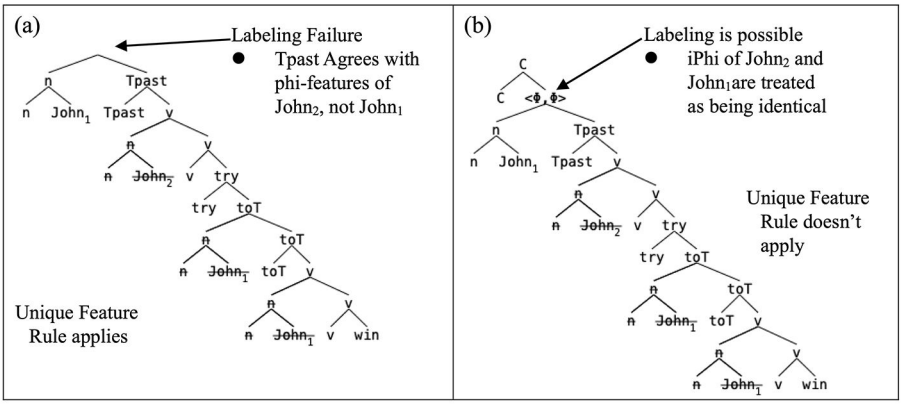
26) Note that this issue only arises when there are multiple arguments in the same phase. For example, in *What did John say that Mary will buy?* (Figure 10), *Mary* is contained within a separate phase from *John*, so *Mary* cannot undergo IM to the matrix clause.

One possibility is that this derivation in Figure 13 crashes because the phi-features of $John_1$ and $John_2$, although identical in terms of person, number, and gender, are treated differently because they are associated with separate arguments. This can be modeled with what I refer to as a **Unique Feature Rule**—$John_1$ comes with iPerson:3rd**1**, iNumber:sg**1**, and iGender:**1**, where the final 1 is a unique feature identifier.[27] $John_2$ comes with person, number and gender features that are identically valued to those of $John_1$, but the unique feature identifier is 2 instead of 1, so the features are iPerson:3rd**2**, iNumber:sg**2**, and iGender:**2**. Utilizing this Unique Feature Rule, this derivation can be ruled out. The uPhi on Tpast Agree with the phi-features of $John_2$. Then after $John_1$ undergoes IM to the TP, Minimal Search finds the phi-features on Tpast and on $John_1$, but they are treated as being different, due to the Unique Feature Rule. This is ruled out as a Labeling failure, shown in Figure 14a.

I also modeled this construction in my computer model without the Unique Feature Rule. When the Unique Feature Rule does not apply, then this derivation converges, as shown in Figure 14b. FormCopy converts all lower instances of John into Copies, and the highest $John_1$ only is pronounced. Minimal Search finds equally valued person, number, and gender features on $John_1$ and on Tpast—it does not matter that Tpast has obtained these phi-features via agreement with $John_2$ instead of $John_1$. Crucially, if this derivation is permitted, there is no problem for Spell-Out—the correct *John tried to win* results.

Figure 14

Completed Derivations of 'John Tried to Win' ($John_1$ in TP and $John_2$ in vP)

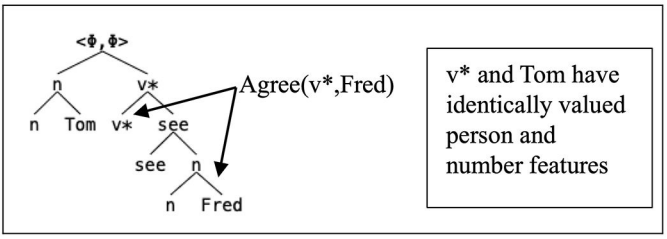


27) Although gender is not important in English, I still assume that it is a feature of a noun. Eliminating gender would not change this analysis.

Although the Unique Feature Rule sounds like an added, and possibly unnecessary complexity, it is necessary. Consider the derivation of Figure 15. Without the Unique Feature Rule, if *Tom* remains within the v*P and does not undergo IM to the matrix TP, Labeling would be possible within the v*P. This is because the uPhi of v* are checked by the phi-features of *Fred*. The person, number, and gender features of *Tom* and *Fred* are identical. If features are not treated as unique, Labeling should be possible within the v*P.

Figure 15

Tom Saw Fred



Note. Labeling possible if features aren't treated as unique.

In order to rule out superfluous Labeling as in Figure 15, the Unique Feature Rule, defined in (28), is required. Features that are valued the same way, but that are associated with different lexical items, are not treated as being identical by the language module.

(28) **Unique Feature Rule**: Features associated with a particular lexical item are unique from identically valued features associated with a separate lexical item. (For example, iPerson:3rd**1** of X are not identical to iPerson:3rd**2** of Y.)

The derivations in Figure 16a–b below involve what would traditionally be referred to as multiple specifiers. In Figure 16a, $John_1$ is initially Merged in theta-position in the non-finite clause. Then $John_1$ undergoes IM to the matrix vP theta position, followed by EM of $John_2$. Assume that there are no problems for theta-role assignment, in accord with Theta Theory (Chomsky, 1981), so $John_2$ is able to obtain a theta role.[28] Then $John_1$ undergoes IM to the TP. Figure 16b is similar. In this case, $John_2$ is successfully Merged in matrix theta position. Then $John_1$ undergoes IM to the vP. $John_2$ undergoes IM to the TP, but Tpast Agrees with the closest NP that it c-commands, $John_1$. In both of these

28) If $John_2$ needs to be directly Merged with v (without anything intervening) in order to obtain a theta-role, then this is potentially a violation of Theta Theory. But whether or not this is the case is not clear. If one assumes that as long as $John_2$ is Merged with v (even if there is an intervening element), a theta-role can be assigned, then $John_2$ can obtain a theta-role.

derivations, Tpast Agrees with a different *John* from the *John* that appears in the TP. These derivations are ruled out by the Unique Feature Rule, so that the phi-features of $John_2$ are treated as being different from the phi-features of $John_1$, as shown in Figure 17.

Figure 16

John Tried to Win (Multiple Specifiers)

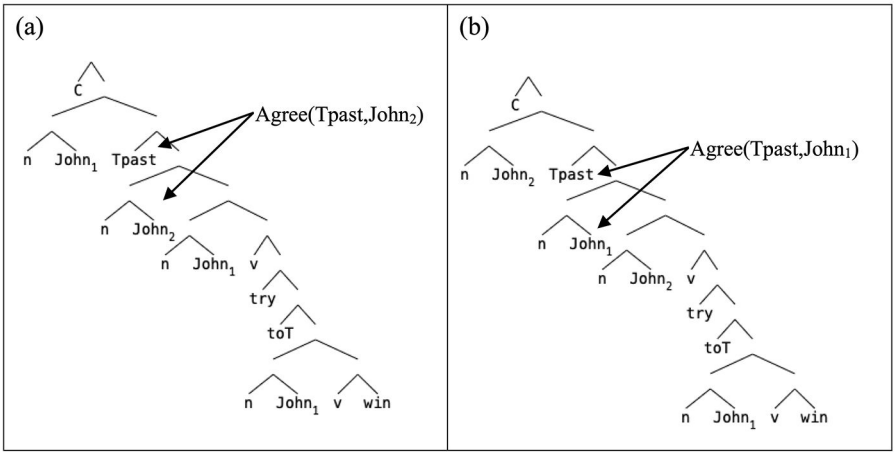
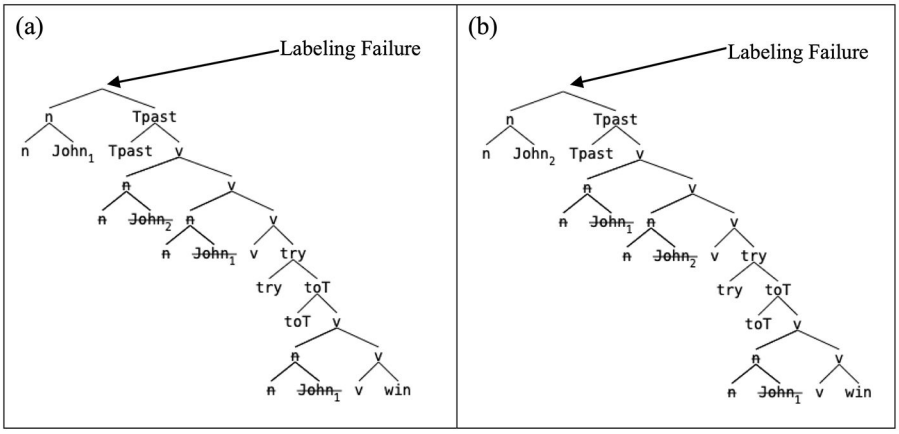


Figure 17

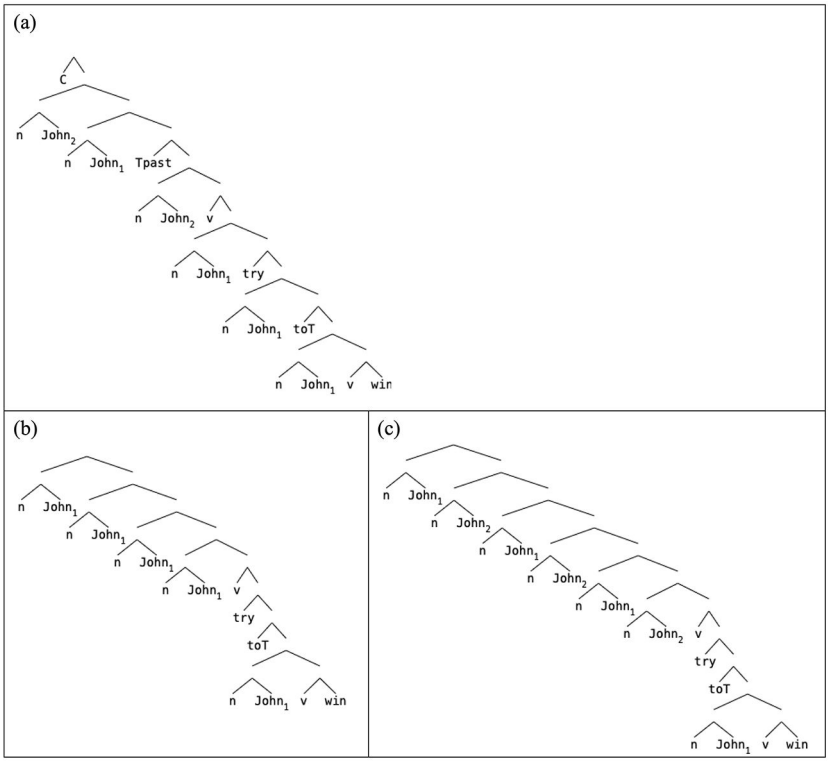
Crashed Derivations: Unique Feature Rule Applies



Free Merge needs to be constrained to prevent multiple instances of IM of identical arguments within a single phase. For example, if Merge of an argument is completely free within a phase, then derivations such as in Figure 18 will arise in which the same argument undergoes IM multiple times with the root of the SO. Thus, it is necessary to prevent an argument from being successively remerged. Note that given FormCopy, all of these could potentially converge with the correct output.

Figure 18

Derivations With Successive Applications of IM for 'John Tried to Win'



To block derivations such as these, which can potentially result in infinite loops, there needs to be a rule that blocks successive IM of multiple arguments to the same phrase. Generation of these ill-formed structures can be blocked by the following rule that simply bans consecutive applications of IM. After one application of IM of an argument, the next operation cannot be IM. This solves the relevant problem and structures such as those in Figure 18 cannot be generated. Thus, I will assume that this constraint No Successive IM holds.

(29) **No Successive IM**: *IM IM (An IM operation cannot directly follow another IM operation.)

Note that if (29) holds and successive applications of IM are not permitted, then constructions in which there are consecutive applications of IM should not appear in language. Whether or not this is truly the case is an open question. Some languages have multiple wh-fronting (e.g. Bulgarian, Serbo-Croatian) that could potentially be formed by multiple applications of IM (e.g., see Boeckx & Grohmann, 2003; Bošković, 2002; Rudin, 1988), as in the following examples.

(30) Koj kogo vižda? (Bulgarian)
Who whom sees
Who sees whom? (Rudin, 1988, p. 449)

(31) Ko kogo vidi? (Serbo-Croatian)
Who whom sees
Who sees whom? (Rudin, 1988, p. 449)

Box Theory offers an explanation. If these arguments are actually in the Box, from where they are accessed, they are not treated like typical arguments that are set-Merged with the core SO. Thus, their presence may be permitted at Spell-Out, with language-related idiosyncrasies that are beyond the scope of this work. They are pronounced together, but they do not actually involve consecutive applications of IM.

If the arguments in this paper are correct, Free Merge of arguments can generally be constrained by Labeling, but Free Merge also produces multiple convergent derivations for target constructions, which I turn to next.

5 Overgeneration

The main problem that Free Merge raises is that of overgeneration. Given Free Merge of arguments, a large number of crashed derivations can occur. Furthermore, a single construction can have multiple convergent derivations. As discussed in the previous sections, I used a computer model to implement Free Merge of arguments within a particular phase. The model also incorporates the Unique Feature Rule, which requires features associated with a particular argument to be uniquely identified, and the No Successive IM rule, which blocks consecutive applications of IM. The total numbers of convergent and crashed derivations for the main sentences generated by the model used for this paper are shown in Table 3–Table 6, which list the numbers of derivations that crash and converge for each target construction. All complete crashed and convergent derivations are available in the Supplementary Appendix (see Ginsburg, 2024).

Table 3

Basic Statements

	Sentence	Crash	Converge
1	Tom saw Fred	2	1
2	Tom read a book. (Chomsky, 2015, p. 10)	2	1
3	He thinks that John read the book.	3	1
4	Mary arrived.	3	5
5	Tom will read a book.	3	2
6	The book was read.	7	9
7	John expects Mary to arrive.	16	5
8	Mary thinks that Sue will buy the book. (Pesetsky & Torrego, 2001, p. 357, originally from Stowell, 1981)	5	2
9	She was reading a book.	3	2
10	She had read a book.	3	2
11	She has been reading a book.	5	4
12	She will have been reading a book.	9	8
13	The book was being read.	15	17
14	The book had been being read.	31	33
15	The book will have been being read.	63	65

Table 4

Control Constructions

Example	Sentence	Crash	Converge
1	John tried to win. (Chomsky, 2021, p. 21)	24	12
2	John tried to finish the work.	25	12
3	Emily forgot to do the homework.	25	12
4	Emily will have forgotten to do the homework.	217	108

Table 5

Wh-Questions

Example	Sentence	Crash	Converge
1	Who do you expect to win? (Chomsky, 2015, p. 10)	4	1
2	What will Mary buy? (Pesetsky & Torrego, 2001, p. 369)	3	2
3	What did Mary buy? (Pesetsky & Torrego, 2001, p. 357)	2	1
4	Who bought the book? (Pesetsky & Torrego, 2001, p. 357)	2	1
5	Bill asked what Mary bought. (Pesetsky & Torrego, 2001, p. 378)	3	1
6	What did John say that Mary will buy? (Pesetsky & Torrego, 2001, p. 370)	5	2
7	*Who do you think that read the book? (Chomsky, 2015, p. 10)	3	1
8	Who do you think read the book? (Chomsky, 2015, p. 10)	3	1
9	What do you think that John read?	3	1
10	What do you think John read?	3	1
11	What did John say Mary will buy? (Pesetsky & Torrego, 2001, p. 370)	5	2
12	*Who did John say that will buy the book? (Pesetsky & Torrego, 2001, p. 371)	5	2
13	Who did John say will buy the book? (Pesetsky & Torrego, 2001, p. 371)	5	2

Table 6

Yes/No Questions

Example	Sentence	Crash	Converge
1	Will Tom read a book?	3	2
2	Does Tom read a book?	2	1
3	They asked if the mechanics fixed the cars. (Chomsky, 2013, p. 41)	3	1
4	Will the book have been being read (by the students)?	63	65

The question arises of whether or not it is reasonable for there to be multiple crashed derivations for a single construction. For example, the following example (see discussion of (27) above) has 63 crashed derivations.

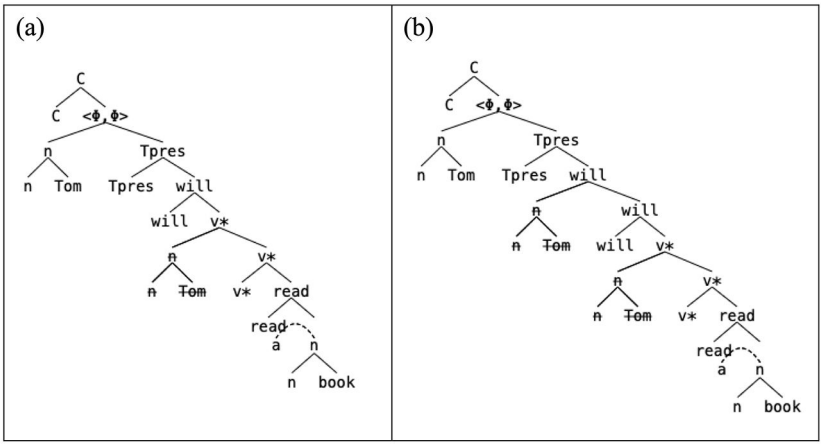
(32) The book will have been being read.

The ideal model would most likely be one that does not generally produce crashed derivations. That said, it is crucial to note that if Merge is free, then derivations of this sort can be generated. Given Labeling though, they generally crash, which is desired.

The second issue related to overgeneration regards well-formed derivations. Many (although not all) of the constructions in Table 3–Table 5 have more than one convergent derivation. For example, for *Tom will read a book* there are two possible convergent derivations that differ with respect to how many times the surface subject *the book* has undergone IM. Since Merge within a phase is free, there is no need for successive cyclic movement, and thus nothing blocks these multiple derivations. In Figure 19a, the subject *Tom* undergoes IM to Tpres. In Figure 19b, *Tom* undergoes IM to *will* and then to Tpres. Both of these options are possible, and the result is the same well-formed output. In the latter case, there is no problem for the structure with respect to Labeling. Since *Tom* undergoes further IM, it isn't visible to Labeling of {Tom, will Tom read the book}, and an unlabelable {XP, YP} structure does not result.

Figure 19

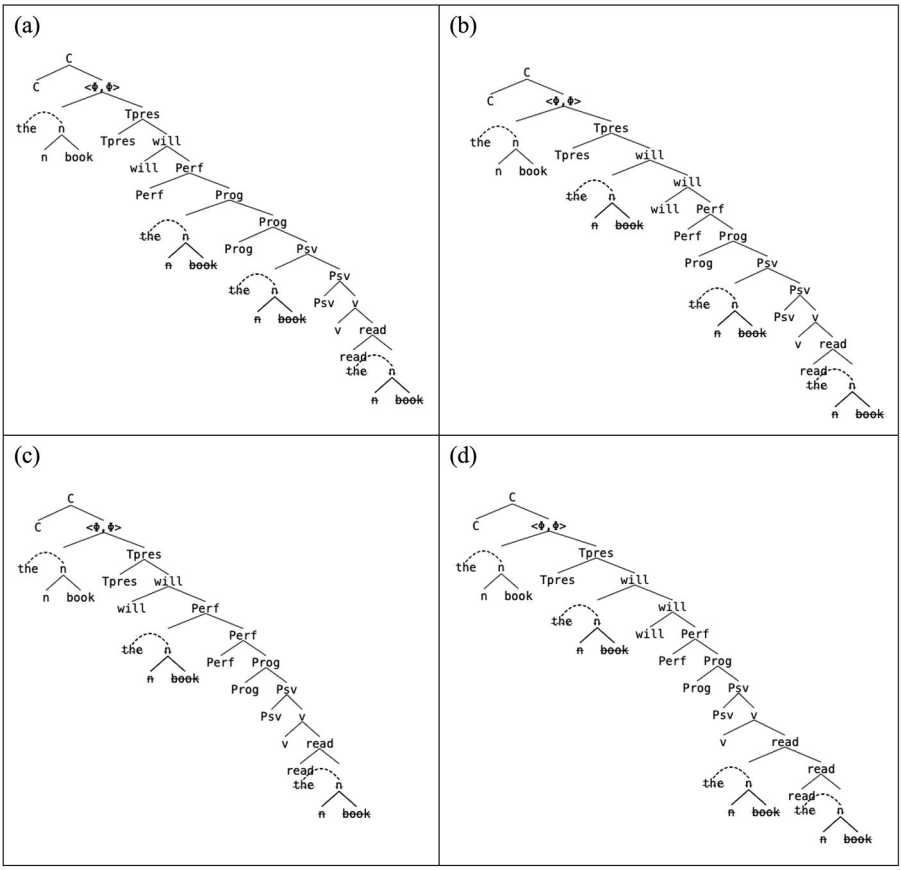
Multiple Derivations of 'Tom Will Read a Book'



Four possible derivations (out of 65) for *The book will have been being read* are shown in Figure 20. In Figure 20a, *the book* undergoes IM to Psv, Prog, and Tpres. In Figure 20b, it undergoes IM to Psv, will, and Tpres. In Figure 20c, *the book* undergoes IM to Perf and Tpres. In Figure 20d, it undergoes IM to *read*, *will*, and Tpres.

Figure 20

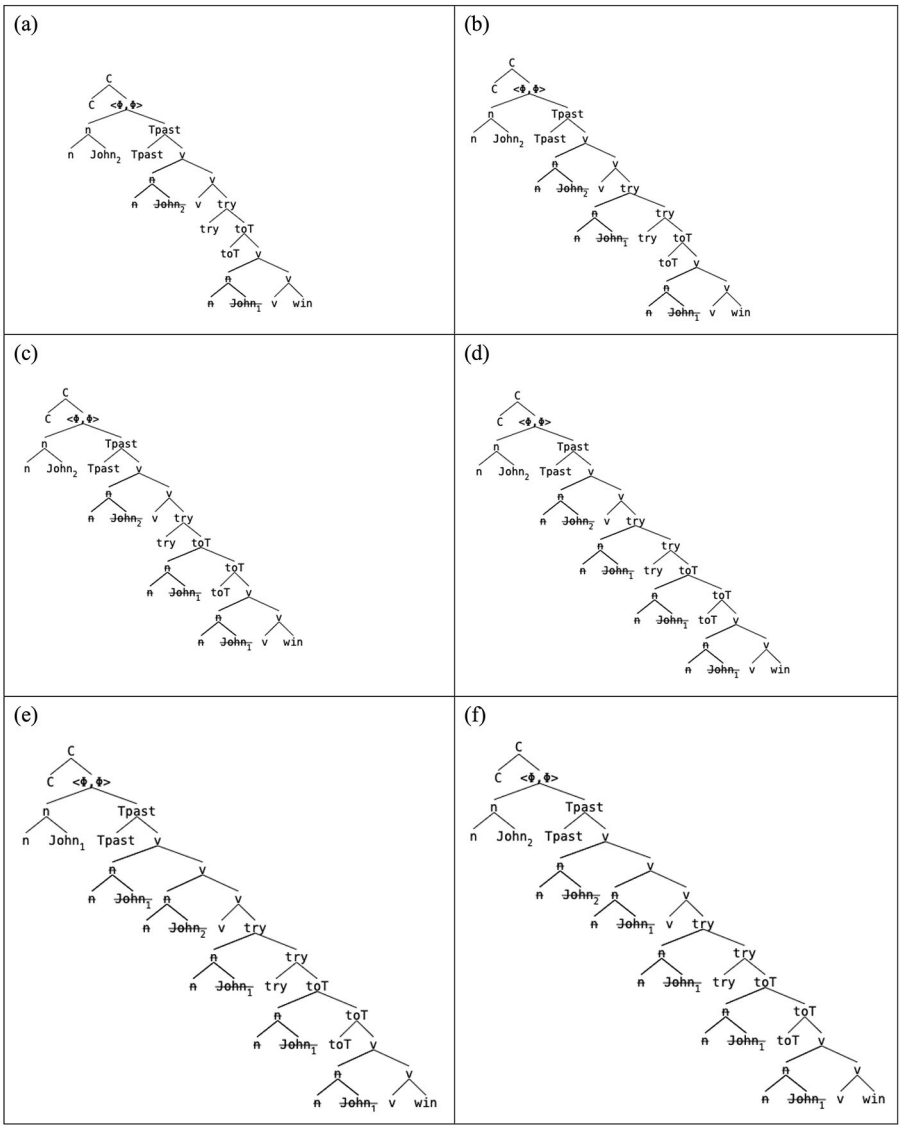
Multiple Derivations of 'The Book Will Have Been Being Read'



Six successful derivations (out of 12) of *John tried to win* are shown in Figure 21. Note that the lower $John_1$ can remain in-situ in theta-position, or it can undergo IM to a higher position. FormCopy converts $John_1$ into a Copy, so there are no Labeling problems in these positions. Both $John_1$ and $John_2$ are Merged in theta-positions and FormCopy results in only $John_2$ being pronounced. The two derivations in Figure 21e–f involve IM of both $John_1$ and $John_2$ with v. Since Tpast Agrees with the same inscription of John that is present in the TP, Labeling is possible (without violating the Unique Feature Rule). None of these derivations cause problems for Labeling, and all converge successfully.

Figure 21

Multiple Derivations of 'John Tried to Win'



Given Free Merge, successive-cyclic movement is not required.

Evidence from quantifier stranding indicates that an argument can internally set-Merge in intervening positions before arriving in subject position. The quantifier *all* can

be stranded as in (33)b–c, which can be accounted for if *all the children* raises through a VP-internal position. Assuming that *arrive* is unaccusative, *all the children* should originate as the complement of *arrive*. In particular, in (33)c, if the adverbial *quickly* is within the VP, then *all* must also be in a VP-internal position.[29]

(33) a. **All** the children arrived.
b. The children **all** arrived.
c. The children quickly **all** arrived.

McCloskey (2000) gives the following examples with quantifier stranding from West Ulster English, which support the idea that there is internal set-Merge (successive-cyclic IM) of an argument in intervening positions.[30]

(34) a. What **all** did he say that he wanted to buy?
b. What did he say **all** that he wanted to buy?
c. What did he say that he wanted **all** to buy?
d. What did he say that he wanted to buy **all**? (McCloskey, 2000, p. 62)

These examples demonstrate that an argument can undergo IM in intermediary positions. A stranded quantifier is an indication that IM has occurred. However, these examples are compatible with the notion that IM can, but need not, occur in intermediary positions, which is what my model predicts. When there is a stranded quantifier, IM in intervening positions has occurred. When there is no stranded quantifier, IM may or may not have occurred.

Another important issue raised by this Free Merge model is overgeneration—it erroneously predicts a few derivations to be well-formed, contrary to fact.

In most cases, Labeling is sufficient to account for the general requirement that subjects appear in the TP in English. Consider the derivation of *She was reading a book*. In the convergent derivation shown in Figure 22, *she* undergoes IM to Tpast, and shared phi-features label. If the subject *she* does not move to the TP, the derivation will crash. For example, in Figure 23a, the subject *she* remains in-situ in the v*P and in Figure 23b,

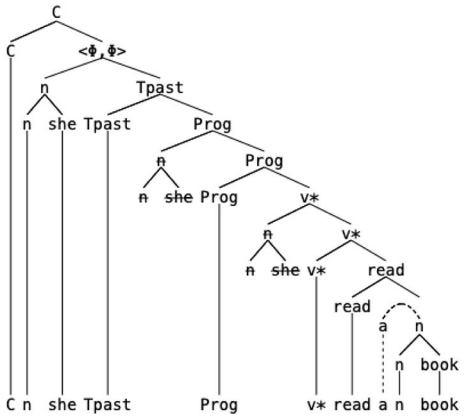
29) See Sportiche (1988) and Stroik (2009), among others, for evidence of successive-cyclic movement of subjects.

30) How exactly Labeling works in cases in which a quantifier is stranded remains unclear, as discussed in Blümel (2018). For example, it isn't clear how *all* and the VP *arrived* are labeled in (33), as they potentially form an {XP, YP} structure {{**all** ~~the children~~}$_1$, {arrive t$_1$}} with no shared features. Blümel discusses several possibilities. One possibility is that the quantifier functions as an adverbial, in which case it could be an adjunct, and thus potentially not cause problems for Labeling. Another possibility, referred to as Distributed Deletion (Fanselow & Ćavar, 2002), is that copies can be selectively pronounced so that "pronunciation of members of a movement chain can be scattered (Blümel, 2018, p. 67)." In my model, the easiest way to implement Labeling with quantifiers would probably be to treat them in a similar manner to determiners, and make them adjuncts that are pair-Merged to an SO. As pair-Merged adjuncts they would not cause problems for Labeling.

the subject undergoes IM with Prog. In both cases, the result is an unlabelable {XP, YP} structure which crashes due to a lack of shared features.

Figure 22

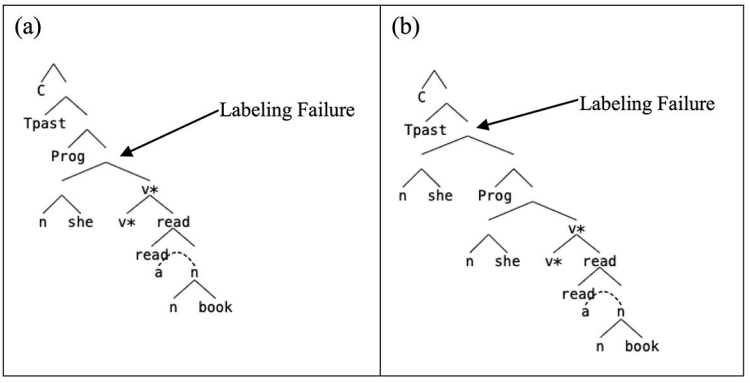
She Was Reading a Book



Spell-Out:	C n she T(Past,3rd,sg) be -ing v* read a n book
PF Rules:	she T(Past,3rd,sg)+be she was -ing+read she was reading a book

Figure 23

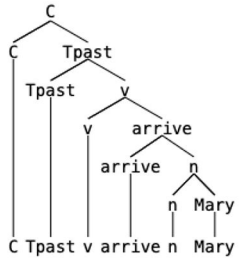
She Was Reading a Book (Crashed Derivations)



There are issues, however, with unaccusative and passive constructions. Given the standard assumption that the surface subject of an unaccusative and passive originates as an object, derivations in which an object remains in situ are not necessarily ruled out. The model incorrectly generates *arrived Mary*, shown in Figure 24, *was read the book* in Figure 25, and *John expects to arrive Mary* in Figure 26. Crucially, all of the other convergent derivations for these examples that are generated by this model result in the well-formed output (4 other derivations successfully converge as *Mary arrived*, 8 other derivations as *The book was read*, and 4 other derivations as *John expects Mary to arrive*). Thus, the model does produce the correct derivations most of the time.

Figure 24

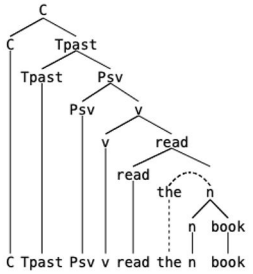
**Arrived Mary (Mary Arrived)*



Spell-Out:	T(Past,3rd,sg) v arrive n Mary
PF Rules:	T(Past,3rd,sg)+arrive arrived Mary

Figure 25

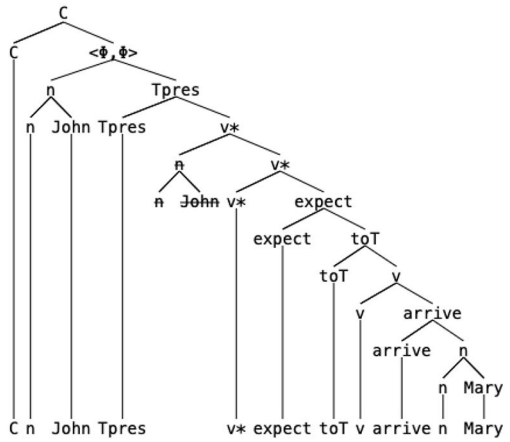
**Was Read the Book (The Book Was Read)*



Spell-Out:	T(Past,3rd,sg) be -en v read the n book
PF Rules:	T(Past,3rd,sg)+be was -en+read was read the book

Figure 26

**John Expects to Arrive Mary (John Expects Mary to Arrive)*

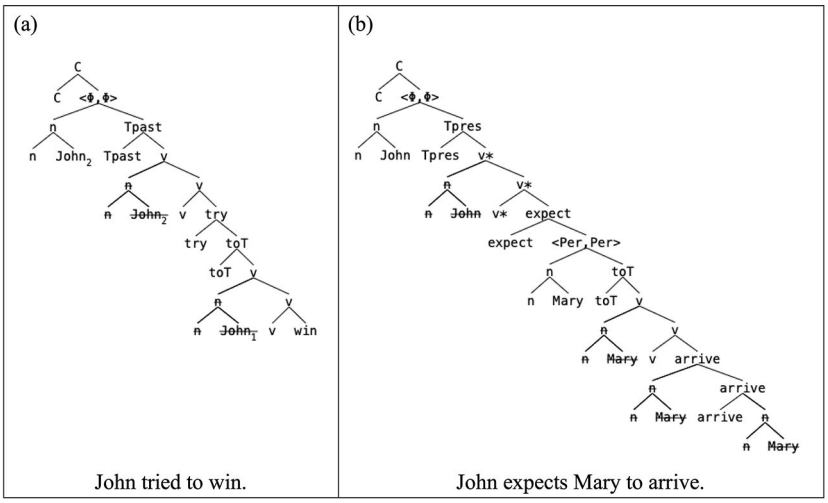


Spell-Out:	C n John T(Pres,3rd,sg) v* expect toT v arrive n Mary
PF Rules:	John T(Pres,3rd,sg)+expect John expects to arrive Mary

In Figure 24–Figure 26, T probes and Agrees with the underlying object (which is the closest argument), and T's uPhi are checked. Then T is not part of an {XP, YP} structure, so if it labels, it must label by itself. As discussed in Section 3.1, Chomsky deals with the EPP requirement for an overt subject in English by relying on strength, with the proposal that English T is too weak to label by itself, and so it requires an {XP, YP} configuration for Labeling. If T is weak and is stipulated to require an {XP, YP} structure for Labeling, then these derivations are correctly ruled out. However, strength is an unclear stipulation, which I do not adopt. See Section 3.7 above. In my model, non-finite T may or may not have an overt specifier. In the derivation for *John tried to win*, shown in Figure 27a, $John_1$ remains in-situ (although it can also raise to toT, where it will be converted into a Copy). In the derivation of *John expects Mary to arrive* shown in Figure 27b, *Mary* appears in the non-finite clause forming an {XP, YP} structure with toT (where shared Person features label). If non-finite T were weak, then toT would always require an overt "specifier", contrary to fact.

Figure 27

Non-Finite T in Control Constructions



The simplest assumption is that there are no strong or weak heads in the syntax, thus suggesting that Labeling Theory does not provide an explanation for the need for a subject to be in the traditional specifier of TP position (at least not in certain cases). I do not have a clear solution to this issue (which is the long-debated problem of the EPP), but one possibility is that the requirement for an overt subject in languages such as English is primarily a constraint on Spell-Out.

Richards (2016) develops what he calls Contiguity Theory, which takes the position that movement operations can be influenced by phonological structures, so that syntax and phonology are heavily connected. Richards proposes that in English, T is a suffix that must follow a metrical boundary, and a subject in the TP provides this metrical boundary. A metrical boundary is the edge of a metrical foot, where a foot contains one or more syllables, one of which receives more stress than the others. In a language such as Spanish, which does not require an overt subject, the vowel that precedes a tense morpheme is stressed, and thus the syllable before the Tense morpheme follows a metrical boundary. In Spanish, a metrical boundary can occur within a word. For example, in (35)a–b, the boldfaced tense morphemes follow a metrical boundary at the end of the verbal root.

(35) a. cant**á**-is (Spanish)
sing Fut Past 2pl
'you (pl.) sing'

b. canta-**rí** -a -is (Spanish)
sing Fut Past 2pl
'you (pl.) would sing (conditional)' (Richards, 2016, p. 12)

Richards (2016, p. 15) writes that in languages such as English, "metrical boundaries occur only on complete words, which are in turn found in specifiers." For example, in (36), there is supposedly a metrical boundary at the edge of the subject *there*, which precedes the verb containing the tense morpheme suffix *-ed*.

(36) **There** arrived a man. (Richards, 2016, p. 22)

Note that Richards argues that phonological constraints have specific effects on a syntactic derivation, and not necessarily only at Spell-Out, writing "the narrow syntax can make reference to, for instance, metrical boundaries" (Richards, 2016, p. 27). Consider how this approach can deal with examples in which T does not overtly follow a specifier at Spell-Out. In (37)a–b, I assume that T raises at Spell-Out. But in the syntactic structure, T still follows the subject, which has a metrical boundary. If the requirement for the affix T to follow a metrical boundary applies at the level of narrow syntax, then these can be accounted for.

(37) a. Will Tom read a book?
b. Does Tom read a book?
c. Who did John say will buy the book? (Pesetsky & Torrego, 2001, p. 371)

Contiguity Theory, as developed by Richards (2016) is complex, and further examination of whether or not it can truly account for EPP effects is warranted, but it may be a promising approach.[31]

31) Contiguity Theory, as developed in Richards (2016) is not necessarily compatible with the model developed in this paper. For example, it makes extensive use of head movement and phrasal movement that my model would consider to be suspect. In addition, the requirement that T follow a metrical boundary in English does not directly rule out a non-argument from appearing in the typical specifier of TP position. Thus, (i)-(ii) are not directly ruled out, although I note that these sound better to my ears than their counterparts without the adverbials; in particular (ii) sounds like it might be possible (to my ears).

(i) *John expects **quickly** to arrive Mary.

(ii) *?**Quickly** was read the book.

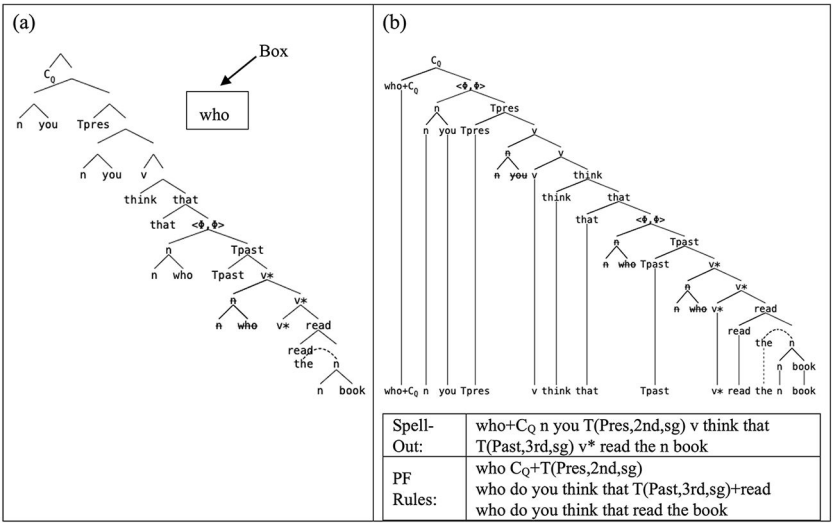
Richards is able to limit the initial specifier position to arguments in English, thus ruling out examples such as (i)–(ii), by means of complex proposals that the English affix T must follow a metrical boundary and that T must be in the same prosodic domain as a goal that it Agrees with. T is a probe and a subject is its goal. If I understand correctly, if the subject does not move to the specifier of the TP then it will be in a different prosodic domain from its probe. If the subject moves to the specifier of TP, then both the probe and goal are in the same prosodic domain and T follows a metrical boundary created by the subject. The notion that phonological constraints can account for perplexing properties of language, such as the EPP property, may be promising, but the complexity of the Contiguity

To summarize, my model potentially predicts that an argument in an unaccusative or passive construction need not move, contrary to fact. I suggested that there might be a Spell-Out based solution, but if not, then the general problem of the EPP still remains. I leave an in-depth analysis of this for future work. I also note that in the vast majority of cases, my model correctly generates the target Spell-Out.

Another remaining issue with my model is that the interaction between Free Merge and Labeling cannot account for the *that*-trace effect. Consider the sentence with the *that*-trace effect in Figure 28. When the embedded CP is completed, the *wh*-phrase *who* is in the Box. In this example, *who* and the Tpast phrase form a labelable {XP, YP} structure. When the matrix C_Q is Merged, C_Q accesses the box and the wh-phrase is pronounced in clause initial position.[32] Thus, this is predicted to be well-formed, contrary to fact.

Figure 28

**Who Do You Think That Read the Book?*



Theory Approach is potentially problematic from the perspective of the Strong Minimalist Thesis. Further analysis is required.

32) Chomsky (2015) develops an analysis of the *that*-trace effect. See Ginsburg (2016) for discussion of potential problems with that analysis. If insertion into the Box requires IM, then it might be possible to claim that *who* undergoes IM to the CP phase edge and then enters the Box, after which it is invisible to Labeling, which creates a problem for Labeling of the TP. Although, as noted above, requiring IM for insertion into the Box seems superfluous. Also, following Chomsky (2015), if one assumes that a root is weak and requires Labeling via an {XP, YP} structure (a position which I do not take), then the question arises of how a wh-phrasal object and a verbal root are labeled after the wh-phrasal object is inserted into the Box. If an object and a root label before the object is inserted into the Box, then it isn't clear why Labeling couldn't also occur before a wh-subject is inserted into the Box.

I suggest that the ill-formedness of *that*-trace effect constructions may have to do with extra-syntactic factors applying at Spell-Out. First of all, when *that* is not pronounced, the *that*-trace effect goes away. Chomsky (2015) relies on dephasing to account for this (in the absence of *that*, the embedded CP is no longer a phase), but dephasing is an extraordinarily complex process that also violates the No Tampering Condition. A simpler assumption is that the *that*-trace effect is simply dependent on whether or not C is pronounced overtly. A promising possibility is that the *that*-trace effect is not syntactic, but has to do with phonological factors. Sato and Dobashi (2016) propose that the *that*-trace effect results from constraints on prosodic phrasing. They propose that there is a "PF condition" that "[f]unction words cannot form a prosodic phrase on their own (2016, p. 1)", and that when *that* is followed by a trace, *that* ends up forming a prosodic phrase by itself, which is not permitted. When *that* is followed by a subject, as well as other types of phrases such as adverbials, it does not form a prosodic phrase on its own, and there is no problem. It is also notable that the *that*-trace effect is not found in certain English dialects as well as in many other languages. For example, Sobin (1987) points out that some English speakers do not find some *that*-trace constructions to be ill-formed. This suggests that the cause of the *that*-trace effect may not be syntactic in nature. If the *that*-trace effect lacks a syntactic cause, then it needs to be accounted for at Spell-Out. I leave in-depth examination of this issue for future work.

6 Conclusion

In this paper, I have discussed Free Merge as implemented by a computer model. I presented the basic components of this model, which attempts to take a "simple" approach (although not necessarily as simple as possible) to language generation, dispensing with complex mechanisms. I have shown that in general, Labeling, combined with phase boundaries, is sufficient to constrain Free Merge. Also, note that Theta Theory and Case Theory play no clear role in ruling out derivations, and Labeling alone is generally sufficient to constrain Free Merge.

There is a certain amount of overgeneration that is a potential problem. In order to deal with overgeneration, I needed to propose the Unique Feature Rule and No Successive IM. If these are truly principles of language, they require further examination. Overgeneration of ill-formed structures ideally should not occur or should be severely limited, probably more so than presented in this paper. Furthermore, overgeneration of well-formed structures is an issue, but the potentially problematic examples discussed in this paper can possibly be eliminated at Spell-Out. The beauty of Free Merge is that IM requires no trigger, and a variety of attested IM operations fall out from the model. Note, however, that feature-driven IM has the advantage of doing away with this overgeneration problem. On the other hand, feature-driven Merge is complicated by the

need for a variety of features to trigger IM. Whether or not feature-driven Merge should truly be eliminated from the theory requires further examination.

Funding: This work was supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (C), Grants #20K00664 and #24K03964.

Acknowledgments: I would like to thank Kleanthes K. Grohmann and two anonymous reviewers for their extremely helpful comments. I would also like to thank Hiroshi Terada and Sandiway Fong for helpful comments and discussion. All errors are my own.

Competing Interests: The author has declared that no competing interests exist.

Author Note: This paper is an extensively-revised and expanded version of Ginsburg (2022), a proceedings paper from the *Joint Conference on Language Evolution (JCole) Kanazawa, Japan 2022.*
Trees for this paper were created with a Tree Drawing Program that I created. This program runs directly in the browser and it is available for anyone to use. Note that there are bugs.
Tree Drawing Program: https://ginsburg-lab.h.kyoto-u.ac.jp/JGTreeDrawingProgram.html

Data Availability: This paper discusses derivations that were modeled with a computer program. The complete derivations that are generated by this computer program are available in the Supplementary Appendix (see Ginsburg, 2024).

Supplementary Materials

The Supplementary Materials consists of webpages (see Ginsburg, 2024) that display:

- information about how the computer model that the author used works
- the complete derivations that were produced by the computer model that is presented in this paper

Index of Supplementary Materials

Ginsburg, J. (2024). *Appendix to "Constraining Free Merge"* [Webpages]. https://ginsburg-lab.h.kyoto-u.ac.jp/JG2024Appendix

References

Abney, S. (1987). *The English noun phrase in its sentential aspect* [Doctoral dissertation]. Massachusetts Institute of Technology.

Ackema, P., & Neeleman, A. (2001). *Context-sensitive spell-out and adjacency* [Unpublished manuscript]. Utrecht University and University College London.

Baker, M. (1988). *Incorporation.* Chicago University Press.

Berwick, R. (2011). All you need is Merge: Biology, computation, and language from the bottom up. In A. M. Di Sciullo & C. Boeckx (Eds.), *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty* (pp. 461–491). Oxford University Press.

Blümel, A. (2018). Q-float in West Ulster English and labeling. *Yearbook of the Poznan Linguistic Meeting, 4*(1), 55–73.

Blümel, A., & Holler, A. (2021). DP, NP, or neither? Contours of an unresolved debate. *Glossa: A Journal of General Linguistics, 7*(1), Article 153. https://doi.org/10.16995/glossa.8326

Bobaljik, J. D. (2008). Where's phi? Agreement as a postsyntactic operation. In D. Harbour, D. Adger & S. Béjar (Eds.), *Phi theory: Phi features across modules and interfaces* (pp. 295–328). Oxford University Press. https://doi.org/10.1093/oso/9780199213764.003.0010

Boeckx, C., & Grohmann, K. K. (Eds.) (2003). *Multiple wh-fronting*. John Benjamins.

Boeckx, C., & Grohmann, K. K. (2007). Putting phases in perspective. *Syntax, 10*(2), 204–222. https://doi.org/10.1111/j.1467-9612.2007.00098.x

Borer, H. (2005a). *In name only* (Structuring sense, Vol. 1). Oxford University Press.

Borer, H. (2005b). *The normal course of events* (Structuring sense, Vol. 2*)*. Oxford University Press.

Borer, H. (2013). *Taking form* (Structuring sense, Vol. 3). Oxford University Press.

Bošković, Ž. (2002). On multiple wh-fronting. *Linguistic Inquiry, 33*(3), 351–383. https://doi.org/10.1162/002438902760168536

Bruening, B. (2009). Selectional asymmetries between CP and DP suggest that the DP hypothesis is wrong. *University of Pennsylvania Working Papers in Linguistics, 15*(1), 27–35. https://repository.upenn.edu/pwpl/vol15/iss1/5

Bruening, B. (2020). The head of the nominal is N, not D: N-to-D Movement, Hybrid Agreement, and conventionalized expressions. *Glossa: A Journal of General Linguistics, 5*(1), Article 15. https://doi.org/10.5334/gjgl.1031

Bruening, B., Dinh, X., & Kim, L. (2018). Selection, idioms, and the structure of nominal phrases with and without classifiers. *Glossa: A Journal of General Linguistics, 3*(1), Article 42. https://doi.org/10.5334/gjgl.288

Carstens, V. (2003). Rethinking complementizer agreement: Agree with a case-checked goal. *Linguistic Inquiry, 34*(3), 393–412. https://doi.org/10.1162/002438903322247533

Cecchetto, C., & Donati, C. (2015). *(Re)Labeling*. MIT Press.

Chen, V. (2018). The raising-to-object construction in Puyuma and its implications for a typology of RTO. *Glossa: A Journal of General Linguistics, 3*(1), Article 111. https://doi.org/10.5334/gjgl.423

Cheng, L. (2000). Moving just the feature. In U. Lutz, G. Müller, & A. von Stechow (Eds.), *Wh-scope marking* (pp. 77–99). John Benjamins.

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1981). *Lectures on government and binding*. Foris.

Chomsky, N. (1986). *Knowledge of language: Its nature, origins, and use*. Praeger.

Chomsky, N. (1993). A Minimalist Program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 1–52). MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on Minimalist syntax in honor of Howard Lasnik* (pp. 89–155). MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (Vol. 3, pp. 104–131). Oxford University Press.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland, & H. Gärtner (Eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the view from syntax-semantics* (pp. 1–30). Mouton de Gruyter.

Chomsky, N. (2008). On phases. In R. Freidin, C. P. Otero, and M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud* (pp. 133–166). MIT Press.

Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language? In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 45–62). Cambridge University Press.

Chomsky, N. (2013). Problems of projection. *Lingua, 130*, 33–49. https://doi.org/10.1016/j.lingua.2012.12.003

Chomsky, N. (2015). Problems of projections: Extensions. In E. Di Domenico, C. Hamann, & S. Matteini (Eds.), *Structures, strategies and beyond: Studies in honour of Adriana Belletti* (pp. 1–16). John Benjamins. https://doi.org/10.1075/la.223.01cho

Chomsky, N. (2021). Minimalism: Where are we now, and where can we hope to go. *Gengo Kenkyu, 160*, 1–41. https://doi.org/10.11435/gengo.160.0_1

Chomsky, N. (2024). The miracle creed and SMT. In M. Greco & D. Mocci (Eds.), *A Cartesian dream: A geometrical account of syntax: In honor of Andrea Moro* (pp. 17–40). Lingbuzz Press.

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic Inquiry, 8*(3), 425–504. https://www.jstor.org/stable/4177996

Chomsky, N., Seely, T. D., Berwick, R. C., Fong, S., Huybregts, M. A. C., Kitahara, H., McInnerney, A., & Sugimoto, Y. (2023). *Merge and the Strong Minimalist Thesis*. Cambridge University Press. https://doi.org/10.1017/9781009343244

Cole, P., & Hermon, G. (2000). Partial wh-movement: Evidence from Malay. In U. Lutz, G. Müller, & A. von Stechow (Eds.), *Wh-scope marking* (pp. 101–130). John Benjamins. https://doi.org/10.1075/la.37.05col

Danon, G. (2011). Agreement and DP-internal feature distribution. *Syntax, 14*(4), 297–317. https://doi.org/10.1111/j.1467-9612.2011.00154.x

Dékány, É. (2018). Approaches to head movement: A critical assessment. *Glossa: A Journal of General Linguistics, 3*(1), Article 65. https://doi.org/10.5334/gjgl.316

Donati, C., & Cecchetto, C. (2011). Relabeling heads: A unified account for relativization structures. *Linguistic Inquiry, 42*(4), 519–560. https://doi.org/10.1162/LING_a_00060

Embick, D., & Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry, 39*(1), 1–53. https://doi.org/10.1162/ling.2008.39.1.1

Embick, D., & Noyer, R. (2001). Movement operations after syntax. *Linguistic Inquiry, 32*(4), 555–595. https://doi.org/10.1162/002438901753373005

Epstein, S. D., Kitahara, H., & Seely, T. 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. https://doi.org/10.1162/LING_a_00163

Epstein, S. D., Kitahara, H., & Seely, T. D. (2016). Phase cancellation by external pair-merge of heads. *Linguistic Review, 33*(1), 87–102. https://doi.org/10.1515/tlr-2015-0015

Epstein, S. D., Kitahara, H., & Seely, T. D. (2022). A simpler solution to two problems revealed about the composite operation Agree. In S. D. Epstein, H. Kitahara, & T. D. Seely (Eds.), *A Minimalist theory of simplest Merge* (pp. 111–115). Routledge.

Fanselow, G., & Ćavar, D. (2002). Distributed deletion. In A. Alexiadou (Ed.), *Theoretical approaches to universals* (pp. 65–107). John Benjamins.

Fong, S., & Ginsburg, J. (2014). A new approach to tough-constructions. In R. E. Santana-LaBarge (Ed.)., *Proceedings of the 31st West Coast Conference on Formal Linguistics* (pp. 180–188). Cascadilla Proceedings Project.

Fong, S., & Ginsburg, J. (2019). Towards a Minimalist Machine. In R. E. Berwick & E. P. Stabler (Eds.). *Minimalist parsing* (pp. 16–38). Oxford University Press. https://doi.org/10.1093/oso/9780198795087.003.0002

Fong, S., & Ginsburg, J. (2023). On the computational modeling of English relative clauses. *Open Linguistics, 9*(1), Article 20220246. https://doi.org/10.1515/opli-2022-0246

Gallego, Á. J. (2010). *Phase theory*. John Benjamins.

Gallego, Á. J. (2017). Remark on the EPP in Labeling Theory: Evidence from Romance. *Syntax, 20*(4), 384–399. https://doi.org/10.1111/synt.12139

Georgi, D., & Müller, G. (2010). Noun-phrase structure by reprojection. *Syntax, 13*(1), 1–36. https://doi.org/10.1111/j.1467-9612.2009.00132.x

Ginsburg, J. (2016). Modeling of problems of projection: A non-countercyclic approach. *Glossa: A Journal of General Linguistics, 1*(1), Article 7. https://doi.org/10.5334/gjgl.22

Ginsburg, J. (2022). Constraining free Merge: Labeling and the theta-criterion. In A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, Y. Jadoul, M. Martins, Y. Oseki, E. D. Rodrigues, O. Vasileva, & S. Wacewicz (Eds.), *Proceedings of the Joint Conference on Language Evolution (JCole) Kanazawa, Japan 2022*, 237–244.

Ginsburg, J., & Fong, S. (2018, June 10). *On constraining Free Merg*e. 43rd Meeting of the Kansai Linguistic Society. Konan University, Kobe, Japan. https://ginsburg-lab.h.kyoto-u.ac.jp/WebPresentations/KLS43Pres-vers7.pdf

Ginsburg, J., & Fong, S. (2019). Combining linguistic theories in a Minimalist Machine. In R. E. Berwick & E. P. Stabler (Eds.), *Minimalist parsing* (pp. 39–68). Oxford University Press. https://doi.org/10.1093/oso/9780198795087.003.0003

Goto, N. (2017). Eliminating the strong/weak parameter on T. In M. Y. Erlewine (Ed.), *Proceedings of GLOW in Asia XI, vol. 2* (pp. 57–71). MIT Working Papers in Linguistics.

Groothuis, K. A. (2015). *The inflected infinitive in romance* [Master's thesis, Leiden University]. https://studenttheses.universiteitleiden.nl/access/item%3A2605923/view

Hale, K., & Keyser, S. J. (2002). *Prolegomenon to a theory of argument structure*. MIT Press.

Harizanov, B., & Gribanova, V. (2019). Whither head movement? *Natural Language and Linguistic Theory, 37*, 461–522. https://doi.org/10.1007/s11049-018-9420-5

Harley, H. (2004). Merge, conflation, and head movement: The First Sister Principle revisited. *Proceedings of the North East Linguistic Society, 34*, 239–254.

Jónsson, J. G. (1996). *Clausal architecture and case in Icelandic* [Doctoral dissertation]. University of Massachusetts.

Kishimoto, H. (2021). ECM subjects in Japanese. *Journal of East Asian Linguistics, 30*, 231–276. https://doi.org/10.1007/s10831-021-09226-y

Koppen, M. v. (2005). *One probe – two goals: Aspects of agreement in Dutch dialects*. [Doctoral dissertation, Leiden University]. https://www.lotpublications.nl/Documents/105_fulltext.pdf

Koppen, M. v. (2017). Complementizer agreement. In M. Everaert & H. Van Riemsdijk (Eds.), *The Wiley Blackwell companion to syntax* (2nd ed., pp. 923–962). John Wiley & Sons. https://doi.org/10.1002/9781118358733.wbsyncom061

Kuno, S. (1976). Subject raising. In M. Shibatani (Ed.), *Syntax and semantics 5: Japanese generative grammar* (pp. 17–49). Academic Press.

Landau, I. (2024). Empirical challenges to the Form-Copy theory of Control. *Glossa: A Journal of General Linguistics, 9*(1), 1–40. https://doi.org/10.16995/glossa.16406

Lasnik, H., & Saito, M. (1991). On the subject of infinitives. In L. M. Dobrin, L. Nichols, & R. M. Rodriquez (Eds.), *Papers from the 27th regional meeting of the Chicago Linguistic Society 1991* (pp. 324–343). Chicago Linguistic Society.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, & A. Williams (Eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium, University of Pennsylvania Working Papers in Linguistics* 4.2 (pp. 201–225). Penn Working Papers in Linguistics.

Marantz, A. (2000). Case and licensing. In E. J. Reuland (Ed,), *Arguments and case: Explaining Burzio's Generalization* (pp. 11–30). John Benjamins.

Matasović, R. (2018). *An areal typology of agreement systems*. Cambridge University Press.

Matushansky, O. (2006). Head movement in linguistic theory. *Linguistic Inquiry, 37*(1), 69–109. https://doi.org/10.1162/002438906775321184

McCloskey, J. (1979). *Transformational syntax and model theoretic semantics: A case study in Modern Irish*. Reidel.

McCloskey, J. (2000). Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry, 31*(1), 57–84. https://doi.org/10.1162/002438900554299

McCloskey, J. (2001). The morphosyntax of wh-extraction in Irish. *Journal of Linguistics, 37*(1), 67–100. https://doi.org/10.1017/S0022226701008775

McDaniel, D. (1989). Partial and multiple wh-movement. *Natural Language and Linguistic Theory, 7*, 565–604. https://doi.org/10.1007/BF00205158

Miyagawa, S. (2005). On the EPP. In M. McGinnis & N. Richards (Eds.), *Perspectives on phases* (pp. 201–236). MIT Working Papers in Linguistics.

Mizuguchi, M. (2017). Labelability and interpretability. *Studies in Generative Grammar, 27*(2), 327–365. https://doi.org/10.15860/sigg.27.2.201705.327

Müller, G. (2004). Phrase impenetrability and wh-intervention. In A. Stepanov, G. Fanselow, & R.Vogel (Eds.), *Minimality effects in syntax* (pp. 289–326). Mouton de Gruyter. https://doi.org/10.1515/9783110197365

Oishi, M. (2015). The hunt for a label. In H. Egashira, H. Kitahara, K. Nakazawa, T. Nomura, M. Oishi, A. Saizen, & M. Suzuki (Eds.), *Untiring pursuit of better alternatives* (pp. 222–334). Kaitakusha.

Pesetsky, D., & Torrego, E. (2001). T-to-C movement: Causes and consequences. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 355–426). MIT Press.

Pesetsky, D., & Torrego, E. (2011). Case. In C. Boeckx (Ed.), *The Oxford handbook of linguistic Minimalism* (pp. 52–72). Oxford University Press.

Pires, A. (2006). *The Minimalist syntax of defective domains: Gerunds and infinitives*. John Benjamins.

Platzack, C. (2013). Head movement as a phonological operation. In L. L. Cheng & Cover, N. (Eds), *Diagnosing syntax* (pp. 21–43). Oxford University Press.

Postal, P. M. (1974). *On raising: One rule of English grammar and its theoretical implications*. MIT Press.

Radford, A. (2016). *Analyzing English sentences* (2nd ed.). Cambridge University Press.

Raposo, E. (1987). Case theory and Infl-to-Comp: The inflected infinitive in European Portuguese. *Linguistic Inquiry, 18*(1), 85–109. https://www.jstor.org/stable/4178525

Richards, M. D. (2006). Object shift, phases, and transitive expletive constructions in Germanic. *Linguistics Variation Yearbook, 6*(1), 139–159. https://doi.org/10.1075/livy.6.07ric

Richards, M. D. (2007). On feature inheritance: An argument from the Phase Impenetrability Condition. *Linguistic Inquiry, 38*(3), 563–572. https://doi.org/10.1162/ling.2007.38.3.563

Richards, M. D. (2011). Deriving the edge: What's in a phase? *Syntax, 14*(1), 74–95. https://doi.org/10.1111/j.1467-9612.2010.00146.x

Richards, N. (2016). *Contiguity theory*. MIT Press.

Roberts, I. (2010). *Agreement and head movement: Clitics, incorporation, and defective goals*. MIT Press.

Roberts, I. (2011). Head movement and the minimalist program. In C. Boeckx (Ed.), *The Oxford handbook of linguistic Minimalism* (pp. 195–219). Oxford University Press.

Rudin, C. (1988). On multiple questions and multiple *wh*-fronting. *Natural Language and Linguistic Theory, 6*, 445–501. https://doi.org/10.1007/BF00134489

Sato, Y., & Dobashi, Y. (2016). Prosodic phrasing and the that-trace effect. *Linguistic Inquiry, 47*(2), 333–349. https://doi.org/10.1162/LING_a_00213

Scida, E. (2004). *The inflected infinitive in Romance languages*. Routledge.

Sigurðsson, H. A. (2006). The nominative puzzle and the low nominative hypothesis. *Linguistic Inquiry, 37*(2), 289–308. https://doi.org/10.1162/ling.2006.37.2.289

Sigurðsson, H. A., & Holmberg, A. (2008). Icelandic dative intervention: Person and number are separate probes. In R. D'Alessandro, S. Fischer, & G. H. Hrafnbjargarson (Eds.), *Agreement restrictions* (pp. 251–280). De Gruyter Mouton. https://doi.org/10.1515/9783110207835.251

Sobin, N. (1987). The variable status of Comp-Trace phenomena. *Natural Language and Linguistic Theory, 5*(1), 33–60. https://doi.org/10.1007/BF00161867

Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry, 19*(3), 425–449. https://www.jstor.org/stable/25164903

Stabler, E. P. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics: Lecture notes in computer science* (pp. 68–95). Springer. https://doi.org/10.1007/BFb0052152

Stabler, E. P. (2011). Computational perspectives on Minimalism. In C. Boeckx (Ed.), *Oxford handbook of linguistic minimalism* (pp. 617–643). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199549368.013.0027

Stowell, T. A. (1981). *Origins of phrase structure* [Doctoral dissertation]. Massachusetts Institute of Technology.

Stroik, T. S. (2009). *Locality in Minimalist syntax*. MIT Press.

Tanaka, H. (2002). Raising to object out of CP. *Linguistic Inquiry, 33*(4), 637–652. https://doi.org/10.1162/002438902762731790

Tóth, I. C. (2000). *Inflected infinitives in Hungarian* [Doctoral dissertation, Tilburg University]. https://pure.uvt.nl/ws/portalfiles/portal/394495/84846.pdf

Turano, G. (1998). Overt and covert dependencies in Albanian. *Studia Linguistica, 52*(2), 149–183. https://doi.org/10.1111/1467-9582.00032

Van Eynde, F. (2006). NP-internal agreement and the structure of the noun phrase. *Journal of Linguistics, 42*(1), 139–186. https://doi.org/10.1017/S0022226705003713