Underspecification and (Im)possible Derivations: Toward a Restrictive Theory of Grammar

by

Yushi Sugimoto

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2022

Doctoral Committee:

Professor Marlyse Baptista, Co-Chair
Professor Acrisio Pires, Co-Chair
Professor Richard Lewis
Professor Terje Lohndal, Norwegian University of Science and Technology

Yushi Sugimoto

yushis@umich.edu

ORCID iD: 0000-0002-7980-5770

ACKNOWLEDGEMENTS

The work that I've done in this thesis is based on my work at University of Michigan for four years. Of course, the results that I achieved in my work could not exist without support from various people at various stages in my academic life.

I would like to express my gratitude to my dissertation committee members of my dissertation: Marlyse Baptista, Acrisio Pires, Richard Lewis, and Terje Lohndal. First of all, I'm indebted to the co-chairs of my committee members, Marlyse and Acrisio. Since Sam's loss, Marlyse and Acrisio have been encouraging me to continue studying linguistics and I couldn't have completed my dissertation without their support.

Acrisio has been generous and patient in listening to my premature analyses and has given me valuable and detailed comments and suggestions on my drafts. Meeting with him and discussing the issues of syntax always stimulates my curiosity. I also should thank him for helping me submitting applications for grants and deal with all kinds of administrative issues through my PhD life at UMICH.

I vividly remember when I first took Marlyse's class. I was a naive first year Ph.D. student who didn't have any idea what the class was all about. At the end of the semester I wrote a term paper, which evolved into a full-fledged chapter in this thesis, a part of which is published as a journal article. It was impossible to complete my work without her constant support, encouragement and without her being patient whenever I said "I don't have empirical evidence!," which always happens whenever someone has a meeting with me. This project and conversation with Marlyse opened my eyes to multilingualism and Creole studies. I would be remiss not to say thank you to Marlyse.

I'm also deeply indebted to Richard Lewis and Terje Lohndal. Rick always has a very clear scientific approach to any fields in cognitive science (at least that's my impression). His comments are always helpful in clarifying what elements are lacking in the ideas and theories that I'm developing. I'm also very grateful for being able to take his graduate class (COGSCI700) and to teach COGSCI200 together. Rick is my ideal teacher; unfortunately, I cannot be like Rick, but I hope I can be a 'best teacher' for my future students. Terje was also a key person to my dissertation. The first time we met was 2015 when I was a masters student at Sophia University. He came to give a talk at Sophia University and my previous advisor gave me an opportunity to talk to Terje. I presented the essence of my MA thesis, and since then we've kept in touch and he's always given me valuable comments, which improve each of my papers.

I would like to talk a bit about the late Sam Epstein, who was my former advisor and mentor at UMICH, and who was the reason I came to UMICH. When I first contacted him, I was about to finish my masters thesis, and his reply was full of excitement about my paper and encouragement to continue to pursue studying theoretical issues in syntax. I wish I could have had more time to work with him, but his spirit is with me; whenever I come up with an idea, I always think of Sam and imagine what he would say about it. In addition to Sam, I would also like to thank Hisatsugu Kitahara and Daniel Seely, who worked with Sam for a long time, and who support my life at UMICH.

Things went drastically in a direction that I didn't expect after I finished my first year at UMICH, wtih Sam's loss and COVID. These challenges gave me time to think about my future. COVID has been making everyone's life hard, but one good thing is that it makes it easier to have a conversation with people outside UMICH and outside the United States via online meetings. Conversation with Nobu Goto and Andreas Blümel for hours for each meeting cheered me up and gave me some hope of keeping studying syntax. Thank you, Nobu and Andreas! I also want to thank the members of "Yavneh" (Noam Chomsky, Bob Berwick, Sandiway Fong, Riny Huybregts, Hisa Kitahara, Andrew McInnerney and Daniel Seely), where we have had meetings each month to discuss the current state-of-the-art mechanisms of syntax. I also would like to express thanks to Alan Ke and Andrew McInnerney. We started a weekly meeting in 2017, if I recall correctly, and since then we've kept in touch and discussed various types of topics in syntax and sentence processing. I was really lucky to have colleagues who are interested in theoretical aspects of syntax, which usually doesn't happen to me so often.

I want to thank other faculty members in the department of Linguistics at UMICH. My thanks go to Natasha Abner, Steve Abney, Emily Atkinson, Pam Beddor, Julie Boland, Jon Brennan, Andries Coetzee, Ezra Keshet, Elaine McNulty, Savi Namboodiripad, Robin Queen, and Rich Thomason for their teaching, advice, and general support.

My thanks should also go to my colleagues at UMICH those who have helped me survive: Aliaksei Akimenka, Danuta Allen, Wyatt Barnes, Marcus Berger, Felicia Bisnath, Rawan Bonais, Dominique Bouavichith, Danielle Burgess, Dominique Canning, Lauretta Cheng, Yu-Chuan Chiang, Jeonghwa Cho, Justin Craft, Sophia Eakins, Wilkinson Daniel Wong Gonzales, Marjorie Herbert, Tamarae Hildebrandt, Demet Kayabaşi, Mathew Alex Kramer, Chia-Wen Lo, Carrie Ann Morgan, L.R. Nikolai, Joy Peltier, Moira Salzman, Yourdanis Sedarous, Cecilia Solís-Barroso, Csilla Tatár, Tzu-Yun Tung, Kelly E. Wright, and Jian Zhu.

I would like to thank Will Nediger for the proofreading of this dissertation.

For my linguistics training in Japanese, I thank the following professors and colleagues for their encouragement and all kinds of support: Naoki Fukui, Norimasa Hayashi, Toru Ishii, Takaomi Kato, Ryoichiro Kobayashi, Roger Martin, Takashi Munakata, Takanobu Nakamura, Takanori Nakashima, Hiroki Narita, Miki Obata, Jun Omune, Yohei Oseki and Masanobu Sorida.

Last, but not least, I would like to thank my family, Mutsumi Sugimoto, Mariko Sugimoto, and Sayaka Sugimoto, for their support. Without their mental and financial support, I wouldn't even exist as a linguist now.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

ABSTRACT

This dissertation develops two main ideas: (i) the range of possible human languages may be partially limited by the possible combinations of functional features, and (ii) Universal Grammar (UG) contains linguistic features and operations whose application is not fully specified. Specifically, some underspecified components are allowed under a restrictive theory of grammar in Minimalism.

In order to convey these ideas, this dissertation consists of three parts. First, we need syntax to construct syntactic structures which are an optimal solution to interfaces/language-external systems (Chomsky, 2000, 2005), though the syntax itself does not have to be involved in the evaluation process to determine whether the sentence is acceptable or not. The derivations that are generated by MERGE, a syntactic structure-building operation (Chomsky, 2021a,b) are, by definition, grammatical, though acceptability is affected by other factors such as working memory, sentence processing difficulty, and semantic-pragmatic factors. Based on Chomsky (2021b), Epstein et al. (2018a), Epstein et al. (2021), and Kitahara & Seely (2021), I focus on (im)possible derivations such as the cases where internal Merge and parallel Merge have been argued to apply. Furthermore, I reanalyze constraints on movement such as remnant movement, improper movement, the freezing effect, the subject condition and the adjunct condition.

The second part of the dissertation is about the underspecification of rule ordering in narrow syntax (Obata et al., 2015). The operations in narrow syntax are MERGE (external Merge and internal Merge/movement), Agree, feature inheri-

tance, Labeling, and Transfer. Obata et al. (2015) analyze complementizer agreement with *wh*-phrases in Haitian Creole and Cabo Verdean Creole. The framework they adopt (Chomsky 2013) cannot hold under MERGE (Chomsky, 2021a,b). I propose alternative, cyclic derivations for these phenomena. I also analyze additional complementizer patterns in Brazilian Portuguese. In addition, I propose that variation in how functional categories are introduced in the narrow syntax can be different from language to language, yielding different head amalgams, which can account for a cluster effect without invoking the notion of (macro-)parameters.

The last part of this dissertation focuses on the possible combinations of functional features in Creoles, where I argue that the underspecification is also in the lexicon (i.e. regarding number features, person feature, etc). In order to argue this point, I adopt a late-insertion-based exoskeletal model of language mixing (Grimstad et al., 2018; Riksem et al., 2019) and propose that this model can capture creole languages as well. Based on this model, I argue that feature recombination (Aboh, 2009, 2015, 2019, 2020) applies to the functional features in Creole languages as an underlying principle of Creole genesis. I apply this model to the analysis of the anterior marker *-ba* in Cabo Verdean Creole (CVC), nominal structures in Saramaccan, and a complementizer in CVC.

CHAPTER I

Introduction

1.1 The goal of this thesis

This thesis explores a possible grammatical model that restricts possible linguistic structures. One main idea in this thesis is that Universal Grammar (UG) contains primitives and operations whose application is not fully specified. That is, some underspecified components are allowed. First, possible human languages are constrained by functional categories, but grammar may allow multiple possible combinations of formal features that functional categories consist of. Second, derivations are generated by MERGE, a syntactic structure-building operation (Chomsky, 2019a,b, 2021a,b; Chomsky et al., 2019), though rule ordering of operations in narrow syntax could be underspecified (Obata et al., 2015) as long as the derivation converges. Thus, there might be multiple optimal derivations and they could account for most linguistic variation.

There are three main proposals in this thesis. Firstly, I will investigate the nature of MERGE as proposed by Chomsky (2019a,b, 2021a,b). MERGE allows structures in the workspace where the accessible elements are restricted under the biological constraint Minimal Yield, which states that MERGE can yield only one new accessible element at once. Since the workspace is not infinite, this constraint limits the number of accessible elements, which is desirable for minimal optimal computation to satisfy descriptive adequacy. As a consequence of MERGE, I will explore a proposal that some cases of remnant movement (movement of a phrase which includes the 'trace' or a lower copy of a previously moved element within the phrase) are generable by MERGE (Epstein, Kitahara & Seely, 2018a), while other cases are ungrammatical. Furthermore, I will argue that improper movement (movement from A′-position to A-position) is in principle ungrammatical due to an independent theory, namely theta theory, which yields the A-/A′-distinction. I will also argue that the freezing effect (movement out of a moved element is blocked) and adjunct condition (extraction of adjuncts is ungrammatical) are not part of narrow syntax. That is, the freezing effect configurations are formed by MERGE and the adjunct island configuration are generable by *Form Set*, a primitive operation of MERGE, though there might be extra-syntactic factors that make the sentence degraded, allowing non-syntactic approaches that are still compatible with the proposed system.

The second part of the thesis deals with the underspecification of rule ordering in narrow syntax. The underspecification of rule ordering wasn't explicitly discussed within minimalism until Obata et al. (2015)[1]. Obata et al. (2015) proposed to derive linguistic variation within narrow syntax by appealing to the underspecification of rule ordering. Obata et al. (2015) dealt with different kinds of linguistic variation, using only syntactic operations and underspecified rule ordering. For example, linguistic variation can take place when the operations Agree and Movement apply in different orders. These two operations give us two possible orderings. Obata et al. (2015) demonstrated the difference between Haitian Creole and Cabo Verdean Creole in terms of C-agreement with the *wh*-phrase, though the framework they adopted was Chomsky (2013), which is no longer assumed. I will

[1]See Huang (1982) and Pollock (1989), which show the optionality of application of rules. See also Roeper (1999) for how the rules are fixed through language acquisition under the minimalist program framework.

implement the possible derivations in a framework that only allows cyclic derivations (Chomsky, 2015, 2020, 2021a,b) and seek rule-ordering differences among Haitian Creole, Cabo Verdean Creole and Brazilian Portuguese. I will also explore the possible combinations of functional categories among languages and derive a cluster effect without assuming macro-parameters as primitives of UG.

The last topic that I will focus on in the thesis is the formation of functional categories in Creoles. The syntactic structure is formed by the structure building operation MERGE and the syntactic configuration determines information such as argument structure, based on *a neo-constructivist approach to mono/bilingual/creole Grammars* (cf. Borer, 2003; Marantz, 1997; Lohndal, 2014; Riksem, 2018). In a language mixing in which the words are mixed within a syntactic domain such as nominal phrases, functional categories (FCs) are determined by one of the two languages and the roots are determined by the other language in some cases (cf. Grimstad et al., 2018). Even if the language has word-internal language mixing, it does not mix the properties of FCs (cf. Grimstad et al., 2018; Riksem et al., 2019). A relevant aspect of this model is its null theory approach; minimal assumptions suffice to explain language mixing as well as monolingual grammars. I will argue that, in Creole languages (unlike language mixing whose FCs are selected by one of the source languages) FCs are formed via feature recombination (cf. Aboh, 2015), resulting in the hybrid nature of FCs. This suggests that the possible syntactic structures of natural languages partially depend on the formation of FCs. This approach leads to the idea that functional categories themselves are not universal, but they can decompose into primitives and the possible combination of primitives is underspecified, in that it's possible to recombine features.

My proposals fit a minimalist architecture of grammar, as shown below in Figure 1.1. Under a restrictive theory of MERGE, the underspecification occurs in two places. When the functional categories are formed, the primitives (e.g., the

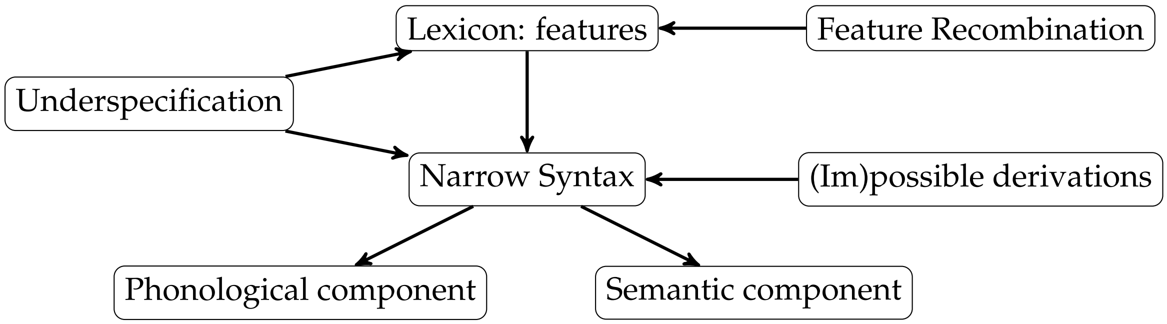


Figure 1.1: The proposed grammatical model

individual features of functional categories) are available by universal grammar, but the combination of the primitives is language-specific, in a way which will be determined through language acquisition. Another underspecification is rule ordering of syntactic operations.

Under the restricted grammatical model that I briefly described above, the components that are available by Universal Grammar are (i) primitive features that can be combined to form functional categories and (ii) syntactic operations such as MERGE. Although syntactic operations themselves are partially restricted by third factor principles (Chomsky 2005, natural law/biological principles, e.g., restrictions on processing information including language in the brain), the rest is underspecified. This will leave room for 'linguistic variation' in grammar.

In the next section, I briefly summarize what will be discussed in each chapter.

1.2 The Chapters

1.2.1 Chapter 2

Chapter 2 elucidates the (il)legitimacy of the derivations formed by the structure-building operation MERGE (Chomsky, 2021a,b) and the constraints on movement that limit the possible derivations for linguistic structures. I will explore (i) remnant movement, (ii) improper movement, and (iii) two syntactic islands (i.e., the subject island and the adjunct island) in this chapter.

The recent development of syntactic theory suggests that the structure-building operation Merge is the simplest operation, which combines two syntactic objects into a set without any labels.

(1) Merge(α,β) = {α,β} (Chomsky, 2013; Epstein et al., 2014; Collins, 2017)

Chomsky (2004) argues that this operation applies freely, either internally or externally. The properties of this version of Merge are that (i) Merge does not include linear order and (ii) Merge does not create a projection/label. Under this simplest Merge hypothesis (Merge forms a set, but not a label), Chomsky (2013) proposes that syntactic objects need to be labeled and that a labeling algorithm (which uses minimal search, a third factor principle) determines the label. The problematic structure is where two phrases are merged, namely, {$_Z$ XP, YP}, because the labeling algorithm finds two heads, namely X and Y, hence the labeling algorithm cannot determine the label for Z. There are two solutions, according to Chomsky (2013). (i) When one of the elements, say XP, moves out of Z, Y becomes the label. (ii) When X and Y have relevant shared features (e.g., ϕ-features, Q-feature), such features label the structure. This theory also implies that the unlabeled structure becomes ungrammatical.

(2) *{XP,YP} where the structure is unlabeled

More generally, (2) applies whenever there are no shared features between the two phrases. In fact, recent literature suggests that (2) is the locus of the ungrammaticality of some syntactic islands (Bošković, 2016; Goto, 2016; Richards, 2019). When it comes to languages where ϕ-agreement does not take place, it's unclear how to label the {XP,YP} structure. Saito (2016, 2018) proposes that case markers on DP/PP and inflectional elements on predicates become anti-labeling devices, which circumvent the labeling problem. Although the necessity of the label is inevitable, I will argue that a labeling analysis of movement is not sufficient to capture different types of movement, such as remnant movement, improper movement, and the freezing effect. I will adopt the more recent version of the structure-building operation MERGE to account for these types of movement.

It has been proposed by several researchers that the simplest Merge (1) can create multi-dominant structures (e.g., Citko, 2005). One might justify this as long as there is no violation of Merge. However, in Chomsky (2017c,d, 2019a,b,c, 2021a,b), he reformulates (1) since the implicit assumptions were not spelled out in (1). The version of Merge that I will introduce in this thesis is MERGE (3).

(3) MERGE

  a. WS = {P, Q, . . . }

  b. MERGE(P,Q,WS) = { {P,Q}, . . . } = WS'

MERGE applies on the *workspace* where syntactic objects and lexical items are available. MERGE proceeds with some steps which have been implicitly assumed in the literature. For example, when P and Q formed a set {P,Q}, it was assumed that the original occurrences of P and Q were deleted. But the question is, deleted from where? Another issue is how to select P and Q. What's the accessible element for Merge? Chapter 2 will discuss these details. Based on this reformulation of Merge, I will analyze different types of movement. The consequences of this version of MERGE (3) will become more clear when I introduce the principles that underlie MERGE and what are called Language Specific Conditions (LSCs, conditions that a language system has to satisfy). I will argue that the interaction between the fundamental principles and LSCs will constrain different kinds of movement. For example, MERGE rules out remnant movement, while the interaction between MERGE and Phase theory (Chomsky, 2000, 2001) allows some types of remnant movement (Epstein, Kitahara & Seely, 2018a, 2021). Another pattern has to do with the interaction between MERGE and theta theory (Chomsky, 2021a,b), where the violation of theta theory is unavoidable in the cases of improper movement, resulting in the wrong instruction being sent to the phonological component. The last part of this chapter is about what is called the CED effect (Huang, 1982), namely, that extraction out of subjects and adjuncts is banned. I will argue that this type of extraction is generable by (3), although there might be other factors that make the sentence degraded, such as semantic principles (Truswell, 2007, 2011; Ernst, 2022) and a sentence processing difficulty (cf. Culicover & Winkler, 2022).

1.2.2 Chapter 3

The aim of this chapter is to explore the possible multiple optimal derivations, which derives a part of the linguistic variation. Based on phase theory (Chomsky, 2000, 2001, 2004, 2008, 2013, 2015), it is assumed that syntactic operations take place at the phase level (except for external merge to build the syntactic structure, cf. Epstein & Shim 2015). Based on Chomsky (2008), the steps of rule application at the phase level are as follows: after the phase head is introduced, unvalued features on the phase head initiate Agree, and feature inheritance takes place, followed by movement. Transfer sends off the complement of the phase head, which renders the domain inaccessible for further computation. At Transfer, the unvalued features are valued, so the interfaces can interpret all syntactic objects, which is stated as *Full Interpretation*.

(4) Full Interpretation

All terms of a syntactic object must be interpreted, none can be ignored.

(Chomsky et al., 2019, 242)

In other words, if syntactic objects have unvalued features, for example, the derivation will be doomed to be ungrammatical. However, as long as valuation takes place properly, the rule ordering between agree, movement, and feature inheritance is, in principle, not necessarily specified in a certain way. That is, there will be multiple ways of carrying out the derivations by different rule ordering applications. Based on Obata et al. (2015), I argue for this underspecification approach in Chapter 3. The interesting cases that I will discuss are derivations where T agrees with the subject or the object. Another case is the interrogative complementizer *ki* in Haitian Creole, Cabo Verdean Creole, and *que* in Brazilian Portuguese.

A new domain of underspecification of rule ordering is also explored, namely the possible combination of functional categories. Functional categories such as *v**, T, and C (Chomsky, 2000, 2001) normally have a specifier position, where spec-head agreement is available, which indeed takes place in English. However, some languages do not show such an agreement relation. I argue that such languages introduce the functional categories as an amalgam where two or more functional categories are formed into one element before merging into narrow syntax. A consequence is that the specifier position is reduced, so that agreement is not available in a spec-head relation. This captures the absence of ϕ-agreement in Japanese, where I propose that *v**, T, and C are amalgamated before merging to the narrow syntax, whereas each functional head is introduced separately in English, so for each head, the specifier position is available for agree (e.g., spec-head agreement). I will also argue that this amalgamation approach provides a solution to the labeling problems discussed in Chapter 2 regarding Saito's (2016) analysis.

1.2.3 Chapter 4

The goal of this chapter is to further our understanding of the nature of functional features in Creoles while focusing on how the functional exponent is morphologically realized, assuming a late-insertion-based exoskeletal model in the language mixing literature (Borer, 2003, 2005a,b, 2013, 2017; Grimstad et al., 2018; Riksem et al., 2019). In language mixing, it is observed that words are mixed within a certain syntactic domain (e.g., DP-NP, VoiceP/*v*P-TP, etc.). For example, in the nominal domain, a determiner D may be from one language and N (or a stem, e.g., root + categorizer, see Alexiadou & Lohndal 2017) may originate from another language. Grimstad et al. (2018) and Riksem et al. (2019) propose that the functional projection FP intervenes between D and N, and both D and F are from one language and N from another language. The phonological exponent of the functional features (e.g., D and F) are assumed to be language specific (i.e., from one language), subject to the subset principle (Halle, 1997). Closer to the case that concerns us, Åfarli & Subbarao (2019) show that through long-term language contact, the functional exponent can be reconstituted and it can be genuinely innovative.

In this chapter I propose that functional features can themselves be recombined and that Creole languages can provide evidence for feature recombination either by virtue of their hybrid grammar (i.e., Aboh 2015) or through the congruent functional categories they display (Baptista, 2020), using a late-insertion-based exoskeletal model. That is, functional features are not individually inherited from one language or another, but can be recombined to form new functional features, allowing a novel functional exponent. To show this, we use synchronic empirical data focusing on the anterior marker *-ba* from Cabo Verdean Creole (CVC), Manjako (one of CVC's Mande substrates), and Portuguese (CVC's lexifier) to show how the recombination may operate, as CVC *-ba* recombines the features it inherited from its source languages, while innovating. I will also address nominal structures in Saramaccan and the complementizer *ki* in CVC to show novel features in these languages.

CHAPTER II

Structure Building and (Im)Possible Derivations

2.1 Introduction

In this chapter, I explore various types of movement, comparing two different theories, namely labeling theory (Chomsky, 2013, 2015) and a recent theory of structure building (i.e., MERGE Chomsky 2019a,b, 2021a,b). I argue that recent developments of the structure-building operation have consequences regarding the constraints on movement and show that labeling theory itself is not sufficient to capture some types of movement in Japanese.

Labeling theory was proposed in Chomsky (2013). Since the simplest Merge forms only an unordered set $\{\alpha,\beta\}$ where α and β are syntactic objects, the syntactic structure does not have a label/projection. In order to interpret the syntactic objects, Chomsky (2013) proposes a labeling algorithm that finds the prominent head/features. For example, in a structure {H,XP}, H is a prominent head found by minimal search, hence becomes the label. Chomsky (2013, 2015) also proposes that unlabeled structures such as {XP,YP} (where each head X and Y is ambiguously found, hence no label) become ungrammatical without additional steps to yield labeling. Based on this, several researchers have been exploring empirical issues related to locality, such as island effects (Bošković, 2016; Goto, 2016; Richards, 2019).

In this first part of this chapter, I discuss issues regarding labeling in Japanese, based on Saito (2016, 2018), and argue that we cannot capture some scrambling phenomena and (im)proper scrambling based on Saito's (2016; 2018) labeling analysis. Rather, I propose that the structure-building operation MERGE (Chomsky, 2019a,b, 2021a,b) and Phase theory (Chomsky, 2000, 2001) provide an explanation for these phenomena by means of remnant movement, based on Epstein, Kitahara & Seely (2021) and Kitahara & Seely (2021). In this framework, the structure-building operation is reformulated in a more precise way. This reformulation reveals that the only subcases of MERGE are external Merge and internal Merge, and MERGE rules out other kinds of extensions of Merge such as parallel Merge (Citko, 2005, 2011; Citko & Gračanin-Yuksek, 2021). This version of Merge restricts the accessible elements in the workspace where the syntactic objects are combined. In improper remnant movement cases, there are multiple identical accessible elements, which results in ambiguous input. MERGE does not allow this. On the other hand, in proper remnant movement, one of the multiple copies is inaccessible due to the phase impenetrability condition (Chomsky, 2000, 2001). In this chapter I also explore how this approach can capture other patterns of remnant movement in German and Japanese. I also propose that improper movement is banned by the theta theory under the MERGE framework (Chomsky, 2021a,b). Finally, I discuss other consequences of MERGE: the subject island configuration is generable by MERGE and there is also a way of generating the adjunct island configuration. The proposed approach to these two islands makes it possible to account for the variation in acceptability identified in the literature, which suggests that (un)acceptability involves multiple sources in addition to grammatical aspects.

The organization of this chapter is as follows. §2.2 introduces the labeling theory that is proposed by Chomsky (2013). §2.3 introduces Saito's (2016; 2018) labeling analysis of Japanese where feature sharing-type labeling does not take place. Although Saito's (2016; 2018) labeling analysis is plausible, I will point out that the his analysis does not seems to predict the (un)grammaticality of sentences when movement takes place, which is discussed in §2.4, where various types of movement are introduced such as improper scrambling, CP scrambling, remnant movement, and the freezing effect in Japanese. §2.5 gives an interim summary of the issues in the labeling analysis of Japanese. Since Saito's (2016; 2018) labeling analysis has issues, an alternative approach to labeling in Japanese is pursued in §2.6. Although §2.6 does not provide a solution to labeling in Japanese, it addresses the problem that the constraints on movement (e.g., improper scrambling, remnant movement, and the freezing effect) cannot be captured just by labeling theory. Instead, §2.7 introduces the MERGE framework and provides an account for constraints on movement in terms of remnant movement. In the literature on remnant movement, various analyses have been proposed. Of particular interest is a linearization approach (Nunes, 2004; Takita, 2010) to remnant movement, which is introduced in §2.7.3. In §2.7.4, I argue that this linearization approach cannot capture the asymmetry that takes place in Japanese Exceptional Case Marking (ECM) constructions where remnant movement is involved, and I argue that MERGE and Phase Theory predict the asymmetry. §2.7.5 introduces another type of a constraint on movement, that is, improper movement. It has been argued that the A′-movement followed by A-movement is banned (the ban on improper movement, May 1979; Fukui 1993a; Obata & Epstein 2011; Safir 2019 among others). Building upon the MERGE framework, I follow Chomsky's (2019b; 2021b) analysis and analyze improper movement in terms of theta theory and what is called *Form Copy*. This framework shows that the interaction between MERGE, theta theory, and *Form Copy* yield incorrect derivational steps that make sentences with improper movement ungrammatical. §2.8 shows the consequences of MERGE, and discusses subject islands and adjunct islands. I will argue that the subject island, a subcase of the freezing effect, is generable by MERGE and the adjunct island is formed by *Form Set*, a basic operation that is involved in MERGE, which makes the adjunct island transparent in some non-finite clause adjuncts. §2.9 concludes this chapter.

2.2 Labeling Theory

One of the unique properties of human language is the capacity to create an infinite number of expressions from a finite system (i.e., the human cognitive system). *Merge*, in the minimalist literature, is the structure-building operation that provides the infinite array of linguistic expressions that are interpreted at the interfaces, a semantic component (SEM) and a phonological component (PHON).

Under the minimalist program for linguistic theory (Chomsky, 1995), the deduction of the principles of Universal Grammar (UG) from what are called third factor principles (e.g., the natural laws, not specific principle to language, Chomsky 2005) has been proposed to satisfy evolvability (Chomsky, 2004, 2017b). Thus, the goal of the minimalist theory in the current development of generative grammar is to minimize the principles in UG itself. Of particular interest here is the structure-building operation Merge that yields hierarchical structures recursively, and which is implemented in UG.

Assuming that the structure-building operation Merge combines two syntactic objects into an unordered set, the simplest version of this does not yield projection or labeling of the structure, nor does it yield word order (Chomsky, 2004, 2013), as shown in (1).[1,2]

[1]See Fukui (2001); Epstein et al. (2015); Fukui & Narita (2014) for the history of the structure-building operation. See also Collins (2002); Seely (2006); Narita (2014); Collins (2017); Collins & Seely (2020) for the projection-free Merge.

[2]It is also assumed that the input of Merge is two elements, hence binary structures are assumed (Chomsky, 2013).

(1) Merge(α,β) = {α, β}

a. the set {α, β} does not have an order.

b. the set {α, β} does not have a label/projection/headedness.

In order to interpret these syntactic objects, interfaces need to know what kind of objects they are. In the original formulation of Merge in Chomsky (1995), the label was also formed by Merge as in γ in (2).

(2) $\{\gamma, \{\alpha, \beta\}\}$

The question is how to determine the label γ. In Chomsky (1995), three possibilities were listed, but Chomsky excludes the possibilities of (3a) and (3b).

(3) a. the intersection of α and β

b. the union of α and β

c. one or the other of α and β (Chomsky, 1995, 224)

Chomsky (1995) argues that (3b) indicates that the label shows a contradictory property (e.g., the structure is both verbal and nominal, etc.). As for (3a), he argues that it is irrelevant for labeling.[3]

Under the simplest Merge, Chomsky (2013) proposes that minimal search, a third factor principle (Chomsky, 2005), finds a relevant element to identify the label of the structure (this is the labeling algorithm: LA) (4).

(4) Labeling Algorithm (LA) (Chomsky, 2013, 2015)

a. {H, XP} $\rightarrow$ MS finds H; the label of this set is H.

b. {XP,YP} $\rightarrow$ MS finds X and Y; an ambiguous situation

i. {XP, YP} $\rightarrow$ $XP_i \ldots$ {t_{XP} , YP} $\rightarrow$ Y becomes the label of the set.

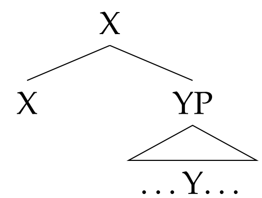
[3]See Zeijlstra (2020) for a possibility that the union of labels could be a label for a structure. As for (3a), the intersection of the labeling is similar to Chomsky's feature-sharing labeling. See the details below.

ii. {XP_F, YP_F} $\rightarrow$ X and Y share relevant features (e.g., Q-feature, ϕ-features). The label of the set becomes the pair of relevant features (e.g., $<Q,Q>$, $<\phi,\phi>$).

LA applies to syntactic structures in a top-down fashion. When LA finds a structure such as (4a), it finds the prominent element H (e.g., X in (5)).

(5) a. {X,YP}

b.

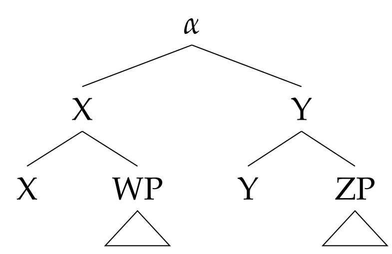


Thus, in (5), X becomes the label of the structure. Note that LA does not representationally attach the label.[4] Rather, minimal search finds and identifies the prominent element for interpretation.

The non-trivial cases are {XP, YP} structures (4b). LA finds X and Y simultaneously, and cannot determine what is the label of this structure (6).

(6) a. {XP, YP}

b.



[4]Just for expository purposes, I'll use tree notation and put the X′-style labels on the tree structures in this thesis. But see Chomsky (2015), suggesting that tree representations would lead us to misunderstand the nature of syntactic structures, as he states in his paper.

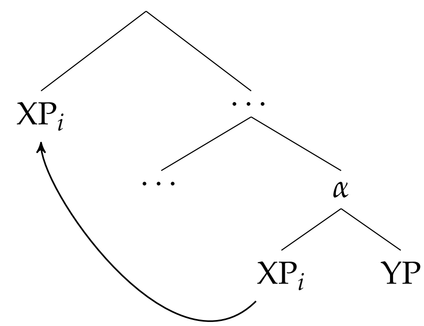
> "It is therefore advisable to abandon the familiar tree notations, which are now misleading. Thus in the description of an [XP, [YP,ZP]] structure, there is no node above either of the two merged constituents. There is no label for the root of the branching nodes (Chomsky, 2015, 6)."

Thus, we should use only set-theoretic notation. However, it is still, I believe, helpful for readers to present the tree notation.

One strategy is to move one of the elements out of the structure (4b-i). Then the head of the other phrase becomes the label of the structure, as shown in (7).

(7) a. $\{XP_i \ldots, \{XP_i, YP\}\}$

b.



The XP that moves out of the structure α leaves a copy under the copy theory of movement (cf. Chomsky, 1995; Nunes, 2004). Chomsky (2013) assumes that the lower copy becomes invisible for labeling. More precisely, some element A is in the domain D if and only if every occurrence of A is a term of D (cf. Chomsky, 2013, 44). Here let us adopt the definition of a term of from Epstein et al. (2014) (cf. Seely 2006, 201).

(8) *A term of*

a. K is a term of K.

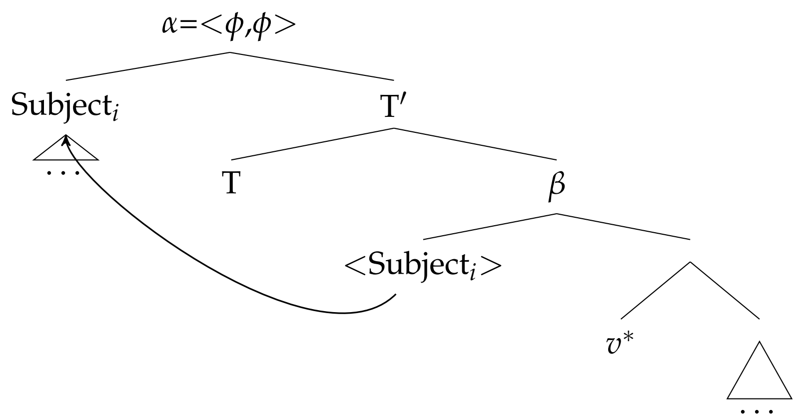
b. If L is a term of K, then the members of L are terms of K.

(Epstein et al., 2014, 466,fn11)

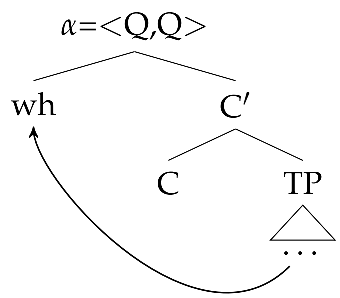
In (7), the domain of α includes only YP since α does not include every occurrence of XP (i.e., the higher copy of XP is out of the domain of α). Hence, XP in α is invisible for labeling; α is labeled by the head of YP, Y.

The other option to label unlabeled {XP,YP} structures is when X and Y share relevant features (4b-ii). For example, in a case of the spec-head configuration of the subject and T (9a), T and the subject in the [spec,TP] have the relevant ϕ-features. Then the structure α is labeled by $<\phi,\phi>$ in (9a). When the {XP, YP} structure is formed by *wh*-movement, the *wh*-phrase and C share the relevant Q-feature, so the structure will be labeled as <Q,Q> (9b), which indicates the type of the sentence (e.g., an interrogative sentence).

(9) a.



b.



Notice that in (9a), the base-generated position of the subject is also an {XP,YP} structure, hence unlabeled. Once the subject moves to the [spec,TP], β is labeled v^* since the subject has moved. Then, the subject moves to the position where feature sharing takes place. Thus, this analysis derives the EPP effect without assuming an EPP-feature.[5] As for *wh*-movement (9b), since the intermediate positions will always become unlabeled {XP, YP} structures, the edge elements have to keep moving up until they are in the configuration where relevant features are shared. Thus, successive cyclic A′-movement is explained without looking ahead in the

[5] EPP stands for Extended Projection Principle (henceforth EPP), which was originally proposed by Chomsky (1982), arguing that "clauses have subjects (Chomsky, 1982, 10)." In the minimalist literature, the [spec,TP] position in a finite clause must be filled by something (typically a subject), which is called the EPP effect. In Chomsky (2000), EPP become a feature, namely an EPP-feature that is on a head which needs a specifier. See Chomsky (2000, 2001); Bošković (2002, 2007) for relevant discussion.

derivation (cf. Chomsky 2000, 2001).

One of the general questions in labeling theory is how to label the structure for languages where agreement is not obligatory. Regarding this point, Saito (2016, 2018) argues that anti-labeling devices (e.g., case markers) play a crucial role in Japanese. Adopting Saito (2016), Miyagawa et al. (2019) argue that there are two ways of labeling: (i) labeling inducer and (ii) labeling blocker. In Japanese, the labeling blocker strategy is available. Thus, the {XP,YP} labeling problem is avoided in Japanese due to availability of case markers, for example. In the following subsection, the labeling analysis for Japanese (Saito, 2016, 2018; Miyagawa et al., 2019) is introduced and discussed in detail.

2.3 Issues Regarding Labeling in Japanese

2.3.1 Anti-labeling/labeling blocker analysis

The Labeling Algorithm in Chomsky (2013) solves {XP,YP} problems such as subject-predicate, subject-raising, and *wh*-movement in English, as briefly reviewed above. However, questions arise when it comes to languages where ϕ-feature agreement is absent. Japanese is such a language, since it does not have ϕ-feature agreement and obligatory *wh*-movement. Saito (2016, 2018) argues that case- marking functions as an anti-labeling device/labeling blocker, i.e., a phrase with a case marker becomes invisible for labeling. Saito (2016) also argues that PP bears a phonologically unrealized case marker, and therefore it also includes the anti-labeling device. He extends this anti-labeling device to inflection on predicates. Including case markers and inflectional elements, Saito (2016) calls these λ-features.

Related to Saito (2016, 2018), Miyagawa et al. (2019) propose the following.

(10) The labeling function of particles

a. [Particle: YS] [a]ttaches to a non-projectable element ⇒induces projection

b. [Particle: YS] [a]ttaches to a projectable element ⇒ blocks projection

(Miyagawa et al., 2019, 2, (1))

Starting from the labeling inducer, Miyagawa et al. (2019) argue that the Q-particle attached to C in Japanese functions to induce labeling.

(11) The function of the Q-particle

The Q-particle attaches to C to give it an independent category status, which induces labeling by C. (Miyagawa et al., 2019, 6, (11))

Since Japanese is a *wh-in-situ* language, this proposal explains the absence of obligatory *wh*-movement in Japanese, since a Q-particle does not need a *wh*-phrase in its specifier for labeling because the Q-particle itself can be a label (11).

In the case of Old Japanese, this is not true.

(12) Old Japanese

Idukuni-ka kimi-ga fune fate kutsa mutsubi-kemu
which-KA you-NOM ship stop grass tie-PST

'Where did you anchor your ship?' (Miyagawa et al., 2019, 5,(9))

Miyagawa et al. (2019) argue that in Old Japanese, the Q-particle (*ka*) was not on C, so the *wh*-phrase had to agree with the Q-particle by focus-movement (i,e, creating a {wh,CP} configuration). In this case, feature-sharing takes place (e.g., <Q,Q>), which is same as modern English.[6]

Miyagawa et al. (2019) also assume with Saito (2016, 2018) that there are labeling blockers such as case markers in Japanese, as discussed below. Thus, a com-

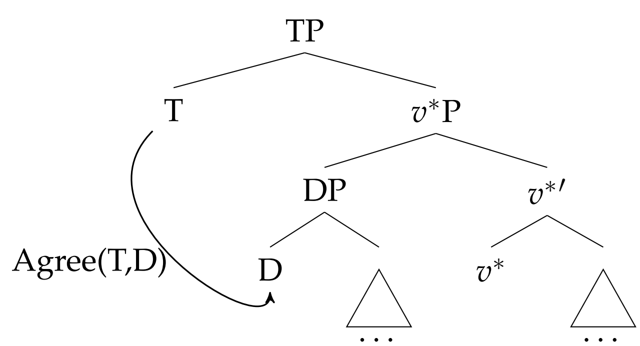
[6]See van Gelderen (2018) for the labeling approach to language change. See also Dadan (2019) for the preference for H-XP structure diachronically.

prehensive theory of labeling could have both strategies, (i) labeling inducer and (ii) labeling blockers.

Although Saito (2016, 2018) and Miyagawa et al.'s (2019) analysis of labeling provides an account of the labeling issues in Japanese, I will argue in this subsection that there are problems with this analysis. The rest of the subsection introduces and explains the issues with Saito's (2016; 2018) labeling analysis of Japanese.

Based on Bošković's (2007) idea of the Agree system, Saito (2018) assumes that case-assignment is independent from ϕ-feature agreement. Consider the following schematic tree representation.

(13)



In Chomsky (2000), T probes to find a goal, namely D, which has valued ϕ-features in (13). As a reflex of agreement, the unvalued feature of case will be valued.

(14) probe-goal based Agree (cf. Chomsky, 2000, 2001)

a. A probe P and a goal Q have a relation in terms of Matching

b. Matching is feature identity

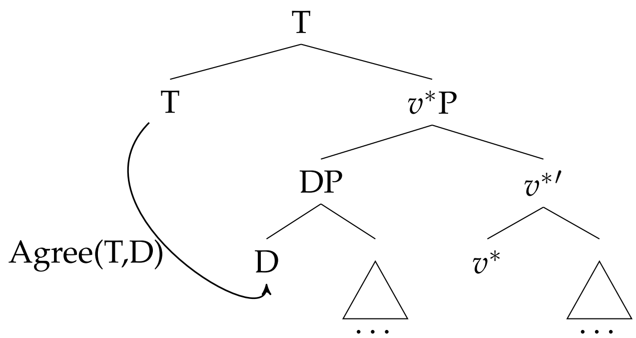
c. The domain of P is the sister of P

d. P finds Q where Q has the closest c-command relation with P

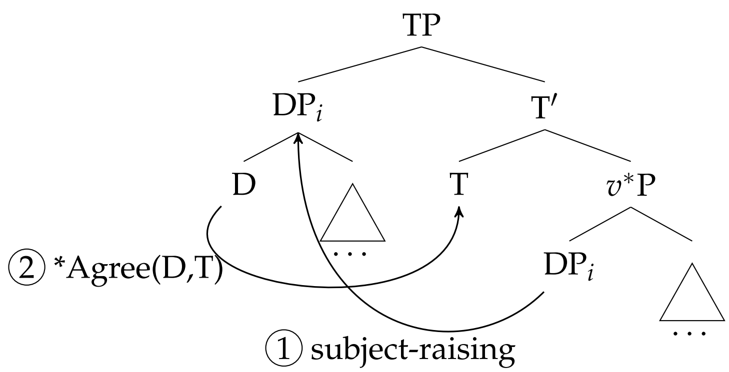
However, Bošković (2007) argues that the case-feature is valued independently from the ϕ-feature agreement, namely after DP moves to [spec,TP]. Since the moved DP in [spec,TP] has an unvalued case feature, this becomes the probe to find a relevant head, namely T, and the unvalued case-feature gets valued.

In Saito (2018), it is argued that Bošković's (2007) probe-goal approach does not hold since D itself, which has the unvalued case-feature, cannot c-command T, as illustrated in (15b).

(15) a.



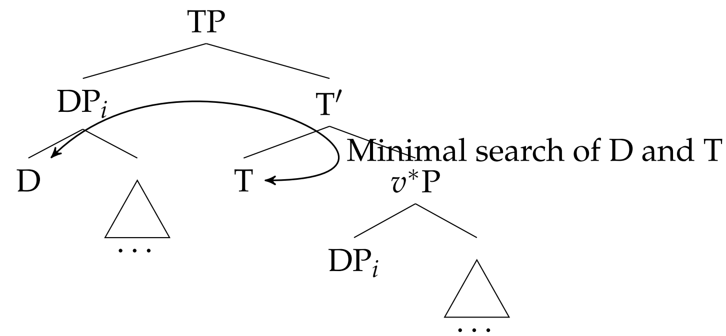
b.



Saito (2018) assumes that minimal search takes place instead, therefore D and T are found simultaneously, as illustrated in (16).[7]

[7] It's not entirely clear, then, why the raised DP has to move to [spec,TP] in the first place if the unvalued feature itself does not require a c-command relation with the goal. Furthermore, Bošković's analysis presupposes subject-raising takes place before Agree, which Saito (2018) adopts. However, it has been argued that subject-raising is, at least, not obligatory in Japanese (cf. Fukui, 1986; Kuroda, 1988; Ishii, 1997). Also, Saito (2018) does not define Search itself (whether Search is a breadth-first search or depth-first search; see the following discussion). See Ke (2019) for more details on the definition of minimal search.

(16)



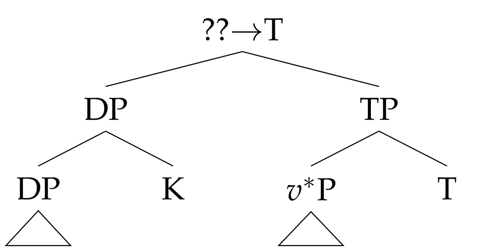
Since Japanese does not have ϕ-feature agreement, the feature-sharing labeling does not take place, but case-valuation can take place in the way discussed above. Saito (2016) proposes that the suffixal case functions as anti-labeling device that makes the phrase to which it is attached invisible for labeling. Thus, the labeling problem does not arise, for subject cases. Notice that although the subject in the [spec,TP] does not agree with T, it has an unvalued feature, namely, a case-feature. Thus, in Saito's (2016) analysis, based on Bošković (2007), the subject still has to move to [spec,TP] to value its unvalued case-feature (but see also footnote 7).

Saito (2018) further proposes the following.

(17) Search $\{\alpha, \beta\}$ for a label. If α is a weak head or search into α yields a weak head, then search on the α side is suspended and it continues only on the β side. (Saito, 2018, 6, (14))

Saito (2018) then assumes K, a head of KP (a case phrase), is a weak head, just as T in English is a weak head.

(18)



By definition, K is a weak head, thus when minimal search finds this head, it needs to suspend the search within the DP structure and instead find another relevant head, which is a T head. Saito (2018) assumes that T is a strong head in Japanese, unlike English. Hence, the label of the whole structure becomes T. Furthermore, minimal search finds these two heads (i.e., K and T), so the case-feature on K is valued as nominative by T.

The head K cannot be a label by definition, but it also functions as an anti-labeling device. So, the head D is invisible for labeling of the entire structure in (18), due to the fact that K is a weak head. At the same time, however, Saito (2018) assumes that the label for the {DP,K} structure should be DP since, for example, V satisfies the selectional relation with D, not with K, when DP is an object of the verb.

This analysis immediately explains the scrambling cases in Japanese. In (19b), the object *sono neko-o* 'the cat-ACC' is scrambled to the front of the sentence, which yields an {XP, YP} structure (here, the structure is {obj-ACC, CP}). Due to the λ-feature on the object, the scrambled object becomes invisible for labeling, by definition.

(19) a. Hanako-ga sono neko-o nade-ta
Hanako-NOM that cat-ACC pet-PAST
'Hanako petted that cat.'

b. sono neko-o$_i$ Hanako-ga t$_i$ nade-ta
that cat-ACC Hanako-NOM t$_i$ pet-PAST
'Hanako petted that cat.'

(20) a. Taro-ga [Hanako-ga sono neko-o nade-ta to] it-ta
Taro-NOM [Hanako-NOM that cat-ACC pet-PAST C] say-PAST
'Taro said that Hanako petted that cat.'

b. sono neko-o$_i$ Taro-ga [Hanako-ga t$_i$ nade-ta to] it-ta
that cat-ACC$_i$ Taro-NOM [Hanako-NOM t$_i$ pet-PAST C] say-PAST
'Taro said that Hanako petted that cat.'

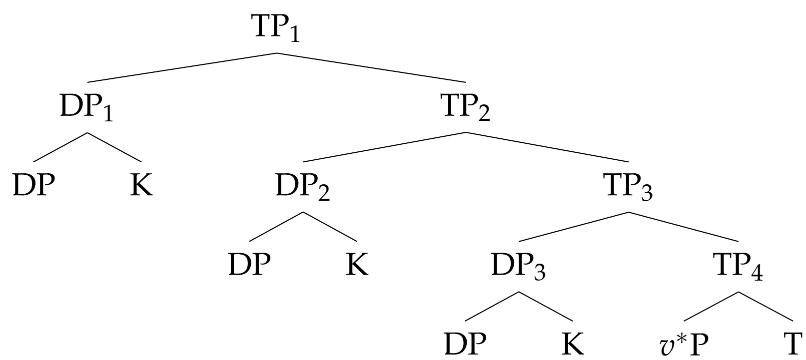
The same is true in (20b). In (20b), long-distance scrambling in Japanese is represented (Saito 1985, 1989, 1992 among many others). Scrambling of *sono neko-o* 'that cat-ACC' to the edge of the sentence makes {XP,YP}, but the object *neko-o* 'the cat-ACC' becomes invisible due to the accusative case marker, so the head of CP, namely C, becomes the label for this structure.[8]

Saito (2018) also discusses the multiple nominative construction in Japanese. The example is from Kuno (1973).

(21) Bunmeikoku-ga dansei-ga heikin-zyuumyo-ga mizika-i
civilized.country-NOM male-NOM average-life.span-NOM short-PRES
'It is in civilized countries that male's average life span is short.'

The structure for the multiple nominative construction in Saito (2018) is as follows.

(22)



For the relevant bottom structure (TP_3), K in DP_3 and T in TP_4 are found by minimal search, so DP_3 gets nominative case. When minimal search finds DP_2 and TP_3, it searches further into these objects. In DP_2, K is found, but it is a weak K. In TP_3, again, K and T are found. The same is true for DP_1 and TP_2. Thus, it is not clear how DP_1 and DP_2 get nominative case from T by minimal search, since for DP_2 and DP_3, T is not unambiguously found by minimal search. Saito (2018) stipulates that the label provider values a case feature. In the structure above, DP_2 gets nominative case by TP_3, since the label of TP_3 is T. DP_1 also gets labeled by TP_2 since the label of TP_2 is T.[9]

[8] But see also Goto (2013a) and Miyagawa et al. (2019) for the treatment of root clause labeling in Japanese. See also §2.6.2 for relevant discussion.

[9] See also Epstein et al. (2020) for relevant discussion. See also Goto & Ishii (2021) for the structure of the multiple nominative construction in Japanese. If Goto & Ishii (2021) is on the right track, the case valuation problem that we discuss here might be orthogonal.

To sum up, the assumptions in Saito (2018) are as follows.

(23) Saito's (2018) assumptions

a. K is a weak head

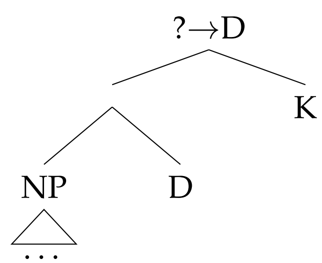
b. For a nominal phrase with a case marker, the label of KP is DP

c. T is a strong head

d. Case valuation takes place when the label provider is a sister of the DP

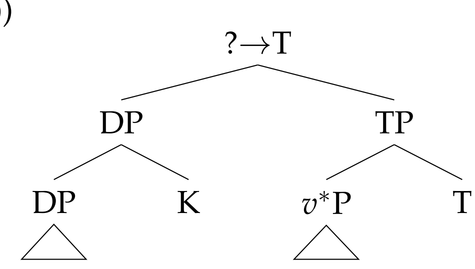
One issue is the labeling of DP. Let us consider the structure in (24) again.

(24)



Minimal search finds K first. However, this label is a weak head. (17) says that once minimal search finds K, then the search in that domain terminates. I assume here that the search domain is {DP, K}, so search into K is terminated. Then the search goes into DP and finds D, which becomes the label of D. Saito (2018) assumes that the label of the whole structure is D. However, when we search a structure such as (18), the search domain becomes {DP,TP}.

(25) (=(18))



Again, at this time, when K is found, minimal search terminates and search continues only into TP. Thus, it seems the determination of the domain of search changes what minimal search can find.

This seems to be different from the minimal search that Chomsky (2013) assumes. However, even assuming Chomsky (2013), it seems we get the right label assuming that T is a strong head and K is a weak head in Japanese. Suppose minimal search applies to the whole structure of (25). Minimal search finds T and K, but K is a weak head, thus T becomes the label. D is too deep to search. However, to label DP itself, D is found after K is found, and D becomes the label.

To summarize, in order to make Saito's labeling analysis in Japanese clearer, we have to say that by definition KP is invisible. However, if KP is invisible, we cannot label the relevant DP structure, which is supposed to be a DP or nominal phrase. Saito (2018) assumes that it can be labeled as D. Thus, within the DP structure D is visible, but outside of DP, it's invisible. Furthermore, in order to label a TP structure, T has to be a strong head, which is an unmotivated assumption.[10] There is a piece of evidence that a DP or nominal phrase is visible for subextraction in Japanese, which we will see in the following sections.

2.3.2 On Movement and Labeling

In the previous section, we discussed the issues in labeling theory in Japanese. In this subsection, the labeling approach to movement is discussed to show how labeling handles the constraint on movement.

As discussed in Chomsky (2013, 2015), labeling theory explains the EPP effect and successive cyclic *wh*-movement. The idea is that the derivation converges when an element moves to the stable {XP,YP} structure positions via feature-sharing. Subsequent work argues that unlabeled {XP,YP} structures make the derivation ungrammatical in terms of movement and extraction (Ott 2015; Goto 2016; Bošković 2016; Richards 2019, among many others). For example, the basic idea in Goto (2016) is that unlabeled structures are opaque for extraction, and he argues that

[10] See Goto (2017) for the relevant discussion about eliminating the strength of T.

this derives the locality effect.[11]

(26) Generalzation of extraction under Labeling (GEL)
An unlabeled {XP,YP} structure is opaque for extraction.

(Goto, 2016, 335)

Bošković (2016) and Richards (2019) also have a similar view, which derives some island effects.

(27) a. Subject island: *{nP,v^*P}
b. Adjunct island: *$\langle v^*$P,CP$\rangle$
c. CNPC: *{nP,CP}

(Richards, 2019, 149)

Although I won't go into the details for each work, the basic idea in Richards (2019) is that merging two phases (such as merging v^*P and CP) confuses the interfaces. Bošković (2016), on the other hand, derives the island effect from the anti-locality principle and the timing of the labeling. Regardless of whether we adopt either of them, the results seem to be identical. Namely, unlabeled {XP,YP} structures make derivations ungrammatical.

There is a much stronger version of the constraint on movement in terms of labeling. Narita (2014) assumes that merging two phrases is banned in the first place.[12]

(28) *{XP,YP}: there can be no merger of two phrasal SOs. (Narita, 2012, 3)

His proposal is to reduce one phrase to a head by multiple transfer. For example, when the external argument is merged with the predicate, it seems it is unavoidable to create an unlabelable structure {XP,YP}. However, Narita assumes that

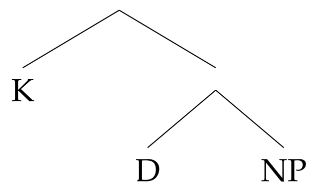
[11] Notice that this generalization is for subextraction. Thus, extraction of XP or YP is possible from an unlabeled {XP, YP} structure, as Chomsky (2013) discussed.

[12] See also Kayne (2011) for a similar idea, but with different assumptions.

the subject, which is a phrase with a phase head K, is reduced to a head K before externally merging with the predicate.

(29) a. Subject: {K,{D,NP}}

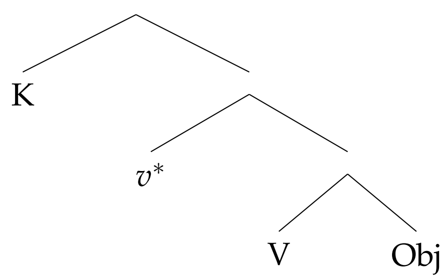
b.



c. noncomplement-reduction: K, where {D,NP} is transferred.

d. Merge K with v^*P: {K, {v^*,{V,Obj}}}

e.



The structure does not include an unlabeled {XP,YP} structure.[13] The consequence is the subject island is deduced, since the complement of the K as in (29b) is trans-

[13]See also Epstein & Shim (2015) for a reanalysis of the spec-head relation to a head-complement.

ferred, which is opaque for subextraction.[14,15]

Although these extensive analyses deduce some minimality effect from labeling theory, these strategies do not seem to work, assuming that case markers are anti-labeling devices/labeling blockers, at least in Japanese (Saito, 2016, 2018; Miyagawa et al., 2019). In Japanese, every movement creates an unlabelable {XP,YP} structure, but it is avoided by the case markers (Saito, 2016, 2018; Miyagawa et al., 2019). That is, Japanese avoids unlabelable {XP,YP} problems only when one of the phrases is case-marked or when it has an inflectional element (i.e., λ-feature). In what follows, however, I will show cases where the labeling problem does not occur, but the sentences are nonetheless ungrammatical, which is not predicted by the labeling analysis of movement in Japanese.

[14]Notice that in Narita's (2014) analysis (see also Lohndal 2014), the absence of subject island effect is also captured by transferring the complement to reduce it to a head. Thus, the interior of the subject is still available for further operations, whereas the interior of the complement is not, which he calls a Complement island.

i. If subextraction applies from the Spec of H, then the complement of H becomes an island for extraction. (Narita, 2014, 119)

Although this approach captures variations in the subject island effect among languages, it heavily relies on multiple transfer, which seems to face an empirical problem. Chomsky et al. (2019: 240-241) points out that if Transfer literally eliminates the structures from the narrow syntax, and if the spell-out takes place where the transfer takes place, the following example (i) cannot be pronounced in the expected way.

i. [$_\alpha$ the verdict [$_\beta$ that Tom Jones is guilty]] seems to have been reached (α) by the jury
ii. The verdict seems to have been reached that Tom Jones is guilty by the jury.

α's base-generated position is the object of the verb *reach*, which raised to the subject. If Transfer of β takes place at the base position, the pronounced sentence would be (ii). This multiple transfer analysis that eliminates the structure does not produce the linear order in (i). Chomsky et al. (2019) concludes that Transfer at least renders the transfer domain inaccessible for further syntactic operations. Considering this discussion, I won't assume the multiple transfer approach to syntax, though there are many empirical consequences of, and other conceptual arguments for this approach. See Narita (2014); Lohndal (2014) regarding these points.

[15]One might wonder, then, why English doesn't use this strategy. See §2.6 for relevant discussion.

2.4 Issues Regarding Movement in Japanese

2.4.1 (Im)proper Scrambling in Japanese

It has been argued that scrambling in Japanese is not feature-driven and scrambling of syntactic objects is relatively free (cf. Fukui, 1993b; Saito & Fukui, 1998).[16] Thus, the scrambling is not driven by ϕ-features or a Q-feature, and gives Japanese a relative free word order. One might think any conceivable scrambling patterns in Japanese should be acceptable.

However, interestingly, Sakai (1994, 1996), based on Saito's (1985) observation, points out that when the landing site of the long-distance scrambling is from the finite embedded clause to the post-subject position in the matrix clause, the sentence becomes degraded.

(30) a. Masao-ga Kumiko-ni [Takashi-ga Boston-e it-ta-to]
Masao-NOM Kumiko-DAT [Takashi-NOM Boston-to go-PAST-C]
it-ta.
say-PAST
'Masao told Kumiko that Takashi went to Boston.'

b. ?*Masao-ga **Boston-e**$_i$ Kumiko-ni [Takashi-ga t$_i$ it-ta to]
Masao-NOM **Boston-to**$_i$ Kumiko-DAT [Takashi-NOM t$_i$ go-PAST C]
it-ta.
say-PAST
'Masao told Kumiko that Takashi went to Boston.' (Sakai, 1996, 129, (10))

This is unexpected from the labeling analysis.[17] The labeling algorithm finds the scrambled element and minimal search ignores this phrase (i.e., the Case suffixal element is a labeling blocker). Note that PP does not have an overt case

[16]Since Japanese is a strictly head-final language, switching the verb position is not available, though right-node raising is possible. In this chapter, I will only discuss leftward movement in Japanese.

[17]As far as I know, Goto (2013a) is the first and the only work to discuss this type of data in terms of labeling.

marker, but Saito (2016) assumes that PP also has an unrealized case marker. Also, Saito assumes that an adverbial element has some sort of inflectional feature (i.e., a λ-feature) as shown below.

(31) PP-scrambling in Japanese

a. Hanako-ga tosyokan-kara hon-o karidasi-ta
Hanko-NOM library-from book-ACC check.out-PAST
'Hanako checked out a book from the library'

b. tosyokan-kara$_i$, Hanako-ga t$_i$ hon-o karidasi-ta
library-from Hanko-NOM t$_i$ book-ACC check.out-PAST
'Hanako checked out a book from the library.' (Saito, 2016, 142, (27))

(32) Adverb-scrambling

a. Taroo-wa sizuka-ni kaet-ta
Taroo-TOP quietness-COP leave-PAST
'Taroo left quietly.'

b. Sizuka-ni Taroo-wa t$_i$ kaet-ta
quietness-COP$_i$ Taroo-TOP t$_i$ leave-PAST
'Taroo left quietly.' (Saito, 2016, 142,(28))

Thus, PP/adverbial scrambling is fine in Japanese. However, again, the labeling theory does not predict (30b).

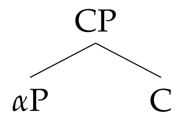
Another property of scrambling is that it's possible to scramble the clause (CP) itself (cf. Saito, 1985).

(33) [Takashi-ga Boston-e it-ta-to]$_i$ Masao-ga Kumiko-ni t$_i$
[Takashi-NOM Boston-to go-PAST-C]$_i$ Masao-NOM Kumiko-DAT t$_i$
it-ta.
say-PAST
'Masao told Kumiko that Takashi went to Boston.'

This example is also unexpected if the case marker makes the element invisible since the scrambled embedded clause does not have a case marker attached to it, which predicts that the root clause with the scrambled CP would remain unlabeled, hence the sentence should be ungrammatical.

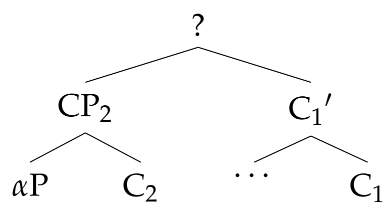
Assuming Saito's (2016) analysis of anti-labeling devices, inflection on the predicate is also an instance of anti-labeling device (i.e., λ-feature). (34) shows when αP and C are externally merged.

(34) (Saito, 2016, 143, (31))



In (34), αP has a λ-feature, just as a DP structure has a case marker. It seems, then, Saito (2016) assumes that C becomes the label. However, the CP scrambling (see (35) below) does not seem to work even if we assume Saito (2016). Let us consider the following tree representation.

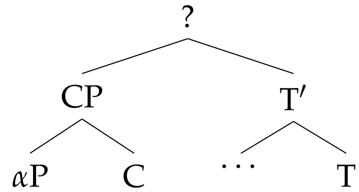
(35)



In this structure, the scrambled CP moves to the specifier of the matrix CP. If C is a strong head, then two Cs are found, which results in the unlabeled {XP, YP} structure. If, on the other hand, C has the λ-feature, then both Cs are weak heads (just like K is a weak head), and it is unclear what the label would be. One might think transfer takes place at the CP level, so the right side of the structure in (35) is C_1, not C_1'. However, the same applies to the scrambled CP, namely, before the scrambling of CP, the complement of C is transferred. Thus, after the transfer, the representation of the structure in (35) will be $\{C_2, C_1\}$. Minimal search will find two heads Cs, hence this is an unlabeled structure (See also §2.6).

The result is not clear even if we assume that CP scrambling is TP adjunction (cf. Saito, 1985). Let us consider the relevant structure below.

(36)



In the scrambled CP, αP has an inflectional element (λ-feature), thus the label of CP is C, assuming that C is a strong head. In the right side of the structure in (36), suppose T is a strong head, by definition from Saito (2018); then C and T are found, which results in an unlabeled {XP,YP} structure.

T usually has an inflectional feature, so suppose it has the λ-feature proposed by Saito (2016) (i.e. an anti-labeling device); then the label of the whole structure will be C since we assume C is a strong head. The result is not what we want since if the label of the whole structure is C, then the scrambled clause becomes the main spine for the whole structure.

Even if we adopt Miyagawa et al.'s (2019) analysis (11), labeling theory cannot label the structure since C is not a Q-particle (i.e., it is a declarative complementizer).

Suppose we don't need to label the root clause (Goto, 2013a,b; Blümel, 2017; Miyagawa et al., 2019).[18] Thus, the problem in (35) and (36) is not a problem anymore (see also §2.6.2). However, this does not solve the problem since CP scrambling can apply to the embedded clause. In fact, if we have more than one embedded clause, it is possible to do such scrambling in Japanese (cf. Saito, 1985).

(37) a. Taroo-ga [$_{CP2}$ Maso-ga Kumiko-ni [$_{CP1}$ Takashi-ga
Taroo-NOM [$_{CP2}$ Masao-NOM Kumiko-DAT [$_{CP1}$ Takashi-NOM
Bosteon-e it-ta-to] it-ta-to] ki-ta
Boston-to go-PAST-C] say-PAST-C] hear-PAST
'Taroo heard that Masao told Kumiko that Takashi went to Boston'

[18] For example, Blümel (2017) proposes that the root clause in a V2 environment in German, the root label has to be unlabeled. See §2.6.2 for relevant discussion.

b. Taroo-ga [$_{CP2}$ [$_{CP1}$ Takashi-ga Bosteon-e it-ta-to]$_i$
Taroo-NOM [$_{CP2}$ [$_{CP1}$ Takashi-NOM Boston-to go-PAST-C]$_i$
[Maso-ga Kumiko-ni t_i it-ta-to]] ki-ta
[Masao-NOM Kumiko-DAT t_i say-PAST-C]] hear-PAST
'Taroo heard that Masao told Kumiko that Takashi went to Boston'

In (37b), the most embedded clause (CP_1) is scrambled to the edge of the next embedded clause (CP_2). Thus, if this scrambling yields the unlabeled {XP,YP} structure, the labeling analysis does not predict this is a grammatical derivation, contrary to fact. That is, the fact is both sentences in (37) are grammatical, though the labeling theory wrongly predicts that (37b) is ungrammatical.

Going back to the post-subject scrambling, if the embedded clause is non-finite and long-distance scrambling to the post-subject position takes place, then the sentence becomes relatively acceptable, as in (38b).

(38) a. Masao-ga Kumiko-ni$_i$ [PRO$_i$ Boston-e iku-yooni] susume-ta.
Masao-NOM Kumiko-DAT$_i$ [PRO$_i$ Boston-to go-so-that] advise-PAST
'Masao advised Kumiko to go to Boston.'

b. Masao-ga **Boston-e**$_j$ Kumiko-ni$_i$ [PRO$_i$ t_j iku-yooni] susume-ta.
Masao-NOM **Boston-to**$_j$ Kumiko-DAT$_i$ [PRO$_i$ t_j go-so-that] advise-PAST
'Masao advised Kumiko to go to Boston.' (Sakai, 1996, 130, (11))

Also, notice that when the scrambled element moves to the left edge of the sentence, it is acceptable as well, as shown below.

(39) **Boston-e**$_i$ Masao-ga Kumiko-ni [Takashi-ga t_i it-ta-to]
Boston-to$_i$ Masao-NOM Kumiko-DAT [Takashi-NOM t_i go-PAST-C]
it-ta.
say-PAST
'Masao told Kumiko that Takashi went to Boston.'

The next set of examples shows that the most embedded clause S_3 is non-finite and the scrambled element moves up to the matrix VP, which makes the sentence unacceptable.

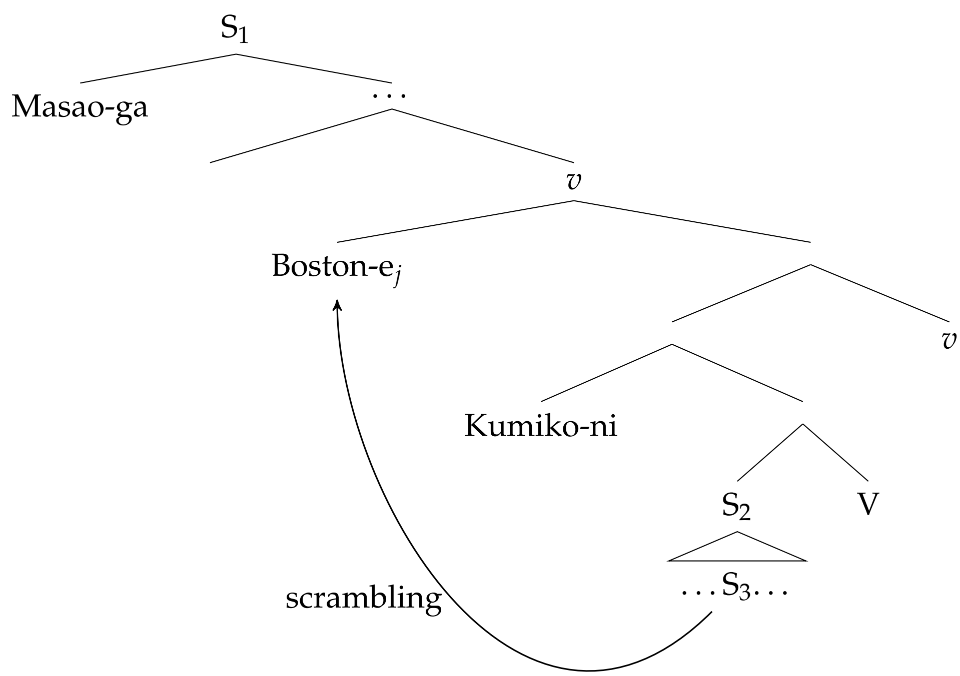
(40) a. Masao-ga Kumiko-ni [$_{S2}$ Takashi-ga$_i$ [$_{S3}$ PRO$_i$ Boston-e
Masao-NOM Kumiko-DAT [$_{S2}$ Takashi-NOM$_i$ [$_{S3}$ PRO$_i$ Boston-to
ikoo-to] keikakushi-te-iru-to] it-ta]
go-so-that] plan-STAT] say-PAST]
'Masao told Kumiko that Takashi is planning to go to Boston.'

b. *[$_{S1}$ Masao-ga **Boston-e**$_j$ Kumiko-ni [$_{S2}$ Takashi-ga$_i$ [$_{S3}$ PRO$_i$ t$_j$
[$_{S1}$ Masao-NOM **Boston-to**$_j$ Kumiko-DAT [$_{S2}$ Takashi-NOM [$_{S3}$ PRO$_i$ t$_j$
ikoo-to] keikakushi-te-iru-to] it-ta]
go-so-that] plan-STAT] say-PAST
'Masao told Kumiko that Takashi is planning to go to Boston.'

(Sakai, 1996, 130-131, (13))

Again, the labeling analysis of Japanese cannot capture the types of sentences that we showed in this subsection. In (40b), the scrambled element that has a λ-feature forms {XP,YP}. Therefore, the structure S_1 should be able to be labeled properly (if this is scrambling to [spec,*v*P], the label is *v* for the relevant {XP,YP} structure in (40b)), as represented in the tree structure in (41).

(41)



Thus, Saito's labeling analysis predicts the sentence (40b) is grammatical, though it is ungrammatical.

In this subsection, the CP scrambling and (im)proper scrambling in Japanese were discussed. We argued that the labeling theory proposed by Saito (2016) does not provide an account for the embedded CP scrambling and (im)proper scrambling.

In the next subsection, we explore remnant movement in terms of labeling.

2.4.2 Scrambling and Remnant Movement

There is another type of derivation that makes the sentence degraded. i.e., remnant movement, movement of the phrase that contains a trace/a lower copy (Takano, 1994, 1995; Müller, 1996; Kitahara, 1997; Hiraiwa, 2010).

(42) a. [$_{TP}$ Ken-ga [$_{CP2}$ Naomi-ga [$_{CP1}$ Yuko-ga gakusei-to at-ta to] omotteiru to] it-ta].
[$_{CT}$ Ken-NOM [$_{CP2}$ Naomi-NOM [$_{CP1}$ Yuko-NOM student-with meet-PAST C] think-PRES C] say-PAST]
'Ken said that Naomi thought that Yuko met with students.'

b. [$_{CP3}$ Ken-ga [$_{CP2}$ **gakusei-to**$_i$ [Naomi-ga [$_{CP1}$ Yuko-ga t$_i$ at-ta to] omotte-iru to]] it-ta.]
[$_{CP3}$ Ken-NOM [$_{CP2}$ **student-with**$_i$ [Naomi-NOM [$_{CP1}$ Yuko-NOM t$_i$ meet-PAST C] think-PRES C]] say-PAST]
'Ken said that Naomi thought that Yuko met with students.'

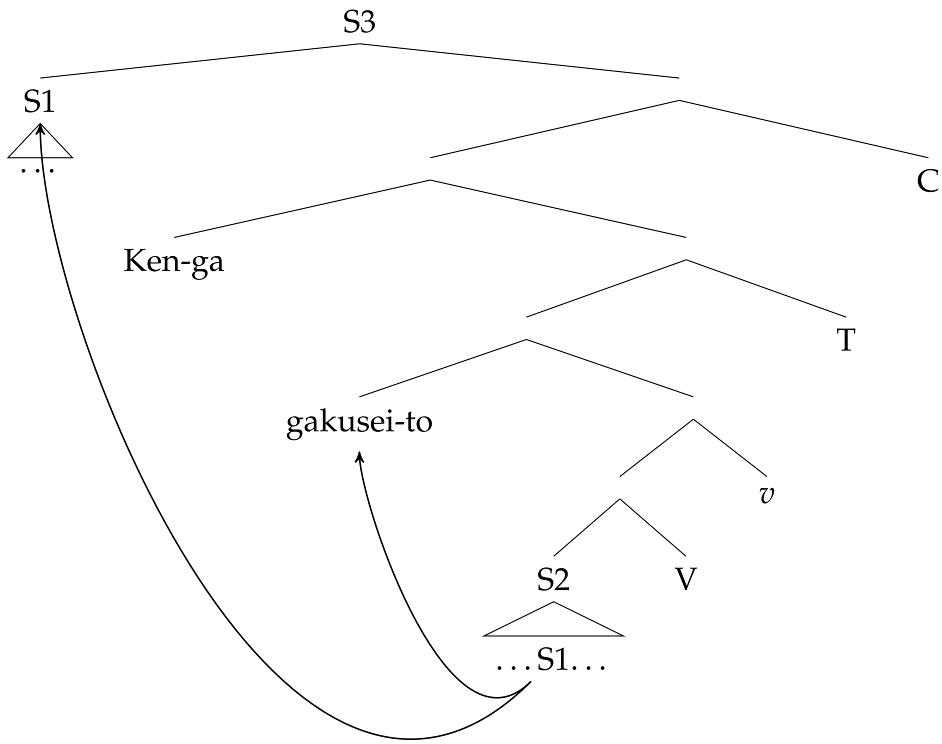
c. [$_{CP3}$ [$_{CP1}$ **Yuko-ga gakusei-to at-ta to]**$_j$ [$_{TP}$ Ken-ga [$_{CP2}$ Naomi-ga t$_j$ omotte-iru to] it-ta]].
[$_{CP3}$ [$_{CP1}$ **Yuko-NOM student-with meet-PAST C]**$_j$ [$_{TP}$ KenNOM [$_{CP2}$ Naomi-NOM t$_j$ think-PRES say-PAST]]
'Ken said that Naomi thought that Yuko met with students.'

d. *[$_{CP3}$ [$_{CP1}$ Yuko-ga t$_i$ at-ta to]$_j$ [$_{TP}$ Ken-ga [$_{CP2}$ gakusei-to [$_{CP2}$ Naomi-ga t$_j$ omotte-iru to]] it-ta]].
[$_{CP3}$ [$_{CP1}$ Yuko-NOM t$_i$ meet-PAST C]$_j$ [$_{TP}$ Ken-NOM [$_{CP2}$ student-with$_i$ [$_{CP2}$ Naomi-NOM t$_j$ think-PRES C]] say-PAST]]
'Ken said that Naomi thought that Yuko met with students.'

e. $[_{CP} \ldots \text{t}_{PP} \ldots] \ldots \text{PP} \ldots \text{t}_{CP} \ldots$ (Hiraiwa, 2010, 135-136, (5))

The first scrambling in (42b) is scrambling PP *gakusei-to* 'with student' out of the embedded clause to the edge of the CP_2, which does not make the sentence degraded. Movement of CP itself is also possible, as already discussed, and is represented in (42c). (42d) shows two steps of movement: (i) PP-scrambling out of CP_1, and (ii) movement of CP_1 itself, which contains the copy of PP. The schematic representation of (42d) is shown below.[19]

(43)



This is called remnant movement, which follows the proper binding condition.

(44) The Proper Binding Condition (PBC) (Fiengo, 1977)

Traces must be bound.

Informally speaking, in (42d), the scrambled CP_1 includes a trace/copy which is not bound by any element. Thus, PBC predicts that this sentence is ungrammatical.

[19]Parts of the structure is omitted. See §2.7.4 for the detail.

Interestingly, it has been observed in the literature that there is an example that does not show the proper binding condition effect.

(45) A-movement followed by remnant-CP scrambling/topicalization: (Hiraiwa, 2010, 141,(16))

a. [$_{TP}$ Ken-ga$_i$ minna-ni [$_{CP}$ t$_i$ baka-da to] omow-arete-iru]
[$_{TP}$ Ken-NOM$_i$ everyone-DAT [$_{CP}$ t$_i$ foolish-PRES C] think-PASS-PRES]
'Ken is thought to be stupid by everyone.'

b. **[$_{CP}$ t$_i$ baka-da to](-wa)$_j$** [$_{TP}$ *Ken-ga$_i$* minna-ni t$_j$ omow-arete-iru]
[$_{CP}$ t$_i$ foolish-PRES C](-TOP)$_j$ [$_{TP}$ *Ken*-NOM$_i$ everyone-DAT t$_j$ think-PASS-PRES]
'Ken is thought to be stupid by everyone.'

(45b) shows that after the passivization (45a), the remnant phrase moves up to the edge of the sentence. PBC wrongly predicts that this is ungrammatical.

Müller (1996, 1998) proposes that a condition on the two movements (i.e., the remnant-creating movement and the remnant movement), which he calls *the Condition of Unambiguous Domination*.

(46) Unambiguous Domination

In a structure ... [$_A$ ... [$_B$ ...] ...] ... , A may not undergo α-movement if B has undergone α-movement. (adopt from Grewendorf, 2015, 5)

This explains (45b), since the remnant-creating movement is A-movement, followed by remnant-CP scrambling (or topicalization).

However, raising to object followed by CP remnant scrambling/topicalization is impossible, as shown in (47b).

(47) A-movement followed by remnant-CP scrambling/topicalization

a. [$_{TP}$ Ken-ga [$_{vP}$ Naomi-o$_i$ kokorokara [$_{CP}$ t$_i$ baka-da to]$_j$
[$_{TP}$ Ken-NOM [$_{vP}$ Naomi-ACC$_i$ really [$_{CP}$ t$_i$ foolish-COP C]$_j$
omot-ta]].
think-PAST]]

'Ken really considered Naomi to be a fool.'

b. *[$_{CP}$ **t$_i$ baka-da to]$_j$** [$_{TP}$ Ken-ga [$_{vP}$ *Naomi-o$_i$* (kokorokara) t$_j$
[$_{CP}$ **t$_i$ foolish-COP C]$_j$** [$_{TP}$ Ken-NOM [$_{vP}$ *Naomi*-ACC$_i$ (really) t$_j$
omot-ta]]
think-PAST]]

'Ken really considered Naomi to be a fool.'

(Hiraiwa, 2010, 141, (18)) (see also Kuno (1976))

The ungrammaticality of (47b) is not explained by (46). The remnant-creating movement is A-movement followed by remnant-CP scrambling (topicalization). Even if we assume Saito's (2016) labeling blocker analysis (cf. Miyagawa et al. 2019), this does not predict the contrast among (42d), (45b), and (47b). For example, in (47a), the scrambling of *Naomi-o* 'Naomi-ACC' should be fine; scrambling of *Naomi-o* yields an {XP,YP} structure, but *Naomi-o* has a case marker (i.e., a λ-feature), thus, it becomes invisible for labeling. There is no labeling problem. In this respect, Saito's (2016) labeling analysis predicts (47a) to be grammatical, which, in fact, it is. However, in the case of CP scrambling, Saito's (2016) labeling analysis does not have a clear prediction, as we discussed. CP is scrambled to the edge of the sentence, and yields an {XP,YP} structure. If we assume that the root can remain unlabeled, as we discussed above, the sentence (47b) should be grammatical, contrary to fact. If we simply say the unlabeled root clause is ungrammatical, we predict (47b) is ungrammatical. However, this predicts that (45b) is also ungrammatical, contrary to fact. Thus, Saito's (2016) labeling analysis does not seem to predict the (un)grammaticality of CP remnant movement in Japanese (i.e., in both (45b) and (47a)).

In next subsection, we discuss another type of movement, namely, the freezing effect. Again, we will see that the labeling analysis does not seem to predict the (un)gramamaticality of the sentences in Japanese.

2.4.3 The Freezing Effect in Japanese

The following set of data shows multiple applications of scrambling, resulting in a freezing effect configuration (a freezing effect: movement out of a moved element is impossible, cf. Wexler & Culicover 1980), yet it does not result in ungrammaticality.

(48) a. [$_{IP}$ John-ga [$_{CP}$ [$_{IP}$ Bill-ga [$_{CP}$ Mary-ga sono hon-o
[$_{IP}$ John-NOM [$_{CP}$ [$_{IP}$ Bill-NOM [$_{CP}$ Mary-NOM that book-ACC
katta to] itta] to] omotteiru].
buy-PAST that] say-PAST] that] think].
'John thinks that Bill said that Mary bought that book.'

b. [$_{IP}$ John-ga [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-ga sono hon-o katta to]$_i$
[$_{IP}$ John-NOM [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary–NOM that book-ACC buy-PAST that]$_i$
[Bill-ga t$_i$ itta]] to] omotteiru].
[Bill-NOM t$_i$ say-PAST]] that] think]
'John thinks that [that Mary bought that book]$_i$, Bill said t$_i$.'

c. [$_{IP}$ sono hon-o$_j$ [John-ga [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-ga t$_j$ katta to]$_i$
[$_{IP}$ that book$_j$ [John-NOM [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-NOM t$_j$ buy-PAST that]$_i$
[Bill-ga t$_i$ itta]] to] omotteiru].
[Bill-NOM t$_i$ say-PAST] that] think].
'That book$_j$, John thinks that [that Mary bought t$_j$]$_i$, Bill said t$_i$.'

(Saito & Fukui, 1998, 465-466, (65))

In (48b), CP scrambling takes place and then the object of the scrambled CP is scrambled to the edge of the sentence (48c).

The same is true for the sentence in which the embedded clause is non-finite.

(49) a. [$_{IP}$ John-ga [Bill-ga Mary-ni$_k$ [PRO$_k$ sono ronbun-o yomu
[$_{IP}$ John-NOM [Bill-NOM Mary-DAT$_k$ [PRO$_k$ that article-ACC read
yooni] susumeta to] omotteiru]
that] advise-PAST that] think]

'John thinks that Bill advised Mary to read that article.'

b. [$_{IP}$ John-ga [PRO$_k$ sono ronbun-o yomu yooni]$_i$ [Bill-ga
[$_{IP}$ John-NOM [PRO$_k$ that article-ACC read that]$_i$ [Bill-NOM
Mary-ni$_k$ t$_k$ susumeta to] omotteiru]
Mary-DAT$_k$ t$_i$ advise-PAST that] think]

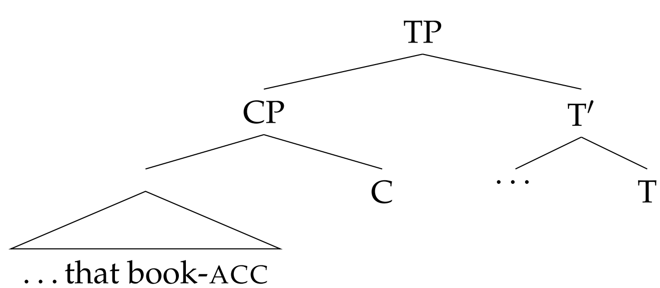
Lit.'John thinks that [to read that article]$_i$, Bill advised Mary t$_i$. '

c. [$_{IP}$ sono ronbun-o$_j$ [John-ga [PRO$_k$ t$_j$ yomu yooni]$_i$ [Bill-ga
[$_{IP}$ that article-ACC$_j$ [John-NOM [PRO$_k$ t$_j$ read that]$_i$ [Bill-NOM
Mary-ni$_k$ t$_i$ susumeta to] omotteiru]
Mary-DAT$_i$ t$_i$ advise-PAST that] think]

'That article$_j$, John thinks that [to read t$_j$]$_i$, Bill advised Mary t$_i$. '

Considering these examples, it's not entirely clear how a labeling analysis deals with CP scrambling. If the scrambled CP is invisible for labeling assuming the C head has a λ-feature, the labeling problem does not arise. However, if it is invisible for labeling, then why is extraction possible from the scrambled CP, given that it is natural to assume that the element which is invisible for labeling due to the λ-feature is inaccessible for further operations?

(50)



Goto (2013b) assumes that an unlabeled {XP,YP} structure becomes an opaque domain for subextraction, which implies that a labeled structure becomes invisible for extraction.[20] It is natural to extend this idea to invisible elements, such as a DP

[20]See §2.3.2 for relevant discussion.

which has a λ-feature. However, assuming Saito's (2018) analysis, the prediction is not clear. Saito (2018) assumes that for labeling, a λ-feature makes the element that the case marker is attached to invisible, but he does not say anything about the interior of the element that becomes invisible. Furthermore, if C is a weak head, as we discussed, the complement of the CP also has a λ-feature, hence, the label of the scrambled CP structure remains unlabeled.

Another complication is the status of C and T, as we discussed. Suppose we assume with Saito (2016, 2018) that T and C are strong heads. C and T are found, and the structure becomes unlabelable. Suppose C is a weak head or labeling blocker. Then, T becomes the label and the complement of C is invisible, hence extraction is predicted to be impossible, contrary to fact.

In this subsection, we discussed the freezing effect in Japanese. We noted that Saito's (2016; 2018) analysis does not seem to predict how exactly the invisibility of labeling works with the opaque domain for subextraction from a CP clause.

2.5 Interim Summary

So far we have discussed the followings:

(51) a. **CP scrambling and Root phenomena:** For Saito (2016, 2018), if the scrambled elements have suffixal elements, there is no labeling problem since the suffixal elements are anti-labeling devices. However, CP does not have a case marker. Thus, when the embedded CP itself is scrambled to the root clause, there is no way of labeling the root. In Miyagawa et al. (2019), a Q-particle attached to C is a label inducer, but the cases we saw involve scrambling of the declarative CP (C without a Q-particle), as opposed to interrogative sentences. In fact, scrambling of CP to the embedded clause is possible, which again cannot be cap-

tured by the current labeling theory. This is due to the unclarity of the status of C. Furthermore, as we discussed in §2.4.1, there will be configurations where both Cs will be found. Thus, regardless of the status of the head C, it is entirely unclear how to label the whole structure when the CP is scrambled to another CP (which is not a root clause, assuming the multiple embedding construction).

b. **Improper Scrambling:** The legitimacy of scrambling to the post-subject position depends on the status of the movement type (Saito, 1985; Sakai, 1994, 1996), but as we discussed, the labeling theory does not predict the ungrammaticality of (im)proper scrambling.

   i. Movement out of a finite embedded clause to the post-subject position in the matrix clause is improper scrambling.

   ii. Movement out of a non-finite embedded clause to the post-subject position in the matrix clause is legitimate.

   iii. For improper scrambling (51b-i) and proper scrambling (51b-ii), labeling theory predicts that both are grammatical since the scrambled DP has a case marker, which makes DP invisible by definition in Saito (2016).

c. **Proper Binding Condition:** The legitimacy of scrambling followed by remnant movement depends on the status of the movement type (cf. Müller, 1996; Takano, 1995). It also depends on the landing site (i.e., whether it is a phasal edge or not, cf. Hiraiwa 2010).

   i. A-type scrambling followed by scrambling/topicalization of the remnant CP is possible.

   ii. The first and second movement (remnant movement) cannot be the same type of movement (Müller, 1996; Takano, 2000; Hiraiwa, 2010).

iii. The examples we saw were A-movement followed by remnant CP scrambling/topicalization. However, the asymmetry between subject raising (45b) and raising-to-object (47b) was observed. Even though the movement types are different in both examples, (45b) is grammatical, whereas (47b) is ungrammatical, which cannot be captured by the current labeling theory. Furthermore, as we discussed, CP scrambling cases are not entirely clear, which also applies to the remnant CP movement.

d. **Freezing Effect:** Japanese does not show a freezing effect.

   i. It is not clear how the labeling analysis in Saito (2016) predicts the example in (49c). First, the labeling of the CP scrambling to the embedded edge position is not clear in Saito (2016, 2018).

   ii. If the scrambled CP is invisible, why is the movement of the interior element in the scrambled embedded CP possible?

The problems for the labeling approach is that it cannot capture the facts in (51) at all, as restated in (52) below.

(52) a. The scrambling of CP and the root labeling are unclear. How does it work? Do we need a label or not?

b. If case-marked elements make the phrases invisible, then there is no labeling problem in Improper Scrambling (51b) and Proper Binding Condition (51c), contrary to fact.

For the purposes of our discussion, what we need to explain are the followings.

(53) a. The (un)grammaticality of improper scrambling

b. The (un)grammaticality of remnant movement

c. The absence of a freezing effect

	grammatical	prediction by labeling theory (Saito, 2016, 2018)
Scrambling (e.g., (19b) and (20b))	✓	✓
Proper scrambling (38b)	✓	✓
Improper scrambling (30b)	*	✓
CP scrambling (33)	✓	*
Remnant movement (RTS, (45b))	✓	?
Remnant movement (RTO, (47b))	*	?
Freezing effect (49c)	✓	?

Table 2.1: The summary of types of movement and the predictions of the labeling theory for Japanese

The summary is in Table 2.1 below.

Before moving on to the analysis of movement in Japanese, I would like to review alternative approaches to labeling in Japanese. Then the proposal to explain the examples above will be introduced in §2.7, where we will discuss how to capture (53).

2.6 Alternative Labeling Analyses

In this subsection, we discuss alternative approaches to Japanese in terms of labeling. Let's step back and see the basic derivation by phases and how labeling works in Japanese.

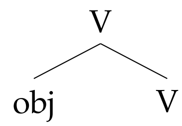
2.6.1 $v*$P Domain

In the v^*P domain, the object and the verb are externally merged then v^* is introduced.

(54) a. Merge (V, obj) = {V, obj}

b. Minimal Search (MS) finds V.

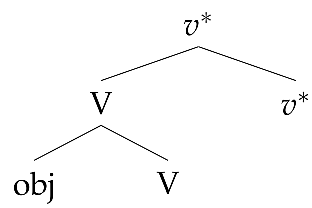
c.



(55) a. Merge (v^*, {V, Obj}) = {v^*, {V, Obj}}

b. Minimal Search (MS) finds $v*$.

c.

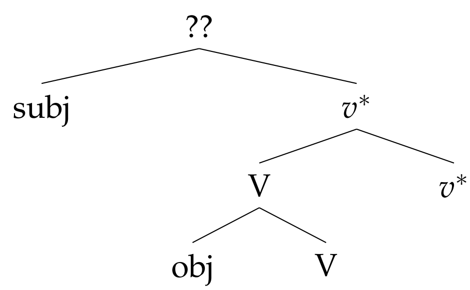


When the subject is introduced, Chomsky's (2013) problem arises. Namely, the subject and v^*P are in an {XP,YP} configuration.

(56) a. Merge (subj, {v^*, {V, Obj}}) = {subj, {v^*, {V, Obj}}}

b. MS finds D/n in *subj* and v^*.

c.



Chomsky's (2013; 2015) solution is to raise the subject to [spec, TP] later. However, here are two problems.

(57) a. **A conceptual problem: The timing of labeling:** If we strictly follow Chomsky's labeling analysis, the labeling applies to the phasal complement. At the v^*P phase, the phase edge is not supposed to be labeled otherwise it always fails to be labeled, i.e., the phasal edge creates the unlabeled {XP,YP} structure.

b. **An Empirical problem:** Some languages do not show clear evidence of obligatory subject-raising to [spec,TP]. At least in Japanese and German, the subject can move to [spec,TP], which creates another {XP,YP} structure at the TP level (e.g., {Subj,T′}) without feature sharing, which results in an unlabeled {XP,YP} structure.[21]

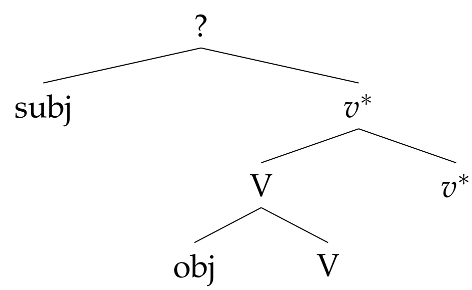
There are some technical ways of avoiding this unlabeled {XP,YP} structure. The first approach is the multiple transfer analysis.

(58) **Multiple Transfer:** Narita (2014); Lohndal (2014); Epstein & Shim (2015); Takita et al. (2016)
The complement of the phase head is transferred at the phase level, then the spec-head relation (e.g., {subj,v^*P}) reduces to the head-complement relation (i.e., {subj,v^*}).

(59) a. Before Transfer:

i. {subj,v^*P}

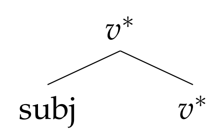
ii.



b. After Transfer:

i. {subj,v^*}

ii.



After transfer of the VP structure, the resultant structure is {sub, {v^*}}.[22] Narita (2014); Epstein & Shim (2015); Takita et al. (2016) suggest that this structure becomes the head-complement relation, namely an {H,XP} structure. However, if

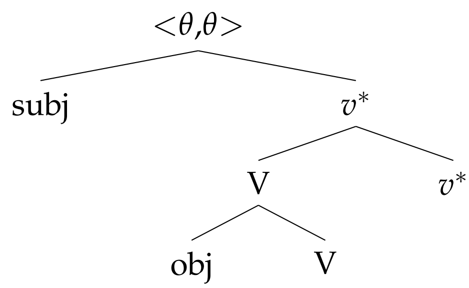
[21]Chomsky (2013) suggests based on Alexiadou et al. (2001) that either the external argument or internal argument moves out of the phase, so an {H,XP} structure is formed. However, as I argue here, there is no feature sharing strategy for the raised XP. Thus, Chomsky's (2013) solution will end up with creating another unlabeled {XP,YP}) structure in the further derivation.

[22]See also Chomsky (2013: 44, fn34) for relevant discussion.

this is possible, English should be able to use this option as well (i.e., the subject can stay *in situ*).

Another possible labeling strategy is feature sharing, not of ϕ-features, but of θ-role features.

(60) **Feature sharing:**[23]



v^* is a theta-assigner and the subject is a theta-assignee. One might think that the label $<\theta,\theta>$ is possible. However, theta roles are not unique to Japanese, but any other language including English also has to satisfy some version of the theta criterion. Again, English should be able to use this option.[24]

Perhaps, it is fine for English to label the structure as v^*P with the two options that we discussed above (i.e. multiple transfer/θ sharing). But in the TP structure, T is, by definition, a weak head, which results in failure to label TP when the subject does not move to [spec,TP]. If feature sharing-type labeling takes place with {subject,v^*P}, the extraction from the structure makes the label of the structure v^*. Thus, in either analysis, the labeling problem at the v^*P level does not arise in English. In Japanese, the movement option does not help, since movement to [spec,TP] or [spec,CP] will create the unlabeled {XP,YP} structure.

[23] Another possibility is that there is functional projection higher than v^* that assigns a theta role for the external argument. See Funakoshi (2009) regarding this approach.

[24] Note that English also allows an in-situ subject in expletive constructions.

i. Which candidate$_i$ were there [posters of t$_i$] all over town? (Lasnik & Park, 2003, 651,(5a))

ii. *Which candidate$_i$ were [posters of t$_i$] all over town? (Lasnik & Park, 2003, 651,(56))

Lasnik & Park (2003) (see also Merchant 2001) argues that when an expletive is inserted, the subject can stay *in situ*, which makes it possible to extraction from the subject in (i).

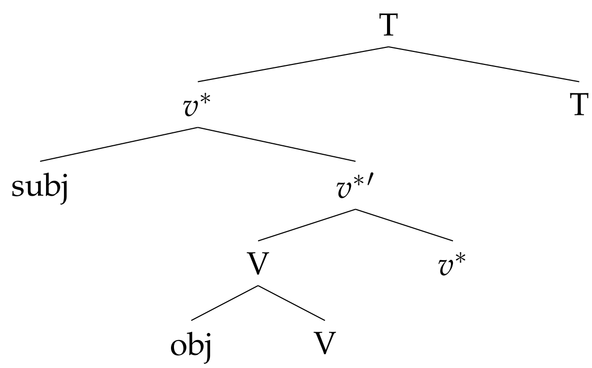
2.6.2 CP Domain

Let's continue the derivation, assuming that the {subj,v^*P} structure is labeled by v^*, as suggested by the analyses considered above.

(61) a. Merge (T, v^*P) = {T,v^*P}

b. MS finds T.

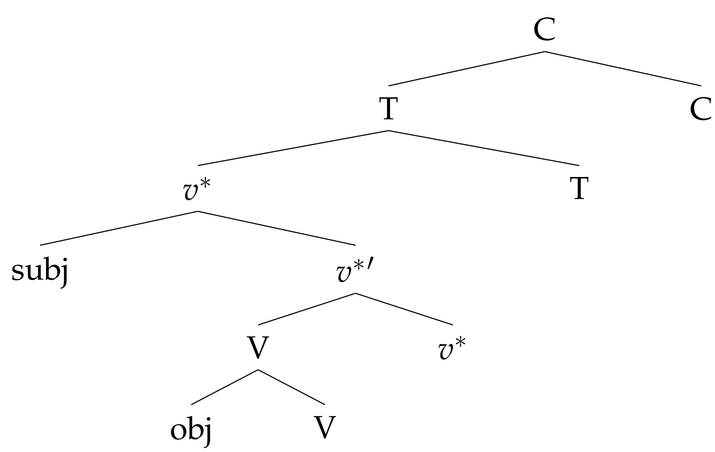
c.



(62) a. Merge (C, TP) = {C, TP}

b. MS finds C.

c.

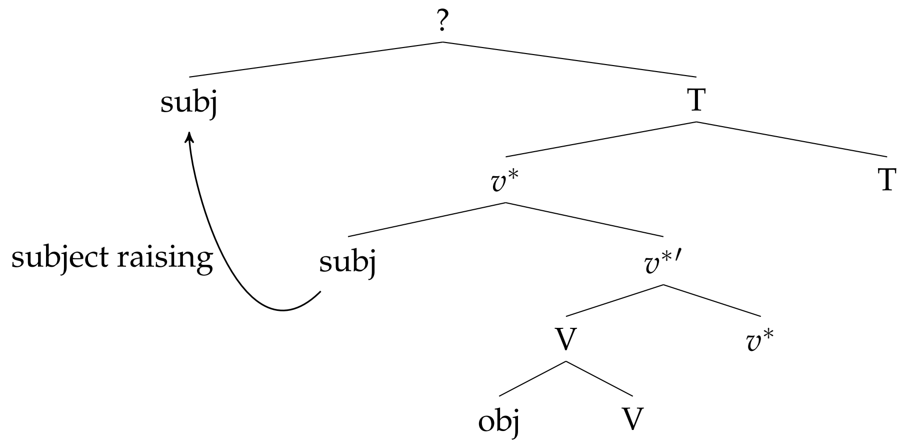


Even if [spec,TP] is available for Japanese, it results in an unlabeled {XP,YP} structure. Supposing the subject moves to [spec,TP], what would be the candidate for labeling?

(63) a. {T, {subj, {v^*, {V, obj}}}}

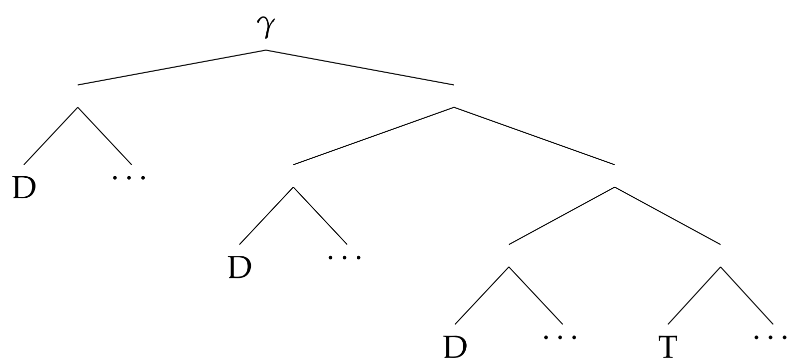
b. IM of subj

c.



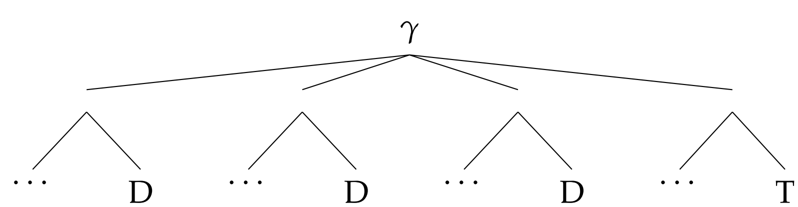
Instead of assuming Saito's (2016) λ-feature, Sorida (2014) suggests that the Case-feature can be a shared label,[25] assuming that the structure-building operation can be more than binary. Considering multiple subject constructions in Japanese, if the structure is binary, the labeling algorithm wrongly predicts that the whole structure γ is labeled by D as represented in (64).

(64)



If we adopt Sorida's (2014) analysis, the structure becomes the following.

(65)



[25]Notice that Saito (2016) also suggests a possibility that the Case-feature can be a label in his paper.

Here, minimal search finds all heads (e.g., all Ds and T). Sorida (2014) assumes that T has a valued Case-feature, whereas D has an unvalued Case-feature. Then he assumes that there is a ϕ-agreement parameter which states that English has ϕ -features, whereas Japanese does not (see also Sorida 2017). As a consequence, he argues that ϕ-feature agreement blocks multiple subjects in English. In Japanese structures such as (65), Sorida argues that valuation of all Ds can take place since T is also found by minimal search. Thus, if we assume Case-feature sharing takes place in Japanese, the subject raising to [spec,TP] is a legitimate derivation, which does not cause a labeling problem, which seems to be incompatible with Saito's (2016) analysis, since in his analysis, the case marker is the locus of λ-features.

Going back to the CP structure, our concern is how to label the whole CP. The issue in (51a) (§2.5), where CP scrambling always makes unlabeled structures, suggests that at least in the root context, the label is not necessary. In fact, some literature already suggests this possibility.

(66) Labeling in root context: (Goto, 2013a,b)
"Labels are necessary for {XP,YP} structures at intermediate positions, but unnecessary for ones at the edge of root CP."

(67) V2 root context: (Blümel, 2017)

a. declarative root clauses must remain label-less

b. prefield-occupation in V2 languages (i.e., [spec,CP] in the declarative root clause) is one strategy to ensure this

(68) Root and labeling (Miyagawa et al., 2019, 10, (24))[26]
The root clause need not be labeled.

These proposals are compatible with Chomsky (2007, 2008), where the assumption is that labels are for further computations in narrow syntax.

[26] See also (Chomsky et al., 2019, 248) for relevant discussion.

If we assume that root clauses do not have to be labeled, this explains why scrambling of CP to the edge of the sentence is possible, though the intermediate CP scrambling is still unclear.

To summarize the discussions of labeling in Japanese, whether we adopt anti-labeling devices/labeling blockers or not, it seems the labeling problems in Japanese have not been solved in the literature. If we adopt Saito's (2016; 2018) analysis, we need to introduce a λ-feature that seems to function only for labeling. If we don not assume Saito's (2016; 2018) λ-features that make the DP invisible, multiple transfer could resolve an {XP,YP} structure to an {XP,Y} structure. However, as we discussed, why then does English not use this option? More justification is needed. The same is true for the feature sharing approach (e.g., sharing case-features or θ-features). In the case-feature sharing approach, parameters have to be involved, though the status of parameters in the minimalist literature is not clear (see chapter 3). Furthermore, as we discussed in §2.4, the labeling analyses do not seem to say much about phenomena such as improper scrambling, CP remnant movement, and the freezing effect, in Japanese (see §2.5). Given the purpose of this chapter, we will not provide a solution to the labeling issues in Japanese (but see §3.8 for a potential solution to these labeling issues.). Rather, in the next section, we discuss an alternative approach to capture the constraints on movement.

2.7 Constraints on Movement

In this section, I would like to argue that improper scrambling and remnant movement are not regulated by labeling, unlike in the proposals reviewed in the previous sections. Rather, I would like to argue that they are instead regulated by general constraints on movement.

As for improper scrambling, I will argue that the A/A'-distinction matters. As for remnant movement, I will argue that recent developments of regarding the structure-building operation MERGE and Phase theory explain the (il)legitimate movement. In the following section, I will introduce Chomsky's (2021a; 2021b) analysis and extend it to remnant movement and improper movement in Japanese.

2.7.1 MERGE

I adopt Chomsky's (2021a; 2021b) framework and explore the consequences of MERGE. In §2.7.2, I will return to remnant movement, which has quite interesting consequences in connection with Chomsky's (2021a; 2021b) framework and MERGE.

Chomsky (2021a,b) revisits the foundations of the architecture of the core system of language. He argues that Universal Grammar (UG) has to satisfy three conditions; *evolvability*, *learnability*, and *universality*. Evolvability suggests that UG has to be simple, assuming that the language faculty evolved quite recently. At the same time, language has to be learnable through language acquisition. According to Chomsky, evolvability and learnability are in conflict with each other, but due to the structure-building operation Merge, structure dependency does not have to be learned through the data, hence there is no learnability problem. Universality says the core system of language (i.e. the language faculty) is uniform across our species. Berwick & Chomsky (2016) suggest that the variety of languages comes from externalization, which is not the core system of language.

Chomsky (2021a,b), then, revisits the structure building operation, since this operation is included in UG. The guiding principle to motivate operations in narrow syntax is called *the strong minimalist thesis*, which "sets as an ideal that all linguistic phenomena can receive genuine explanation in this broader sense (Chomsky, 2021b, 12)." SMT also serves as a disciplinary function (69a) and an enabling function (69b).

(69) Strong Minimalist Thesis (SMT)

a. SMT's disciplinary function: it restricts the mechanisms that are available for description of language (Chomsky, 2021a).

b. SMT also serves an enabling function: it provides options and systems for language that would have no reason to exist if language did not abide by the SMT (Chomsky, 2021a).[27]

SMT is a guideline for constructing a better theory of language: we're constructing a minimal theory of linguistic system (, but not, for example, a proof theory).

Another important notion is *Resource Restriction*. This third factor principle is an overarching principle that the human brain is subject to. Since the computation takes place in a finite domain (i.e., the brain), the structure of the brain limits the information that we can deal with (cf. Fong et al., 2019).

(70) Resource Restriction (RR)

RR restricts the accessible items in the workspace:

a. Phase Impenetrability Condition (PIC, cf. Chomsky 2000, Chomsky 2001)

A complement of a phase head becomes inaccessible (see §2.7.2)

b. Minimal Search (cf. Chomsky, 2013, 2015)

An operation "searches as far as the first element it reaches and no further" (Chomsky, 2021b, 18)

c. Accessible terms are elements in the workspace (not only members of WS, but terms of WS are also accessible for MERGE)

"X is a term of Y if X is a member of Y or a member of a term of Y" (Chomsky, 2021b, 17)

[27] In his other work, Chomsky (2000) formulates SMT as "Language is an optimal solution to legibility conditions (Chomsky, 2000, 96)," where "legibility conditions" can be read as interface conditions. See Freidin (2021) for details.

In the minimalist literature, we already assume that there are mechanisms to minimize the domain that narrow syntax operates on, such as phase theory and minimal search. Based on these SMT-RR principles, Merge is reformulated as follows.

(71) a. WS = {P, Q, . . . }

b. MERGE(P,Q,WS) = { {P,Q}, . . . } = WS'

The operation MERGE applies to the workspace where syntactic objects and lexical items are available. What MERGE does is create a new accessible element and add it to the workspace. Based on RR, *Minimal Yield* is assumed as follows.

(72) Minimal Yield (Chomsky, 2021b)

MERGE adds only one accessible element to the workspace.

When MERGE applies to P and Q (the elements in the workspace), and {P,Q} is formed, the set {P,Q} is the only element that is added to the workspace. Moreover, when the workspace is updated by MERGE, which adds a new accessible element {P,Q} nothing should be deleted or added in ". . . " in (71b), where MERGE does not apply. This simply states that accessible items never disappear, and the computation does not suddenly delete or add something without applying MERGE (Chomsky, 2019b).

This SMR/RR-based MERGE only yields Internal MERGE (IM) and External MERGE (EM).

A case of EM is shown below.

(73) a. WS_1 = { a, b }

b. MERGE(a, b, WS_1) = { {a,b} } = WS_2

The original definition of Merge (Chomsky, 1995) includes the operations *Remove* and *Replace*. Namely, remove *a* and *b*, and replace them with a new syntactic object. Under MERGE (Chomsky, 2019b, 2021a,b), *a* and *b* are removed from the workspace by virtue of *Minimal Yield*; MERGE can add only one new element, namely {a,b}. Notice that *a* and *b* in {a,b} are accessible.

Suppose we don't remove anything. Suppose MERGE combines P and Q, as shown in (74). After MERGE applies to the workspace, the updated workspace includes the identical copy P as a member (74a) in addition to the P in {P,Q}. MERGE can proceed and builds a structure on the structure that includes P (74b). Suppose MERGE applies recursively to the structure arbitrarily, and creates a new object Y.

(74) a. MERGE(P,Q) → {P,Q}, P, Q

b. build up structure with P such as [. . . [Z . . . [. . . P]]] = Y

c. WS = [P, Q, Y, {P,Q} . . .]

d. MERGE(P,Y) = [{P,Y}, P, Q, Y, {P,Q} . . .], where Y includes P.

Since some *P*s are members of the workspace, MERGE can apply to P and Y, which is another member of WS as the result of EM (74c). The new accessible element {P,Y} includes an identical copy P, as a term of the workspace. This situation yields lethal ambiguity and allows derivations that violate any kinds of locality constraints. Thus, *Remove* is needed to remove P and Q from the workspace. Again, because of the property of RR, we don't have to assume the operation Remove, because of Minimal Yield.

Let's move on to the IM case.

(75) a. WS_2 = { {a,b} }

b. MERGE(a_1,{a,b}, WS_2) = { {a_2,{a_1,b}} } = WS_3

MERGE apparently adds two new accessible elements, namely, a_2 and {a_2,{a_1,b}}. That is, there are two copies of *a*, the IMed a_2 and a_1 in the original position.[28] Notice that minimal search (MS (70b), a third factor principle, freely available) finds the higher *a* and the search terminates (i.e., a_1 becomes inaccessible; no further

[28]See Chomsky (2021a,b) for a treatment of copies. See also §2.7.5 for relevant discussion.

search after MS finds a_2). This is another property of RR (70). Therefore, we add only one new accessible element to the workspace.

Parallel Merge is apparently similar to IM in that two copies are generated by MERGE and both are terms of WS, as shown below.

(76) Parallel Merge (cf. Citko, 2005)

a. WS_1 = { {a,b}, c }

b. MERGE(b,c,WS_1) = { {a,b},{b,c} } = WS_2

The difference between internal Merge and parallel Merge is that in the parallel Merge case, *b* in {a,b} and *b* in {b,c} do not c-command each other, hence, both *b*s are accessible, which yields lethal ambiguity. This is a violation of Minimal Yield since there are multiple accessible items. This indicates that parallel Merge is not a subcase of MERGE. Hence, we need additional mechanisms to generate multi-dominant structures, which violates SMT.[29]

2.7.2 Remnant Movement

There is another interesting set of data where multiple copies may or may not cause a problem for the core system of language. The following examples include remnant movement.

(77) a. * [which picture of t_i]$_j$ do you wonder who$_i$ Mary likes t_j

(Saito, 1989, 187, (17b))

b. [$_{CP}$ [$_{Pred}$ t_i How proud of Bill]$_j$ is [$_{TP}$ John$_i$ t_j]]?

(Takano, 1995, 332, (15))

c. Max asked [how likely to win t_i the race]$_j$ John {expected, believed} Oscar$_i$ to be t_j

[29]See Kitahara & Seely (2021) for illustrations of different kinds of extensions of Merge and demonstrations that MERGE does not yield these illegitimate derivations.

It has been argued that this type of movement is not allowed since it violates the *Proper Binding Condition (PBC)*, which states that a trace must be bound. This condition uniformly disallows remnant movement since the phrase including a trace moves up higher than the phrase that is the antecedent of the trace. Kuno (2001) and Saito (2003) argue that remnant movement is allowed when the trace is created by A-movement, based on the argument in Lasnik (1999) that A-movement does not reconstruct to the original position, hence an A-trace is deleted. Once it is deleted, the remnant phrase does not have a trace.

However, there is both conceptual problem with and empirical evidence against the deletion of A-traces. The conceptual problem is that the narrow syntax needs to have the operation of trace/copy deletion, which is not allowed in the current MERGE system. Nothing can be modified (added/deleted) outside of MERGE.

When linearization takes place after the narrow syntax, the deletion of copies is available. In the case of remnant movement, a c-command relation does not hold between the raised copy and the copy in the original position (now moved further up), which is the reason why remnant movement violates PBC. For linearization purposes, we cannot determine which copy should be deleted because there is no asymmetric c-command relation if we follow Kayne (1994). Furthermore, Fox (1999) shows that A-movement can also reconstruct to the original position in some cases. Considering these issues, we need a principled way to explain the remnant movement.

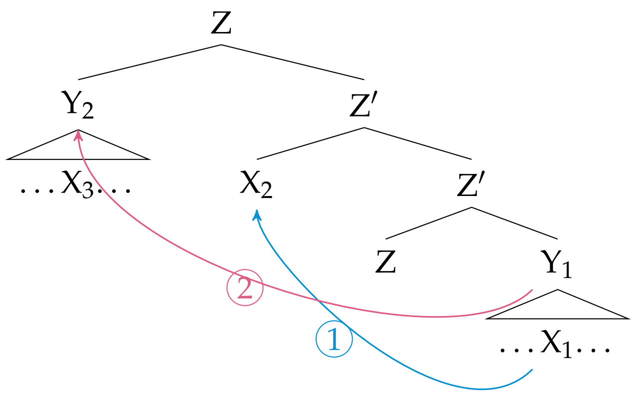
Epstein, Kitahara & Seely (2018a, 2021); Kitahara & Seely (2021) propose that MERGE and Phase Theory give an explanation for the two patterns of remnant movement that contrast in terms of grammaticality in (77).

Let us consider remnant movement in a very schematic way first. In the schema below, X is contained in Y, which is a phrase (78a). Suppose Z, a phase head, is introduced, and then X moves out of Y, yielding copy X_2. Then the phrase Y which includes X moves to the specifier of Z (78b).

(78) a. WS_i= { $\{_Z \ldots X_2 \ldots Z \ldots \{_Y \ldots X_1 \ldots \}\}$ }

b. WS_{i+1} = { { $\{_{Y2} \ldots X_3 \ldots \}, \{_Z \ldots X_2 \ldots Z \ldots \{_{Y1} \ldots X_1 \ldots \}\}\}$ }

c.



In (78c), the higher copy Y_2 c-commands the lower copy Y_1. Thus, the lower copy Y_1 becomes inaccessible for MERGE. However, in terms of the copies of X, the raised X_2 does not c-command X_3 in the higher copy Y_2, and X_3 in the remnant phrase Y_2 does not c-command the raised X_2. That is, both copies (X_3 and X_2) are accessible, since minimal search can find them.

This situation of remnant movement in which multiple copies are accessible is equivalent to parallel Merge.

(79) a. WS_i = { {a, b}, c }

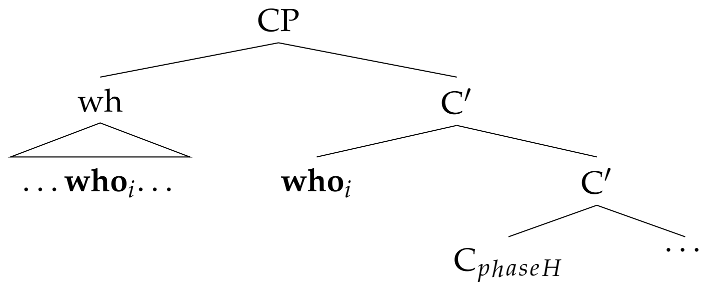
b. WS_{i+1} = { {a, b}, {b, c} }

In (79), both copies of *b* are accessible since there is no c-command relation between these two copies.

Based on this analysis, we could say that a multiple copies situation such as remnant movement or parallel merge results in ungrammaticality. An example of remnant movement is below, where we are looking at the embedded CP.

(80) a. * [which picture of t_i]$_j$ do you wonder who$_i$ Mary likes t_j

b.

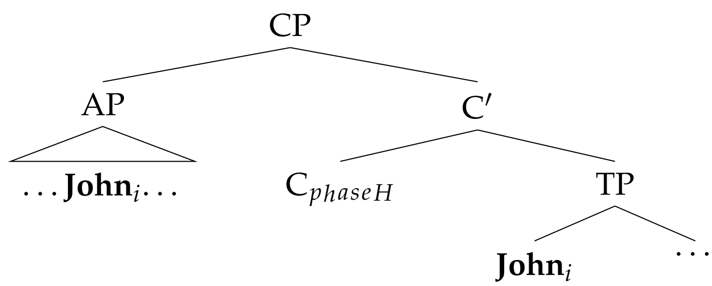


In the embedded CP, two *who*s are accessible, hence the violation of RR.

However, there is a case of proper remnant movement, as shown below.

(81) a. [$_{CP}$ [$_{Pred}$ t_i How proud of Bill]$_j$ is [$_{TP}$ John$_i$ t_j]]?

b.



John moves to [spec,TP], then the remnant phrase moves to the embedded [spec,CP]. Epstein, Kitahara & Seely (2018a, 2021) and Kitahara & Seely (2021) argue that PIC plays a role in this case.

(82) Phase impenetrability Condition (PIC): (cf. Chomsky, 2000, 108)

a. Given HP = [α [H β]], take β to be the *domain* of H and α to be its *edge*

b. In phase α with head H, the domain of H is not accessible to operations outside α, only H and its edge are accessible to such operations

The notion of *phase* is a minimal unit consisting of (i) a phase head, which is typically v* or C,[30] (ii) a complement of a phase head, and (iii) the edge of the phase. The phasehood is defined by a phase head (or unvalued features on a phase head,

[30]The status of D in a nominal phrase is not so clear. See Chomsky (2005); Oishi (2015).

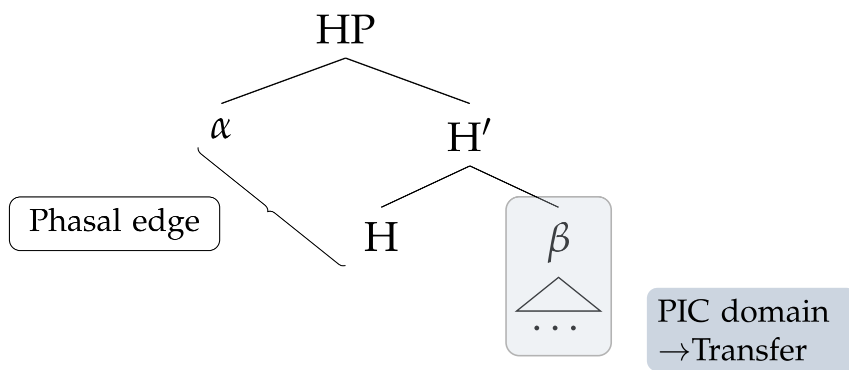


Figure 2.1: Phase Theory and PIC

Chomsky 2015).[31] Once the phase is complete, the complement of the phase head is sent to the interfaces (*Transfer*, Chomsky 2004) (see Figure 2.1 above).

Given the PIC, once syntactic objects are transferred, they are inaccessible for further syntactic operations.[32]

In (81a), *John* in [spec,TP] is in the PIC domain, therefore, it is inaccessible (assuming only the phase head and its edge are accessible for further computation). For the further operations, *John* in the remnant phrase is the only accessible copy. There is no violation of RR. Going back to the ungramamtical example (80a), *who* and the remnant phrase are in [spec,CP] positions (i.e., phase edge positions). Thus, neither of the copies is in the PIC domain, which does not resolve the multiple copy situation. Hence, (80) is predicted to be ungrammatical.

If this approach is on the right track, the generalization is that phase-internal movement such as A-movement followed by remnant movement can be created by MERGE (83). The tree representation of this generalization is shown in (84).[33]

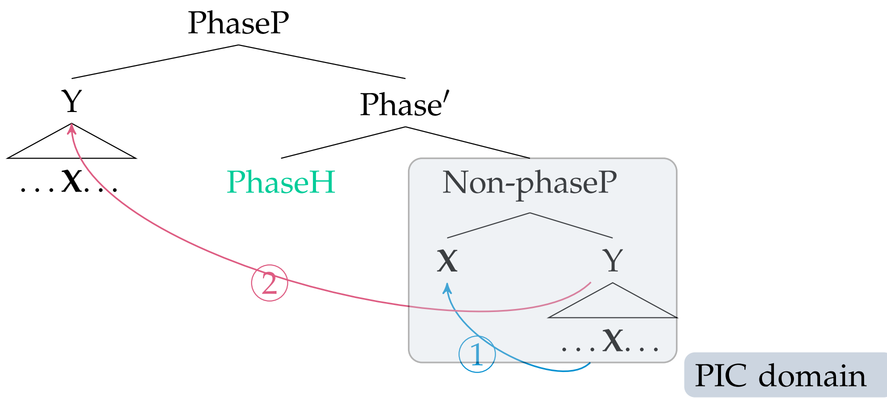
(83) A generalization on Remnant movement : the remnant-creating movement and remnant phrase cannot be both at the phasal edge or phase-internal.

[31] But see also Chomsky (2000, 2001); Grano & Lasnik (2018) for the possibility that the phase is the convergent point of the derivation, where the valuation of unvalued features is completed.

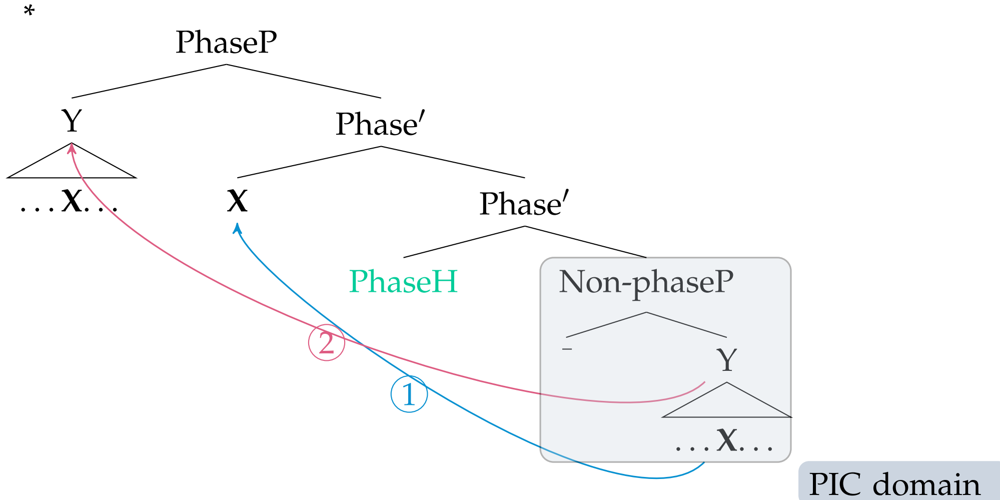
[32] However, see Bošković (2007) for cases where Agree can violate PIC.

[33] Notice that this generalization is derived by MERGE, MY and PIC, not as an independent principle in the Universal Grammar.

(84) a.



b. *

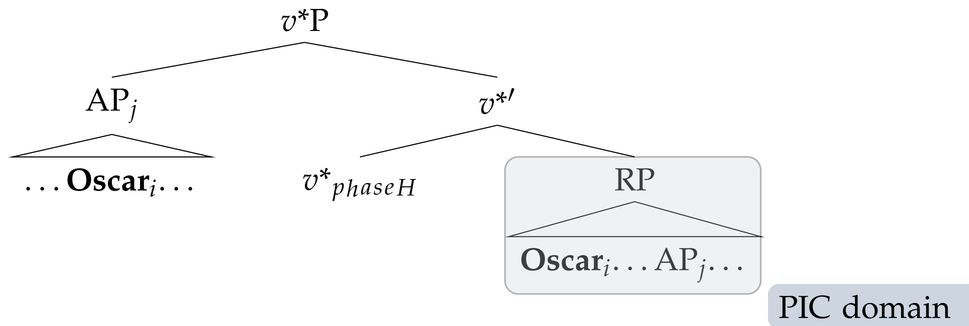


Also, if this is on the right track, what applies to the CP level also applies to the v^{*}P level. The following example tests whether raising-to-object can be followed by remnant movement. The tree representations in (85b) and (85c) show the embedded v^{*}P domain.

(85) a. Max asked [how likely to win t_i the race]$_j$ John [$_{v^*P}$ {expected, believed} Oscar$_i$ to be t_j]

b. [$_{v^*P}$ [$_{AP}$ ... Oscar$_i$ ...]$_j$ [$_{v^{*\prime}}$ v^* [$_{RP}$ Oscar$_i$ [$_{R'}$R (expect, believe) ... AP$_j$]]]]

c.



In (85a), assuming that the landing site of raising-to-object is VP/RP, the raised object is in the PIC domain. Thus, we predict that the derivation is grammatical.

There is another pattern of remnant movement. In German, long-distance scrambling out of control infinitives is possible, as shown in (86a) and passivization of the infinitival clauses is possible, as shown in (86b).

(86) German control infinitives (cf. Grewendorf, 2015)

a. long scrambling out of subject control infinitives

weil Maria den Studenten$_i$ vergeblich [PRO t_i zu küssen] versuchte
since Mary-NOM the student-ACC in-vain [PRO t_i to kiss] tried
'since Mary tried in vain to kiss the student'

b. passivization of infinitival clauses

weil [PRO den Studenten zu küssen] von Maria vergeblich versucht wurde
since [PRO the student-ACC to kiss] by Mary in-vain tried was
'since to kiss the student was tried by Mary in vain'

c. *weil [PRO t_i zu küssen]$_j$ den Studenten$_i$ von Maria t_i vergeblich
since [PRO t_i to kiss] the student by Mary t_j in-vain
versucht wurde
tried was
'since to kiss the student was tried by Mary in vain'

However, the combination of these two (scrambling + passivization), where the passivized infinitival clause subject is a remnant phrase, is impossible (86c).

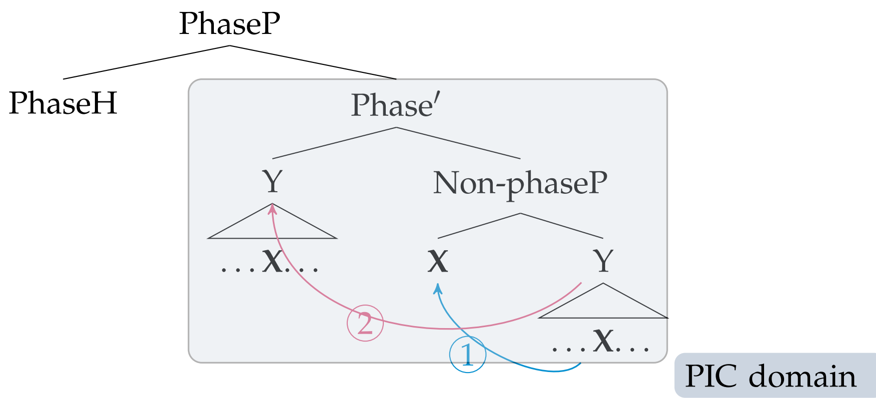
Note that this cannot be explained by (87) or (88), which have been proposed in previous literature, since the movement types are different (i.e., scrambling and passivization).

(87) Unambiguous Domination: Müller (1996)
In a structure ... [$_A$ ... [$_B$ ...] ...] ..., A may not undergo α-movement if B has undergone α-movement

(88) In a derivation yielding the configuration ... [$_X$ ... t_i ...]$_j$ ... Y_i ... t_j ..., movement of Y and movement of X may not be of the same type (cf. Takano, 1994)

If the generalization on remnant movement (83) based on Epstein, Kitahara & Seely (2018a, 2021) and Kitahara & Seely (2021) is on the right track, the following is a possible account, which is an unexplored domain in Epstein, Kitahara & Seely (2018a, 2021): Both types of movement (scrambling and A-movement) are phase-internal movement. The remnant phrase is within the clause because it follows *weil* 'since', but precedes the passivized subject.

(89) *



Thus, the PIC domain includes multiple non-commanding/commanded related copies, which causes lethal ambiguity for language-external systems (e.g., linearization).

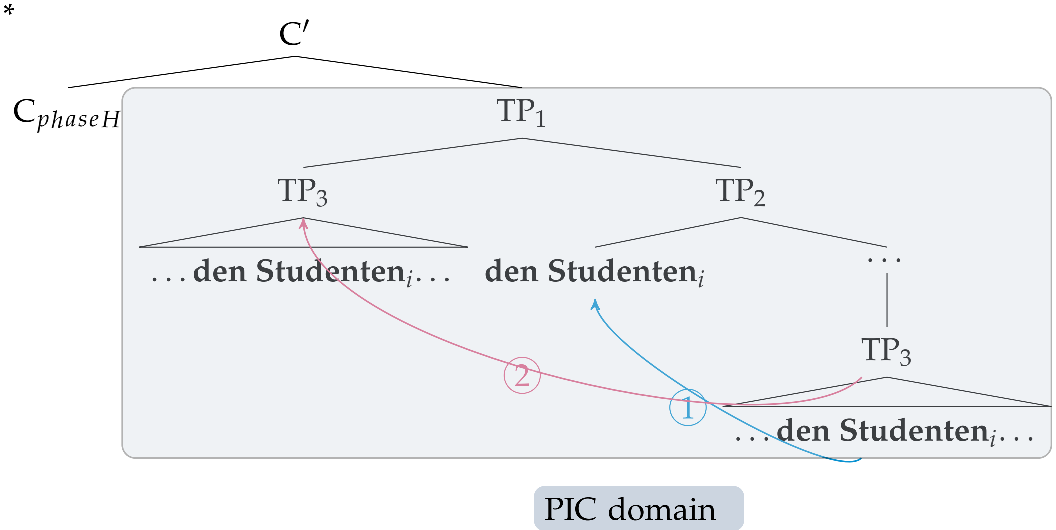
Going back to the concrete example in German regarding remnant movement, the remnant TP movement is followed by scrambling of *den Studenten*. The relevant structure is represented in (91) (some details are omitted).

(90) (=(86c))

*weil [PRO t_i zu küssen]$_j$ den Studenten$_i$ von Maria t_i vergeblich versucht wurde
since [PRO t_i to kiss] the student by Mary t_j in-vain tried was

'since to kiss the student was tried by Mary in vain'

(91) *



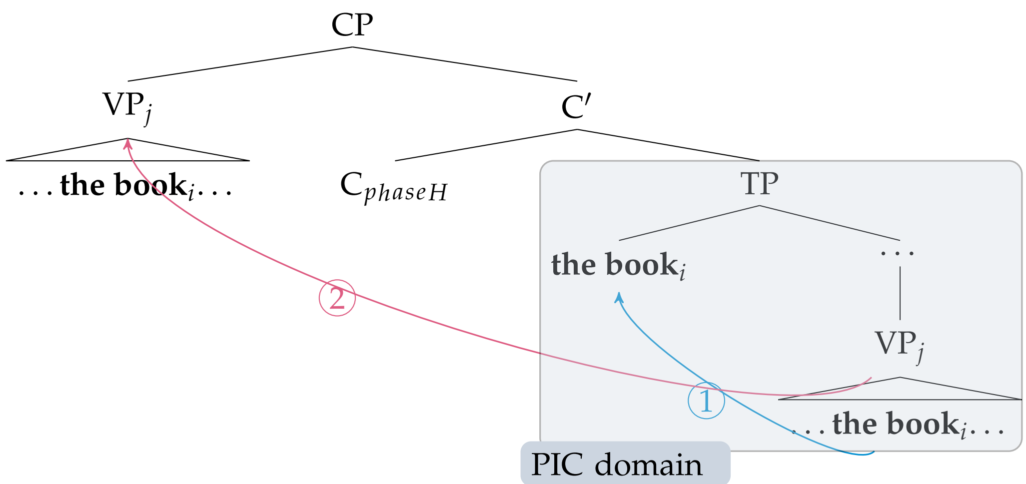
In (91), the scrambled element *den Studenten* is in [spec,TP_2] and, the remnant TP_3 is in [spec,TP_1], so both are in the C phase complement. Thus, *den Studenten* in [spec,TP_2] and *den Studenten* in the remnant TP_3 will be transferred together, resulting in ungrammaticality, as we predict by (76).

Notice that proper remnant movement in German can be captured by the proposed analysis as well.

(92) $[_{VP}$ t_i Gelesen$]_k$ hat $[_{TP}$ $[_{NP}$ das Buch$]_i$ $[_{TP}$ keiner $t_k]]$
$[_{VP}$ t_i read$]_k$ has $[_{TP}$ $[_{NP}$ the book-ACC$]_i$ $[_{TP}$ no one $t_k]]$
'No one has read the book.' (Müller, 1996, 356, (2), with a slight modification)

The remnant-creating phrase (*das Buch*) is scrambled out of VP, then the remnant phrase is topicalized to the Verb-second environment.

(93)



The scrambling of *das Buch* is within the PIC domain and topicalization of the remnant phrase is out of the PIC domain, which does not violate RR. Thus, this example is explained by a phase-based MERGE analysis, which follows the generalization on remnant movement (83).

In summary, we observed that Epstein et al. (2018a; 2021) and Kitahara & Seely's (2021) analysis of remnant movement based on MERGE + RR + PIC correctly explains the sets of relevant data. Notice that there are other kinds of proposals in the literature on remnant movement, which adopt extra assumptions. Under an SMT + RR approach, the mechanism that we can use is very restricted.

An unclear issue is the status of RR and MERGE. RR is a condition on the application of MERGE, otherwise the lethal ambiguity problem arises. At the same time, the proper cases of remnant movement suggest that MERGE can yield the remnant movement (because it is an instance of IM) and could violate RR until PIC applies. However, once PIC appplies, and MERGE checks the workspace, there is no lethally ambiguous situation anymore. On the other hand, in cases of the improper remnant movement and parallel Merge (and other kinds of extensions of Merge), the lethally ambiguous situation is not solved by PIC or minimal search. Hence, further operations can target multiple identical copies, which we want to avoid (otherwise this would allow all kinds of locality violations, such as, one-fell swoop movement of *wh*-phrases across phases).

One way of avoiding this unclarity would be to say that RR is a principle that is involved in MERGE, but not necessarily a condition on the output of MERGE. For example, in a case of proper remnant movement, after the remnant phrase moves to the phasal edge and other operations such as Agree take place, PIC applies. At this point, the lower copies are unavailable. When MERGE applies to the structure in question, there are no multiple copies. When MERGE finds accessible elements, it only finds a higher copy, which is accessible. Thus, Search for MERGE observes the RR restriction, not MERGE itself.

So far, we have discussed Epstein et al.'s (2018a; 2021) analysis of remnant movement. In what follows, I will show that the previous approaches to the linearization of remnant movement in Japanese do not seem to work, and a MERGE-based system can explain remnant movement in Japanese. Here is a brief overview:

(94) a. Linearization approach 1 (§2.7.3.1): Nunes (2004) assumes Chain Reduction and Formal Feature Elimination, suggesting that in a proper case of remnant movement, when the remnant-creating phase moves to [spec,TP], the Case-feature is checked and deleting such a position becomes more costly than deleting the lower copy. This information does not change after remnant movement. Putting aside the details for now, Nunes's mechanism allows a representation in which the Case-feature-checked position is pronounced and other copies are deleted, based on LCA and economy considerations. However, this approach does not explain German and Japanese examples, since they don not have obligatory movement to [spec,TP] for the nominative case.

b. Linearization approach 2 (§2.7.3.2 and §2.7.3.3): Based on Fox & Pesetsky (2005), Takita (2010) suggests that remnant movement causes a problem for linearization, whereas the licit remnant movement is derived by parametrizing the spell-out domain. He shows that remnant movement is impossible for ECM/small clause cases in Japanese. However, this analysis does not capture Hiraiwa's (2010) examples of subject-raising + remnant movement.

c. Based on Epstein, Kitahara & Seely (2018a, 2021); Kitahara & Seely (2021), the asymmetry between subject-raising and object-raising with remnant movement in Japanese can be captured (§2.7.4).

In the following sections, I will review Nunes (2004) in §2.7.3.1, Fox & Pesetsky (2005) in §2.7.3.2, and Takita (2010) in §2.7.3.3, and show how these theories deal with proper/improper remnant movement.

2.7.3 Linearization and Remnant Movement

Since the structure-building operation Merge does not provide information about linear order (the output of Merge is always an unordered set), some sort of algorithm is needed. The most well-known one is Kayne's (1994) Linear Correspondence Axiom (LCA).

(95) Linear Correspondence Axiom (LCA) (Kayne, 1994; Uriagereka, 1999)

a. Base step: If α c-command β, then α precedes β.

b. Induction Step: If γ c-commands β and γ dominates α, then α precedes β.

According to this approach, linearization is defined by asymmetric c-command.[34] This approach motivates X'-theory and suggests that head-final languages such as Japanese are also SVO in the narrow syntax and the object has to move to up to produce the intended word order.[35] Kayne (2011), furthermore, proposes that the

[34] I will put aside the details, which are beyond the scope of this thesis. See also Abels & Neelman (2012); Sheehan (2013) for issues regarding LCA.

[35] See also Fukui & Takano (1998) for relevant discussion.

generalized pair-Merge is the structure-building operation, forcing the specifier-subject-object order as a canonical structure universally.

Although it is plausible to assume that asymmetric c-command is crucial, LCA seems to apply out of the narrow syntax (phonological component), as Chomsky (1995) suggests. Thus, we will maintain the approach that the narrow syntax only forms the hierarchical structures without information about linear order. Thus, we could adopt LCA as an axiom to fix the linear order at PF or language-external systems (Chomsky, 1995).[36]

2.7.3.1 A Copy Theory of Movement and Remnant Movement: Nunes (2004)

Under the minimalist assumptions, internal Merge does not leave a trace in the lower position. Rather, internal merge yields copies, hence, after the movement that creates the multiple copies takes place, all copies are in a c-command relation; the higher copy c-commands the lower copies, not vice versa (Chomsky, 1995).

Nunes (2004) explores the precise algorithm to guarantee which copy is pronounced based on LCA. Syntactic structures have multiple copies, whereas in general only one copy is pronounced. Nunes (2004) also explains why there are languages where multiple copies are actually pronounced, though this is orthogonal to the purpose of this thesis.

The core idea behind pronouncing one copy is *Chain Reduction*:

(96) Chain Reduction (Nunes, 2004, 27, (44))

Delete the minimal number of constituents of a nontrivial chain CH which suffices for CH to be mapped into a linear order in accordance with the LCA.

Chain Reduction itself does not tell which copy has to be deleted. Nunes (2004) also proposes *Formal Feature Elimination*:

[36] See also Narita (2012, 2014) for an LCA-free syntactic approach.

(97) Formal Feature Elimination (FF-Elimination) (Nunes, 2004, 31-32)

Given the sequence of pair $\sigma = < (F,P)_1, (F,P)_2, \ldots, (F,P)_n >$ such that σ is the output of Linearize, F is a set of formal features, and P is a set of phonological features, delete the minimal number of features of each set of formal features in order for σ to satisfy Full Interpretation at PF.

An assumption here is the checking theory (Nunes, 2004), where a [-interpretable] formal feature becomes invisible at PF when it is checked. When a [-interpretable] formal feature is checked, FF-Elimination does not have to apply. But if there is still such a feature, it will be deleted.

Suppose we have a passive sentence in English.

(98) a. was arrested John

b. <John> was arrested John

c. John was arrested <John> (where < > indicates deletion)

Chain Reduction can delete one of the copies of *John*. If it deletes the higher copy, the sentence becomes (98b), whereas if the sentence deletes the lower copy, it becomes (98c). In (98c), the higher copy is the case-checking position. Thus, FF-Elimination does not need to apply. However, in (98b), the lower copy has a [-interpretable] case-feature. Thus, FF-Elimination has to apply. Therefore, the option (98b) is not available in this case, given the economy considerations. The option (98b) is more costly than (98c).

Let us move on to the remnant movement case in (99). The original position of the phrase α, which is represented as t_j in (99a), is c-commanded by the higher copy in the embedded [spec,CP], which is in a *wh*-feature-checking position. Thus, α in [spec,CP] is going to be pronounced, but we are interested in which copy of *John* is pronounced.

(99) a. I wonder $[_{CP}$ $[_{\alpha}$ how likely to t_i win$]_j$ John$_i$ is $t_j]]$

b. I wonder how likely to <John$_2$> win John$_1$ is

c. I wonder how likely to John$_2$ win <John$_1$> is

If Chain Reduction deletes a copy of *John* in the remnant phrase α, the sentence becomes (99b). If it deletes the other copy, the sentence becomes (99c). The option (99b) is more economical than (99c). When Chain Reduction applies to *John*$_1$ right before the copula, which is a case position, *John*$_2$ in the remnant movement will be pronounced. However, since this copy is not in the checking position which has a [-interpretable] case feature, FF-Elimination needs to apply in (99c). Thus, deletion of *John*$_1$ and FF-Elimination applies to *John*$_2$ in (99c). The option (99b) requires fewer applications of deletion since Chain Reduction takes place only at *John*$_2$ and FF-Elimination does not have to apply to *John*$_1$, since it is in a case position.

Although this approach solves the linearization issue (how to linearize the non-c-commanding/c-commanded copies), it does not work for Japanese or German remnant movement, since these languages do not necessarily apply A-movement to [spec,TP] for case (see Kuroda 1988 for Japanese; Wurmbrand 2006 for German). Although there is no strong case-feature in Japanese and German, scrambling is possible. As we discussed in (92), which repeated in (100) below, some cases of remnant movement are allowed in German, where remnant-creating movement does not have a case-feature. Consider (100), repeated from (92).

(100) (=(92))

$[_{VP}$ t_i Gelesen$]_k$ hat $[_{TP}$ $[_{NP}$ das Buch$]_i$ $[_{TP}$ keiner $t_k]]$
$[_{VP}$ t_i read$]_k$ has $[_{TP}$ $[_{NP}$ the book-ACC$]_i$ $[_{TP}$ no one $t_k]]$

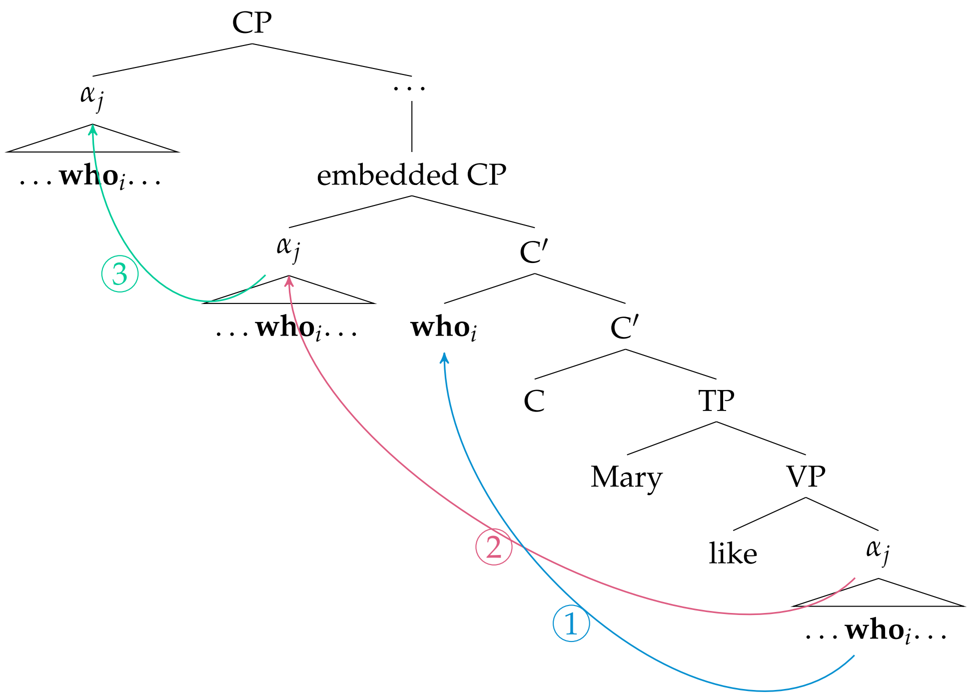
'No one has read the book.' (Müller, 1996, 356, (2))

The remnant phrase VP has a copy of *das Buch*, which is represented as t_i in (92), and it does not c-command the remnant-creating copy in [spec,TP]. Also, both copies *das Buch* already have interpretable features since case can be assigned *in situ* (Wurmbrand, 2006).[37] Thus, there is no way of determining which pattern of deletion is more economical. Furthermore, the improper remnant movement seems to predicted to be grammatical in Nunes' (2004) analysis.

(101) a. * which picture of do you wonder who Mary likes?

b. wonder [$_{CP}$ [$_{\alpha}$ which picture of who$_i$]$_j$ who$_i$ C [$_{TP}$ Mary … t$_j$…]

c.



Who moves from the original position, the complement of VP to the embedded [spec,CP]. Then the remnant phrase α moves to the embedded [spec,CP], higher than the remnant-creating phrase. In the embedded CP, the remnant-creating phrase *who* is in [spec,CP], and the relevant feature is checked with C, whereas the remnant phrase α does not check its feature at the embedded CP, but at the matrix clause. Thus, the remnant-creating phrase *who* is pronounced at the embedded [spec,CP] position and other copies of *who* are deleted. Then, the remnant phrase is pronounced at the matrix clause. It seems the problem regarding which copy

[37] We will discuss more details of Japanese in §2.7.3.3.

should be pronounced does not arise, though the sentence is ungrammatical.

Although Nunes' (2004) analysis predicts most of the linearization correctly, it does not correctly predict the result of proper remnant movement of German and improper remnant movement in English. Thus, I do not adopt Nunes' (2004) analysis of remnant movement in this thesis, and instead seek an alternative approach to capture remnant movement. In the next two subsections, I will introduce an alternative approach to remnant movement in terms of linearization and argue against this approach. §2.7.3.2 introduces Fox & Pesetsky (2005), and based on this linearization model, §2.7.3.3 discusses Takita's (2010) analysis of Japanese remnant movement.

2.7.3.2 Cyclic linearization: Fox & Pesetsky (2005)

Fox & Pesetsky (2005) propose a cyclic linearization algorithm that applies phase by phase (i.e., multiple spell-out). Under this proposal, the order within the spell-out domain should be preserved for further spell-out domains.

(102) Order Preservation (Fox & Pesetsky, 2005, 6)
Information about linearization, once established at the end of a given Spell-out domain, is never deleted in the course of a derivation.

Suppose D is the spell-out domain with the order of X, Y, and Z as follows.

(103) Spell-out of D: [$_D$ X Y Z]

a. ordering
   i. X $<$ Y
   ii. Y $<$ Z ("$<$" refers to "precedence", meaning "Y $<$ Z" is read as "Y precedes Z")

Within the domain D, X precedes Y, and Y precedes Z.

To implement the idea that movement affects phonology, Fox & Pesetsky (2005) assume the following statement.

(104) The relation "<"

An ordering statement of the form $\alpha < \beta$ is understood by PF as meaning that the last element dominated by α and not dominated by a trace precedes the first element dominated by β and *not dominated by a trace*. (Fox & Pesetsky, 2005, 10, (11))

The next example is movement of X, which moves from the phasal edge to the next phase.

(105) Movement of X (leftward movement from a left-edge position) (Fox & Pesetsky, 2005, 11, (13))

a. $[_{D'} \ldots \text{X}\ \alpha\ [_{D}\ \text{t}_X\ \text{Y Z}]]$

b. Ordering: $\text{X} < \alpha$ $\quad$ $\text{X} < \text{Y}$

$\alpha < \text{D} \rightarrow \alpha < \text{Y}$ $\quad$ $\text{Y} < \text{Z}$

Under this scenario, X precedes Y and Y precedes Z. Also, X precedes α, meaning that the last element of X precedes the first element of α. α is pronounced before Y. Thus, the linear order becomes $\text{X} < \alpha < \text{Y} < \text{Z}$, which satisfies the order preservation (102) and the relation "<" (104).

The next example is movement from a non-phasal edge to the next phase.

(106) Movement of Y (leftward movement from a non-left-edge position)(Fox & Pesetsky, 2005, 11, (14))

a. $^*[_{D'} \ldots \text{Y}\ \alpha\ [_{D}\ \text{X}\ \text{t}_Y\ \text{Z}]]$

i. Ordering: $\text{Y} < \alpha$ $\quad$ $\text{X} < \text{Y}$

ii. $\alpha < \text{D} \rightarrow \alpha < \text{X}$ $\quad$ $\text{Y} < \text{Z}$

Y moves to the next phase. The result is that Y precedes α and α precedes X. This creates a contradiction from what we have from the domain D. In D, the order was

X < Y < Z. Thus, we have an order where Y precedes Y and X precedes Y. Thus, this linearization approach forces successive cyclic movement to the phasal edge and does not allow the non-phasal element to move out of the spell-out domain.

2.7.3.3 Cyclic Linearization and Remnant Movement in Japanese: Takita (2010)

Adopting Fox & Pesetsky (2005), Takita (2010) argues that remnant movement is a linearization problem. The examples in Japanese are as follows.

(107) PBC effect on Japanese scrambling

a. *[$_{CP}$ Hanako-ga t$_i$ i-ru to]$_j$ [$_{PP}$ Sooru-ni]$_i$ Taroo-ga t$_j$
[$_{CP}$ Hanako-NOM t$_i$ be-PRES that]$_j$ [$_{PP}$ Seoul-in] Taroo-NOM t$_j$
omottei-ru (koto)
think-PRES (fact)
'(lit.) [That Hanako lives t$_i$]$_j$ [in Seoul]$_i$, Taroo thinks t$_j$'

b. *[PRO t$_i$ ik-u koto]-ga$_j$ [$_{PP}$ Sooru-made]$_i$ Taroo-ni t$_j$
[PRO t$_i$ go-PRES fact]-NOM [$_{PP}$ Seoul-to]$_i$ Taroo-to t$_j$
meizi-rare-ta (koto)
order-PASS-PAST (fact)
'(lit). [To go t$_i$]$_j$, [to Seoul]$_i$, was ordered Taroo t$_j$' (Takita, 2010, 35, (49))

In both cases, the PP moves out of the embedded clause and the remnant phrase moves up. The possible ordering processes are as follows. In (107a), the first spell out domain is the embedded *v*P, and the PP remains there. In (107b), on the other hand, the PP moves to the edge of the embedded *v*P, the phasal edge. Both patterns are possible. The point is that the PP 'in Seoul' precedes the verb *-i* 'be'.

(108) Possible ordering statements for (107a)

a. *Sooru-ni* stays *in situ* $\rightarrow$ Spell-out of the embedded vP

$[_{vP}$ Hanako-ga $[_{VP}$ Sooru-ni i]]

Ordering Table, *Hanako-ga* $<$ *Sooru-ni* $<$ i(be)

b. Movement of *Sooru-ni* $\rightarrow$ Spell-out of the embedded vP

$[_{vP}$ Sooru-ni$_i$ Hanako-ga $[_{VP}$ t$_i$ i]

Ordering Table, *Sooru-ni* $<$ *Hanako-ga* $<$ *i* (be) (Takita, 2010, 36, (50))

However, when remnant movement takes place and the root CP is spelled out, an ordering conflict arises: *Hanako-ga* $<$ i(be) $<$ C $<$ ***Sooru-in*** $<$ *Taroo-ga* $<$ *omotte*. In this order **Sooru-ni** follows *be*, whereas in (108), ***Sooru-ni*** precedes *be*.

(109) a. Ordering Table at embedded vP for (108)

i. Option 1 (108a): *Hanako-ga* $<$ **Sooru-ni** $<$ ***i*** (be)

ii. Option 2 (108b): ***Sooru-ni*** $<$ *Hanako-ga* $<$ ***i*** (be)

b. Ordering Table at the root CP (the final order) for (108)

Hanako-ga$<$ ***i*** (be) $<$ C $<$ ***Sooru-in*** $<$ *Taroo-ga* $<$ *omotte*

Thus, the phrase *Sooru-ni* has to precede and follow the verb *i* 'be' simultaneously in (107a), which is ungrammatical.

The same analysis applies to (107b).

(110) Possible ordering statements for (107b)

a. *Sooru-made* stays *in situ* $\rightarrow$ Spell-out of the embedded vP

$[_{vP}$ PRO $[_{VP}$ Sooru-made ik]]

Ordering Table, *Sooru-made* $<$ *ik*

b. Movement of Sooru-ni Spell-out of the embedded v^*P

$[_{vP}$ Sooru-made$_i$ PRO $[_{VP}$ t$_i$ ik]

Ordering Table, *Sooru-made* $<$ *ik* (Takita, 2010, 36, (51)) Again, the ordering conflict will arise. At the root spell-out domain, the verb *ik* 'go' precedes *Sooru-made* 'to Seoul', whereas in the embedded *v*P, *Sooru-made* 'to Seoul' precedes the verb *ik* 'go'.

(111) a. Ordering Table at the embedded *v*P for (107b)

(110a) and (110b): ***Sooru-made*** < ***ik*** (go)

b. Ordering Table at the root CP (the final order) (107b)

ik < *u* < *koto-ga* < ***Sooru-made*** < *Taroo-ni* < *meizi* < *rare* < *ta*

Thus, this conflict makes the sentence ungrammatical.

Takita (2010) also explains licit remnant movement in English and German.

(112) a. [Criticized t_i by his book]$_j$, John$_i$ has never been t_j

b. [How likely t_i win the game]$_j$ is Mary$_i$ t_j?

c. [t_i Zu lesen]$_j$ hat keiner [das Buch]$_i$ t_j versucht
[t_i to read]$_j$ has no.one [the book]$_i$ t_j tried
'No one has tried to read the book'

Takita (2010) proposes that the spell-out domain of *v*P is parametrized.

(113) Spell-out Domain Parameter for *v*P

a. Linearize the whole *v*P, including the elements on its edge, or

b. Linearize the complement of v^0.

Takita (2010) assumes that the Japanese spell-out domain of *v*P includes the phasal edge (113a). In English and German, the spell-out domain of *v*P only includes the complement of v^0 (113b).

For example, in (112a), *John* moves out of VP, so the spell-out domain does not include *John*.

(114) a. Criticized by his book, John has never been.

b. [$_{vP}$ John$_i$ [$_{VP}$ criticized t$_i$ by his boss]]

c. Ordering Table: criticized < by < his < boss

Then *John* moves to [spec,TP].

(115) a. [$_{TP}$ John$_i$ has never been [$_{v^*P}$ t$_i$ [$_{VP}$ criticized t$_i$ by his boss]]]

b. Ordering table: criticized < by < his < boss

The *v*P is then fronted.

(116) a. [$_{CP}$ [$_{vP}$ t$_i$ [$_{VP}$ criticized t$_i$ by his boss]]$_j$ [$_{TP}$ John$_i$ has never been t$_j$]]

b. Ordering Table: criticized < by < his < boss

criticized < by < his < boss < John < has < never < been

Thus, in the linearization process, there is no point where *John* precedes the *v*P. The same also applies to the German example.

Going back to Japanese examples, Takita (2010) argues that remnant movement of ECM/small clause complements is impossible due to the linearization problem. It has been argued since Kuno (1972, 1976) that the subject in the embedded clause moves up to the matrix clause when it gets an accusative case marker (i.e., *Ziroo-o* 'Ziroo-ACC') in ECM.

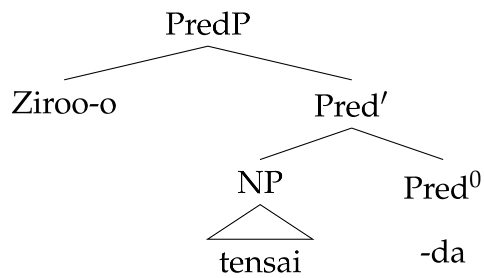
(117) ECM in Japanese

a. Taroo-ga Ziroo-o tensai-da to sinzitei-ru (koto)
Taroo-NOM Ziroo-ACC genius-COP that believe-PRES fact
'Taroo believes Ziroo to be a genius'

b. Taroo-ga Ziroo-o kasikoku omottei-ru (koto)
Taroo-NOM Ziroo-ACC smart consider-PRES fact
'Taroo considers Ziroo smart'

Similar patterns can be observed in small clauses (SC) in Japanese. For the structure of ECM/SC, Takita (2010) assumes the following structure for ECM/SC.

(118) Structure of ECM/SC predicates (Takita, 2010, 57, (92a))



Going back to the linearization problem, the relevant examples are below.

(119) a. *Tensai-da to Tarro-ga Ziroo-o sinzitei-ru (koto)
genius-COP that Taroo-NOM Ziroo-ACC believe-PRES (fact)
'(lit.) To be a genius, Taroo believes Ziroo.'

b. Ziroo-o tensai-da to Taroo-ga sinzitei-ru (koto)
Ziroo-ACC genius-COP that Taroo-NOM believe-PRES (fact)
'(lit.) Ziroo to be a genius, Taroo believes' (Takita, 2010, 51-52, (78))

(119b) shows that the entire PredP is fronted, which does not create a linearization conflict, whereas (119a) shows remnant movement of the ECM complement. Assuming that PredP is the ECM complement, *Ziroo-o* is included in the Spell-out domain of PredP.

(120) *Ziroo-o* $<$ *tensai* $<$ *da*

Although *Ziroo* precedes the predicate *tensai da* 'be a genius' in this spell-out domain, *Ziroo* will follow the predicate at the root CP, which will yield a conflict after the movement of the remnant ECM/SC complement to the edge of the sentence (119a).

2.7.4 An asymmetry between raising-to-subject and raising-to-object and remnant movement

Although Takita (2010) shows a way of capturing the PBC effect via linearization, there are examples that do not seem to be captured by his analysis.

Hiraiwa (2010) shows the contrast between raising-to-subject and raising-to-object in terms of remnant movement.

(121) Raising-to-subject + remnant movement in Japanese (Hiraiwa, 2010, 141, (16))

a. [$_{TP}$ Ken-ga$_i$ minna-ni [$_{CP}$ t$_i$ baka-da to] omow-arete-iru]
[$_{TP}$ Ken-NOM everyone-DAT [$_{CP}$ t$_i$ foolish-PRES C] think-PASS-PRES]
'Ken is thought to be stupid by everyone'

b. [$_{CP}$ t$_i$ Baka-da to](-wa)$_j$ [$_{TP}$ Ken-ga$_i$ minna-ni t$_j$
[$_{CP}$ t$_i$ foolish-PRES C-TOP [$_{TP}$ Ken-NOM everyone-DAT t$_j$
omow-arete-iru]
think-PASS-PRESS]
'Ken is thought to be stupid by everyone'

(122) Raising-to-object + remnant movement (Hiraiwa, 2010, 141, (18))

a. [$_{TP}$ Ken-ga [$_{v^*P}$ Naomi-o$_i$ kokorokara [$_{CP}$ t$_i$ baka-da to]$_j$
[$_{TP}$ Ken-NOM [$_{v^*P}$ Naomi-ACC really [$_{CP}$ t$_i$ foolish-COP C]$_j$
omot-ta]].
think-PST
'Ken felt that Naomi was pretty.'

b. *[$_{CP}$ t$_i$ baka-da to](-wa)$_j$ [$_{TP}$ Ken-ga [$_{v^*P}$ Naomi-o$_i$
[$_{CP}$ t$_i$ foolish-COP C]-TOP [$_{TP}$ Ken-NOM [$_{v^*P}$ Naomi-ACC
(kokorokara) t$_j$ omot-ta]]
really t$_j$ think-PST]
'Ken really considered Naomi to be a fool.'

Takita's (2010) analysis seems to say nothing about this contrast since the ECM complement is the spell-out domain, which will yield a linearization problem in his analysis, as we discussed in the previous subsection.

Since the embedded clause is PredP in (121b), the ordering table for the spell-out domain of PredP is *Ken-ga* $<$ *baka* $<$ *da*. However, at the root CP, the subject *Ken-ga* follows *baka* $<$ *da*.

(123) a. At the PredP for (121b)

Ordering table: ***Ken-ga*** $<$ ***baka*** $<$ ***da***

b. At root CP for (121b)

Ordering table: ***baka*** $<$ ***da*** $<$ *to(-wa)* $<$ ***Ken-ga*** $<$ *minna-ni* $<$ *omow* $<$ *arete* $<$ *iru*

Thus, an ordering conflict arises in (121b), which wrongly predicts that the sentence is ungrammatical. In (122b), the ordering table for the PredP spell-out domain is *Naomi-o* $<$ *baka* $<$ *da*, but, at the root CP, again, *Nomi-o* follows *baka* $<$ *da*.

(124) a. At the PredP for (122b)

Ordering table: ***Naomi-o*** $<$ ***baka*** $<$ ***da***

b. At root CP for (122b)

Ordering table: ***baka*** $<$ ***da*** $<$ *to(-wa)* $<$ *Ken-ga* $<$ ***Naomi-o*** $<$ *(kokorokara)* $<$ *omot* $<$ *ta*

Again, these two orderings in (124) are in conflict with each other, hence, the sentence in (122b) is ungrammatical. Thus, Takita's (2010) linearlization approach predicts that both examples in (121b) and (122b) are ungrammatical, though (121b) is, in fact, grammatical.

However, once we adopt Epstein et al.'s (2018a; 2021) analysis, these contrasts can be captured. Based on Epstein et al.'s (2018a; 2021) analysis, I generalized the their proposal as follows.

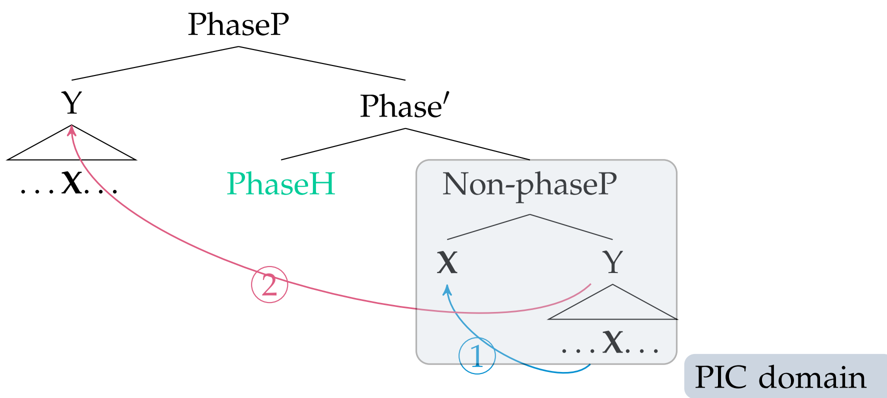
(125) (=(83)) Generalization on Remnant movement: remnant-creating movement and the remnant phrase cannot be both at the phasal edge or phase-internal.

I showed in §2.7.2 that multiple copies which are not in a c-command relation yield a violation of MY since both copies are accessible, which leads to an ambiguity for further computations (i.e., a computation cannot determine which copy to use). In addition to this, phase theory is involved. PIC makes the complement of the phase head inaccessible, which reduces the domain of the computation.

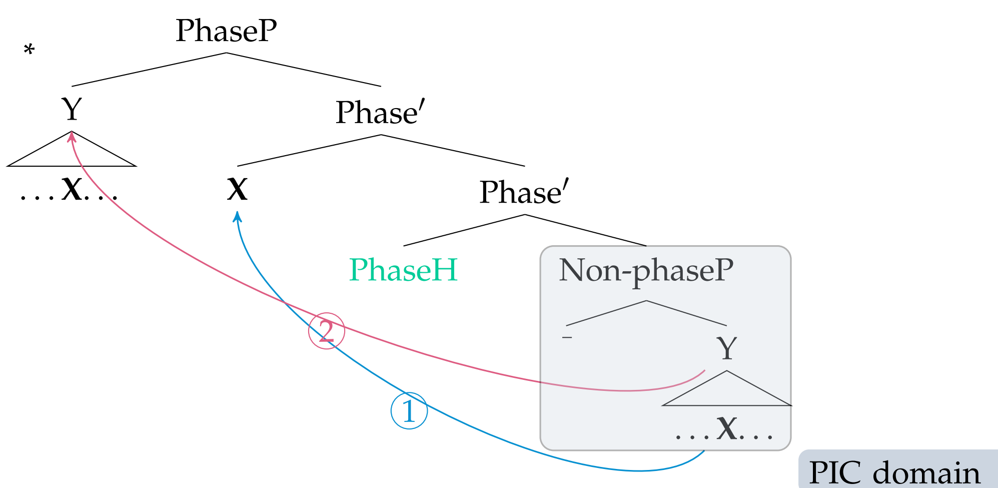
The relevant representation is shown below.

(126) (=(84))

a.



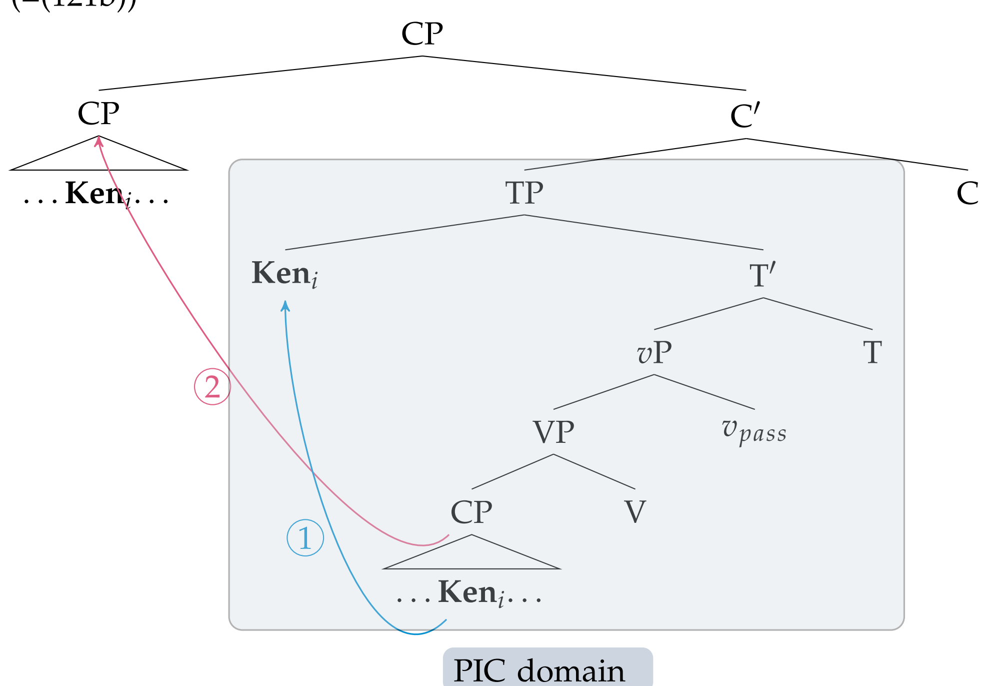
b. *



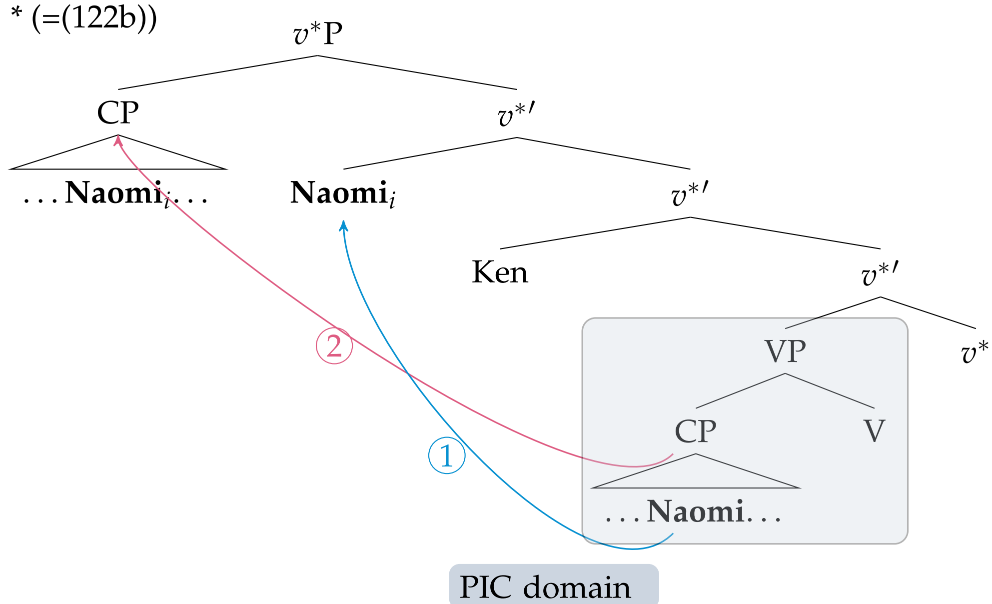
I showed in §2.7.2 that when remnant-creating element and the remnant phrase that includes a copy of the remnant-creating element are both on the phasal edge, MY is violated, since there are two structurally identical copies that are accessible for computation. However, if the remnant-creating phrase is in the domain of PIC, no MY violation takes place. After PIC makes the complement of the phase head inaccessible, the relevant accessible element includes only a copy of the remnant phrase, and the remnant-creating phase is no longer accessible.

Going back to the Japanese remnant movement examples (121b) and (122b), we assume with Tanaka (2002) and Hiraiwa (2010) that raising-to-object in Japanese targets [spec,*v*P], a phasal edge, whereas the subject raising targets [spec,TP]. Thus, the remnant-creating phrase ('Ken-NOM') in (121b) moves to [spec,TP], which is in the PIC domain, whereas the remnant-creating phrase ('Naomi-ACC') in (122b) moves to [spec,*v*P], the edge of the *v* phase. Notice that the matrix verb in the subject raising example (121b) is a passive verb where the phasehood is weak (Chomsky, 2001), whereas the matrix verb in the raising-to-object example (122b) is an active verb. Thus, in (122b), the remnant-creating phrase ('Naomi-ACC') is indeed on a phasal edge.

(127) (=(121b))



(128) * (=(122b))



Thus, the remnant movement can directly target [spec,CP] at the matrix clause (127), whereas the subject *Ken-ga* is located in [spec,TP], the domain of PIC, which makes it inaccessible, so there is no MY violation. On the other hand, in (128), the raised object *Naomi-o* and the remnant CP are at the edge of the matrix *v*P, which causes a MY violation (, that is, there are two copies of *Naomi* at the phasal edge of the *v* phase).

To summarize, we started by exploring linearization approaches to remnant movement (§2.7.3.1). I pointed out that Nunes' (2004) linearization approach cannot explain remnant movement in Japanese and German since the remnant-creating movement is not case-driven, and thus is not subject to FF-Elimination (97). Based on Fox & Pesetsky (2005), Takita (2010) proposes that remnant movement in Japanese can be deduced from the linearization process: If there is a conflict between ordering tables, the derivation will become ungrammatical. In §2.7.3.2, I reviewed Fox & Pesetsky (2005), and in §2.7.3.3, I pointed out that Takita's (2010) analysis wrongly predicts that a derivation in which a remnant-creating phrase raises to the subject position ([spec,TP]), followed by a remnant movement, is ungrammatical. I, instead, have argued that Epstein et al.'s (2018a; 2021) analysis of remnant movement correctly predicts the contrast between (121b) and (122b).

In the rest of this chapter, §2.7.5 explores another aspect of MERGE and accounts for improper movement. §2.8 introduces the status of two islands that can be properly captured by the MERGE framework.

2.7.5 Improper Movement

2.7.5.1 MERGE, Copies and Markovian Derivations

Chomsky (2021a,b) assumes that the derivation is strictly Markovian, indicating that when the derivation is interpreted by a language-external system (the "interpretive system" INT), that system can only see the representation. INT only sees the representation and determines the copy information (e.g., which occurrence of X moves to where? Or are they different elements?), via an operation called *Form Copy* (FC). Thus, INT cannot see whether the identical elements are formed by IM or EM.

Suppose we have the following derivation.

(129) $\{John_1, \{saw, John_2\}\}$

MERGE forms the structure link in (129) and at some point, INT interprets the structure. Since INT can only see this representation, it applies FC and can determine that these two *John*s are copies. Then the lower copy is deleted in the phonological component. However, (130) is ungrammatical.

(130) *John saw

(intended meaning, John saw himself)

Theta theory plays a role here. Based on Chomsky (2021a,b), let us define theta theory as follows.

(131) Theta Theory[38]

a. Theta theory is univocal.

b. An argument cannot be assigned two different roles from one predicate.

c. Each argument has to have either a theta role or a semantic role.

MERGE has to satisfy LSCs such as theta theory. In case of (129), when INT applies FC to the structure, it violates theta theory. Both *John*s are in theta positions and they get different theta roles from a same predicate. $John_1$ is an agent of the event (i.e., John saw someone) and $John_2$ is a theme of the event (i.e., John was seen by the agent).

As for semantic roles, Chomsky (2021b) defines θ-linked positions as follows.

(132) a. X is θ-*linked* to $P(\tau)$ if a copy of X occupies $P(\tau)$.

b. A θ-assigner τ assigns one and only one θ-role elements θ-linked to $P(\tau)$. (Chomsky, 2021b, 26-27)

In addition to a θ-position, Chomsky (2021b) argues that there is a semantic role position where A-movement is relevant. The evidence of a semantic effect comes from an existential presupposition in the subject position (133b).

(133) a. to seem to be intelligent is hard

b. there is a fly in the bottle/a flaw in the proof

c. a fly is in the bottle/*a flaw is in the proof (Chomsky, 2021b, 27)

(133a) is a raising construction where the infinitival clause is raised to the [spec,TP] position and gets a semantic role. However, the subject raised example in (133c)

[38] There is a fundamental question regarding theta theory in recent literature: How is the argument structure realized and what is the source of the argument structure? One approach is what is called the *Constructivist* approach where some researchers argue that syntactic structures are independent from lexical items, suggesting that a verb does not discharge a theta role, but the syntactic structure itself determines the corresponding grammatical properties. See Borer (2003, 2005b,a, 2013); Marantz (1997, 2013); Ramchand (2008); Lohndal (2014); Embick (2015) among many others, See also Lohndal (2019) for an overview of this approach.

(*a flaw is in the proof*) is ungrammatical, showing the lack of an existential presupposition in the subject position in this construction. The main idea here is that a θ-role position is introduced by EM and a semantic role position is created by A-movement (in this case, the [spec,TP] is the semantic role position).

Let us consider the next example (i.e., obligatory control). In (134), $John_1$ is introduced by EM and it gets a theta role from *win*, and moves to $John_2$. $John_3$ is introduced by EM and it gets a theta role from *tried*, and it moves up to $John_4$.

(134) a. John tried to win

b. {John_4, {T, {John_3, {tried, {John_2, {to, {John_1, win}}}}}}}

Since INT can only see the representation, it can determine all occurrences of *John*s are copies. When FC applies to all occurrences of *John*, there is no violation of theta theory since $John_1$ and $John_3$ get theta roles from separate predicates (i.e. *win* and *tried* respectively) (131b). $John_4$ is in a semantic role position, which is θ-linked to $John_3$ and $John_2$ is a semantic role position, which is θ-linked to $John_1$.[39]

To sum up, the system in Chomsky (2021a,b) works in this way: (i) MERGE, which applies to the workspace, forms the structurally identical inscription when IM takes place, which results in copies, i.e., forms IM-configuration when a relation is in a c-command configuration. There are two types of IM-configuration; (a) an IM-configuration that is formed by internal Merge and (b) an IM-configuration that is formed by FC.[40] (ii) MERGE generates the structure derivationally, whereas INT cannot see the derivational history, thus, INT does not know whether the copies are formed by (a) or not, but INT can freely assign them as copies. In the case of

[39]Notice that this captures the fact that the raising construction and the control construction are similar, but not identical (Hornstein, 1999). See more details in Chomsky (2021b).

[40]One might wonder (a) and (b) are identical. However, because of the Markovian nature of the derivation, this could yield a control construction. (b) is an example of a control construction where the surface subject in the matrix clause is merged in the matrix [spec,*v**P]. If INT could see the history of the derivation, control constructions would not exist (assuming the control construction is derived by raising the embedded subject to the matrix clause, just as in the raising construction), which conforms to SMT.

(b), as long as the theta criterion is met, the derivation is interpreted, which yields obligatory control as in (134).

How about the next sentence?

(135) a. *John seems (that) is happy

b. {John$_4$, {T, {seems, {John$_3$, {C, {John$_2$, {is, {John$_1$, happy}}}}}}}}

(135a) is a case of hyper-raising; the subject moves from the finite embedded clause to the matrix subject position (e.g., Ura, 1994). It has been argued in the literature that this type of movement is banned because the derivation includes the movement from an A′-position to an A-position (i.e., movement from *John*$_3$ to *John*$_4$) (May, 1979; Fukui, 1993a; Chomsky, 1995; Obata & Epstein, 2011; Safir, 2019).

Suppose INT applies FC and determines all occurrences of *John* in (135b) are copies. (136) will be the representation.

(136) {John$_4$, {T, {seems, {<John$_3$>, {C, {<John$_2$>, {is, {<John$_1$>, happy}}}}}}}}

As for the theta criterion, *John*$_1$ has a theta role and rest of the copies do not have theta roles. Compared to the obligatory control example in (134), the difference is that (135) includes an A′-position ([spec,CP] in the embedded clause). In (136), there is apparently no problem with CF, though the sentence is ungrammatical (135).

Notice that Chomsky (2019b) discusses the copies and repetitions in terms of c-command. Consider the example (137a). Descriptively speaking, *which boy*$_1$ (a θ position) is internally merged to the embedded [spec,CP$_1$] (*which*$_2$), and *which boy*$_3$(a θ position) is internally merged to the matrix [spec,CP$_2$](*which*$_4$). Again, INT does not have information about the history of the derivation. Thus, INT has to determine which one is a copy or a repetition only using the representation and theta theory. In addition to these, Chomsky (2019b) argues that the c-command relations among copies in (137a) matter.

(137) a. which boy did John ask which boy Bill met?

b. [$_{CP2}$ which boy$_4$ did John [$_{vP}$ ask which boy$_3$ [$_{CP1}$ which boy$_2$ Bill met which boy$_1$]]]

In (137a), *which boy*$_3$, which is in a theta position, c-commands *which boy*$_2$, which is in an A′-position (in the embedded specifier of C$_1$). Chomsky (2019b) argues that in this case, *which boy*$_3$ and *which boy*$_2$ are repetitions.

Let us go back to the hyper-raising example.

(138) (=(135))

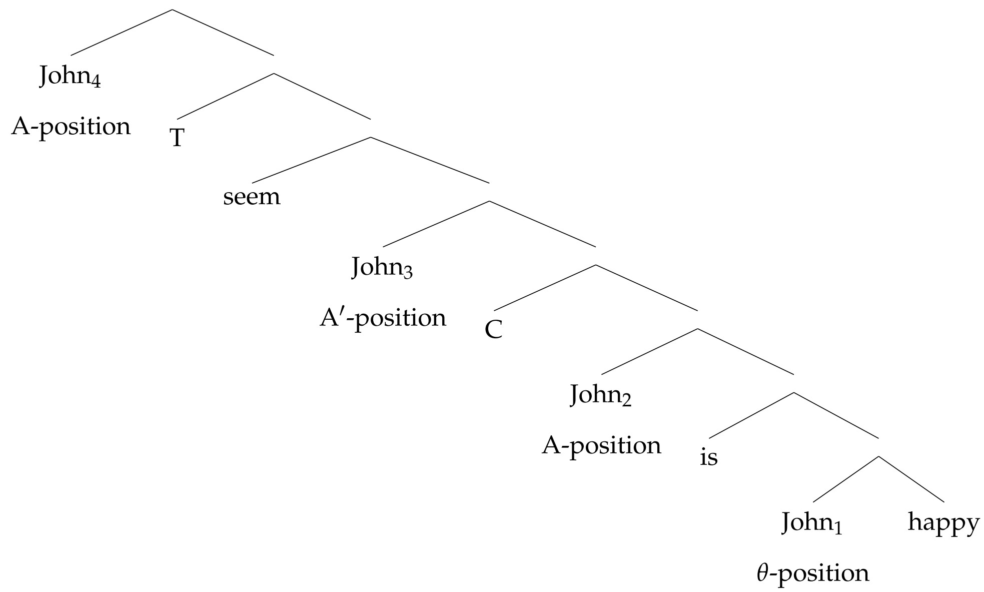
a. *John seems (that) is happy

b. {John$_4$, {T, {seems, {John$_3$, {C, {John$_2$, {is, {John$_1$, happy}}}}}}}

John$_4$ is in an A-position and *John*$_3$ is in an A′-position. *John*$_4$ c-commands *John*$_3$, so they are repetitions. INT checks this (again, the assumption here is that INT cannot see the history of the derivation, but can check the c-command relation representationally) and FC applies only to *John*$_3$, *John*$_2$, and *John*$_1$. Thus, *John*$_3$, *John*$_2$, and *John*$_1$ are copies but *John*$_4$ is not (i.e., it is a repetition). Then, what is pronounced is as follows.

(139) a. *John seems John (that) is happy.

b.



Informally speaking, for the chain (*John*$_3$, *John*$_2$, *John*$_1$), there is no theta violation. However, *John*$_4$ is a trivial chain (i.e., a repetition), and it does not have a theta role, hence, it violates theta theory. Based on the derivation for hyper-raising, the only way of creating copies is (139a). Thus, the hyper-raising derivation in English is never correctly pronounced in the intended way, as in (135a=138a).[41,42]

[41] One might wonder whether to appeal to case theory. If so, we would need to assume the activity condition (Chomsky, 2001), which was abandoned in Chomsky (2007). See Nevins (2005) for details. He deals with hyper-raising without using an activity condition approach. Furthermore, Chomsky (2021a,b) suggests that Case might be outside of the narrow syntax (perhaps, the Agree operation might also be outside of narrow syntax; see also Epstein et al. (2021))). If these are on the right track, it's unclear how the activity condition works in narrow syntax.

[42] The discussion here only applies to the derivation where A-movement is followed by A′-movement. It has been argued that some languages allow hyper-raising (cf. Ura, 1994). These languages have been analyzed using different assumptions; for example, Zulu, one of Bantu languages, allows hyper-raising because the finite embedded CP itself has ϕ-features, whereas it does not allow raising constructions such as *John seems to be happy* (Halpert, 2016, 2018). Since we are interested in an account of improper movement, variation of the phenomenon "hyper-raising" across languages is less relevant here. But see Ura (1994); Bruening (2002); Carstens & Diercks (2010); Petersen & Terzi (2015); Zyman (2017); Fong (2018); Halpert (2016, 2018); Pires & Nediger (2018) for the details of various hyper-raising phenomena across languages.

2.7.5.2 Improper Movement in Japanese

Finally, we come back to improper scrambling in Japanese. I will argue that improper scrambling is banned by CF and theta theory, whereas there is no such violation for proper scrambling.

As we discussed in §2.4.1, the crucial difference between (140b) and (142b) is whether the embedded clause is finite or not. I will assume here that scrambling out of the finite clause will be A′-type movement (Saito, 1992; Sakai, 1996).

(140) a. Masao-ga Kumiko-ni [Takashi-ga Boston-e it-ta-to]
Masao-NOM Kumiko-DAT [Takashi-NOM Boston-to go-PAST-C]
it-ta.
say-PAST
'Masao told Kumiko that Takashi went to Boston.'

b. ?*Masao-ga **Boston-e**$_i$ Kumiko-ni [Takashi-ga t$_i$ it-ta to]
Masao-NOM **Boston-to**$_i$ Kumiko-DAT [Takashi-NOM t$_i$ go-PAST C]
it-ta.
say-PAST
'Masao told Kumiko that Takashi went to Boston.'

(Sakai, 1996, 129, (10))

The analysis is as follows. In (140b), PP moves to [spec,CP] in the embedded clause, which is a A′-position. Then it moves to the A-type position (i.e. [spec,*v*P] (Sakai, 1994, 1996)). This is the improper movement. When INT applies FC, it regards PP in the [spec,*v*P] in the matrix clause and PP in the embedded [spec,CP] as repetitions since PP in the matrix [spec,*v*P] is in an A-position, which c-commands PP in the embedded [spec,CP], which is in an A′-position. Thus, after FC takes place, the result is (141).

(141) Masao-ga [$_{vP}$ **Boston-e** Kumiko-ni [$_{CP}$ **Boston-e** Takashi-ga it-ta to it-ta]].

This is not what we want, and this is what we saw in improper movement in English in (139a). Thus, the derivation in (140b) is degraded. How about the grammatical case?

(142) a. Masao-ga Kumiko-ni$_i$ [PRO$_i$ Boston-e iku-yooni] susume-ta.
Masao-NOM Kumiko-DAT$_i$ [PRO$_i$ Boston-to go-so-that] advise-PAST
'Masao advised Kumiko to go to Boston.'

b. Masao-ga **Boston-e**$_j$ Kumiko-ni$_i$ [PRO$_i$ t$_j$ iku-yooni] susume-ta.
Masao-NOM **Boston-to**$_j$ Kumiko-DAT$_i$ [PRO$_i$ t$_j$ go-so-that] advise-PAST
'Masao advised Kumiko to go to Boston.' (Sakai, 1996, 130, (11))

In this case, the scrambling is out of a non-finite clause, thus there is no [spec,CP] position that serves as an A′-type position. Thus, this scrambling does not involve an A′-position at all (Sakai, 1996), and is acceptable.

Notice that Nemoto (1995) shows the difference between the embedded finite clause and non-finite clause in term of A-/A′-scrambling (cf. Tada, 1993). In English, A-movement remedies a strong crossover (SCO) violation as shown in (143b).

(143) a. *[Whose$_i$ teacher]$_j$ did he$_i$ hit t$_j$?

b. [Whose$_i$ teacher]$_j$ t$_j$ seems to him$_i$ t$_j$ to be intelligent.

The same is true for Japanese, as shown in (144). (144) shows passivization, which remedies SCO violations, whereas (145) shows an SCO violation after the scrambling of the object to the front of the sentence.

(144) [dare$_i$-no sensei]-ga soitu$_i$-ni t$_j$ syookais-are-ta no
[whose$_i$ teacther]$_j$-NOM HE$_i$-to t$_j$ introduce-PASS-PAST Q
'Whose teacher was introduced to him?' (Nemoto, 1995, 260, (6b))

(145) *[dare$_i$-no sensei]$_j$-o soitu$_i$-ga Taroo-ni t$_j$ sookaishita no
[whose$_i$ teacher]$_j$-ACC HE$_i$-NOM Taro-DAT t$_j$ introduced Q
'Whose teacher did he introduce to Taroo?' (Nemoto, 1995, 261, (7b))

Bearing this in mind, let us consider scrambling out of a finite clause.

(146) a. *Hanako-wa soitu$_i$-ni [Taroo-ga dare$_i$-no sensei-o nagutta to] itta no
Hanako-TOP HE$_i$-DAT [Taroo-NOM whose$_i$ teacher-ACC hit C] say-PAST Q
'Hanako said to him that Taro hit whose teacher'

b. *Hanako-wa dare$_i$-no sensei-o$_i$ soitu$_i$-ni [Taroo-ga t$_i$ nagutta to] itta no
Hanako-TOP whose teacher-ACC HE-DAT [Taroo-NOM t$_i$ hit C] say-PAST Q
'Hanako, whose teacher, said to him that Taro hit' (Nemoto, 1995, 265, (15))

In (146b), the SCO violation is not remedied by scrambling. This indicates that scrambling out of a finite clause is A'-type movement, which yields an SCO violation. However, Nemoto (1995) shows that when the embedded clause is a control clause, SCO violation (147a) can be remedied by the scrambling out of the non-finite clause, as in (147b).

(147) a. *Hanako-wa soitu$_i$-ni [PRO dare$_i$-no sensei-o naguru yoo(ni)] tanonda no
Hanako-TOP HE$_i$-DAT [PRO whose$_i$ teacher-ACC hit C] ask-PAST Q
'Whose teacher did Hanako ask him to hit?'

b. Hanako-wa [dare$_i$-no sensei-o]$_j$ soitu$_i$-ni [PRO t$_j$ naguru yoo(ni)] tanonda no
Hanako-TOP [whose$_i$ teacher-ACC]$_j$ HE$_i$-DAT [PRO t$_j$ hit C] ask-PAST Q
'Hanako, whose teacher, asked him to hit' (Nemoto, 1995, 266,(16))

At least, Nemoto's argument, which is based on Tada (1993), shows that scrambling out of the non-finite clause is A-type movement when the scrambling targets the intermediate position of the matrix clause, such as in (147b).

Based on Nemoto's arguments reviewed here, the scrambling to the matrix post-subject position from the embedded clause could be A or A′-movement. When the embedded clause is a non-finite clause, the movement to the matrix post-subject position becomes A-movement. When the embedded clause is a finite clause, the movement to the matrix post-subject position becomes A′-movement. However, the latter argument is not compatible with the examples that we observed in (148).

(148) (=(30b))

*[$_{S1}$ Masao-ga **Boston-e**$_j$ Kumiko-ni [$_{S2}$ Takashi-ga$_i$ [$_{S3}$ PRO$_i$ t$_j$
[$_{S1}$ Masao-NOM **Boston-to**$_j$ Kumiko-DAT [$_{S2}$ Takashi-NOM [$_{S3}$ PRO$_i$ t$_j$
ikoo-to] keikakushi-te-iru-to] it-ta]
go-so-that] plan-STAT] say-PAST

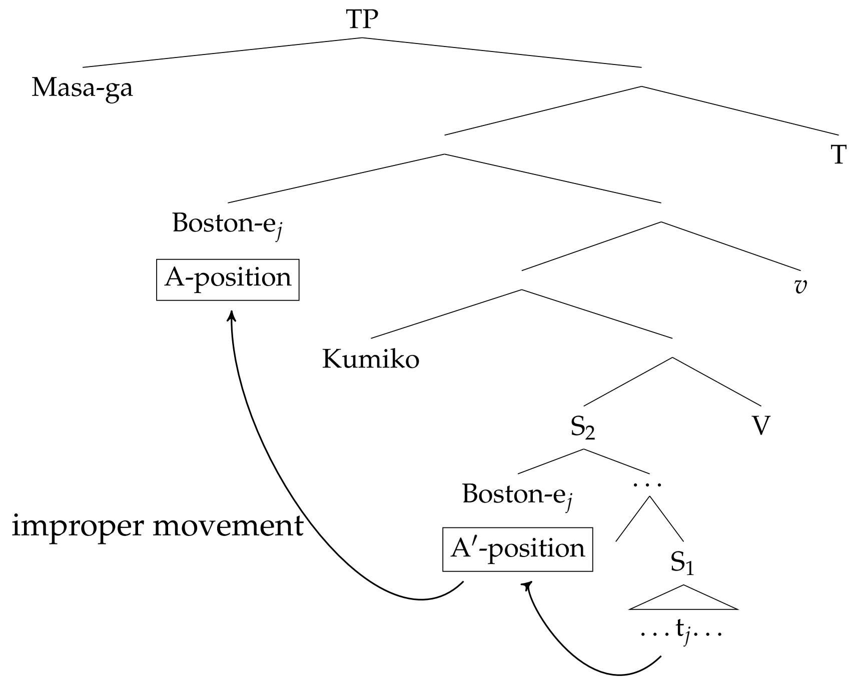
'Masao told Kumiko that Takashi is planning to go to Boston.' (Sakai, 1996, 130-131, (13))

In general, long-distance scrambling across a finite CP is A′-movement, which is grammatical, as shown in (149).

(149) [$_{S1}$ **Boston-e**$_j$ Masao-ga Kumiko-ni [$_{S2}$ Takashi-ga$_i$ [$_{S3}$ PRO$_i$ t$_j$
[$_{S1}$ **Boston-to**$_j$ Masao-NOM Kumiko-DAT [$_{S2}$ Takashi-NOM [$_{S3}$ PRO$_i$ t$_j$
ikoo-to] keikakushi-te-iru-to] it-ta]
go-so-that] plan-STAT] say-PAST
'Masao told Kumiko that Takashi is planning to go to Boston.'

Thus, if (148) is A′-movement, then it should be grammatical, contrary to fact. Thus, we assume here that the locus of the ungrammaticality in (148) is that the matrix post-subject position (i.e., [spec,*v*P]) is an A-movement, which is an improper movement, since it passes the embedded [spec,CP]. I interpret Nemoto's observation in (146b), as being due to the fact that the scrambled element moves to the embedded [spec,CP], which is A′-movement. The relevant structure is as follows.

(150) *



Since S_2 is a finite clause, scrambling out of it makes the movement A'-type movement, and scrambling to the post-subject position is A-movement (Sakai, 1994, 1996). Thus, again it becomes improper movement. As we discussed in (141), which is repeated below as (151), this derivation yields a "wrong" instruction based on FC. Since the copy *Boston-e* in [spec,*v*P] c-commands the copy in [spec,CP], they become repetitions, hence (151) is what will be pronounced.

(151) Masao-ga [$_{vP}$ **Boston-e** Kumiko-ni [$_{CP}$ **Boston-e** Takashi-ga it-ta to it-ta]].

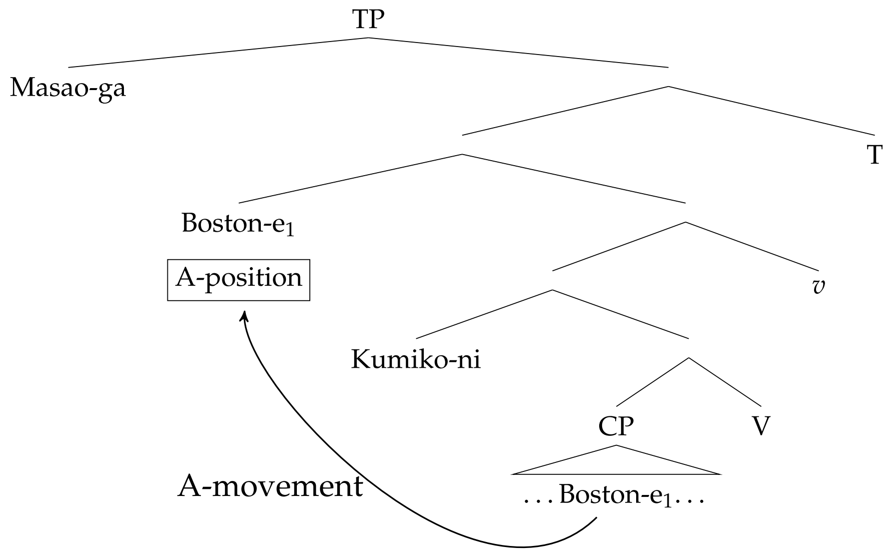
Going back to (142b), which is repeated as (152) below, suppose there is no movement to the embedded [spec,CP], since it is not a phase (as it is a non-finite clause). Based on Nemoto's (1995) discussion, scrambling out of a non-finite clause is not A'-movement.

(152) (=(142b))

Masao-ga **Boston-e**$_j$ Kumiko-ni$_i$ [PRO$_i$ t$_j$ iku-yooni] susume-ta.
Masao-NOM **Boston-to**$_j$ Kumiko-DAT$_i$ [PRO$_i$ t$_j$ go-so-that] advise-PAST

'Masao advised Kumiko to go to Boston.' (Sakai, 1996, 130, (11))

(153)



Thus, the movement is from the original position in the embedded CP to the intermediate position in the matrix clause (i.e., [spec,*v*P]), hence there is no improper movement. FC finds two copies of *Boston-e*, namely a copy in the matrix [spec,*v*P] (*Boston-e*$_1$) and a copy in the embedded clause (*Boston-e*$_2$). *Boston-e*$_1$ c-commands (*Boston-e*$_2$), thus *Boston-e*$_1$ will be pronounced, as follows.

(154) Masao-ga [$_{vP}$ **Boston-e** Kumiko-ni [iku-yooni] susume-ta].

In this section, we explored an account for improper movement, based on Chomsky's (2021b) analysis of the interaction between the INT system, theta theory and FC. MERGE satisfies theta theory and forms syntactic structures, but INT, the interpretive system, cannot see the history of the derivation since the derivation is Markovian. Thus, INT checks the representation of the derivation and it has to determine which element is a copy and which element is a repetition. As a result of the interaction of INT, theta theory, and FC, I argued that the derivation for hyper-raising results in establishing a 'wrong chain' and gives a 'wrong' pronunciation of copies (e.g., ***John*** *seems* ***John*** *that is happy*), since improper movement includes A'-movement followed by A-movement, which disconnects the chain. Furthermore, I argued that proper scrambling in Japanese (which is a grammatical derivation) does not include A$'$-movement, thus, the problem with hyper-raising in English does not occur.

2.8 The CED effect

In the following subsections, I deal with two syntactic islands, the subject island and the adjunct island. In §2.4.3, we noted that Japanese does not show the freezing effect. In §2.8.1, I argue that MERGE can generate the freezing effect configuration without any violation of conditions on MERGE. Since the subject island and the adjunct island are treated by a uniform principle such as the Condition on Extraction Domain (Huang, 1982) and the multiple spell-out analysis (Nunes & Uriagereka, 2000) in minimalist literature, I also address the adjunct island. In §2.8.2, I propose that adjuncts are introduced by *Form Set*, which is a more primitive version of MERGE (combining any elements in the workspace and putting them into a set), so extraction out of adjuncts is, in principle, possible, but it needs to be licensed by a semantic condition (Truswell, 2011; Ernst, 2022), an extrasyntactic principle. Since both islands have different aspects, I will conclude that we need to explain them separately by syntactic or extrasyntactic factors.

It has been argued in the literature that extraction from non-complement may not be possible, which is called the *Condition on Extraction Domain* (CED, Huang 1982).

(155) Condition on Extraction Domain (Huang 1982: 505, (118), cf. Cattell 1976)
A phrase A may be extracted out of a domain B only if B is properly governed.

For example, subextraction from the subject in (156a) and extraction out of the adjunct in (156b) are impossible.

(156) a. * Who$_i$ did [$_{subj}$ stories about t$_i$] terrify John? (Chomsky, 1977, 106)

b. * Which celebrity$_i$ did Mary eat an ice cream [$_{adj}$ before she saw t$_i$]? (Huang, 1982, 503)

In the minimalist literature, it is assumed that all theoretical machinery has to be motivated by conceptual necessity, and the notion of government is not assumed anymore. However, the generalization driven by CED seems to be assumed in the minimalist literature (Uriagereka, 1999; Nunes & Uriagereka, 2000).

However, it has been claimed that exceptions to the CED effect exist. One of the exceptions to the subject island is from Chomsky (2008).[43]

(157) a. * Of which car$_i$ did [$_{TP}$ [the driver t$_i$]$_j$ [$_{vP}$ t$_j$ cause a scandal]? (Chomsky, 2008, 147, (6b))

b. Of which car$_i$ is [the driver t$_i$] likely [t$_j$ to [t$_j$ cause a scandal]]]]? (Chomsky, 2008, 153,(18b))

The set of data (157a) and (157b) already suggests that a uniform analysis of subject island such as CED may not be supported for empirical reasons.

As for the adjunct island, Truswell (2007, 2011) discusses some exceptions to the adjunct island.

(158) a. What$_i$ did John arrive [whistling t$_i$]? (Truswell, 2007, 1357, (4b))

b. Who$_i$ did John get upset [after talking to t$_i$]? (Truswell, 2011, 129, (1b))

Notice that finiteness of the adjuncts seems to matter, as shown in (159).[44]

[43] This exception does not hold for some languages. See Broekhuis 2006, for example.

[44] See Truswell (2011: Chapter 4) for details. He assumes a structure in which the matrix VP includes Op, which determines the event denotation. The finite adjunct has another Op in VP that determines the domain of the event, which yields multiple-event readings. In this dissertation, the finiteness is not discussed in detail, but see also Michel & Goodall 2012; Bondevik et al. 2021; McInnerney & Sugimoto 2022.

(159) * What$_i$ did John die [after he kicked t$_i$]?

(Borgonovo & Neeleman, 2000, 203, (12b))

However, if the CED effect holds regardless of the internal structure of adjuncts, extraction out of the adjunct should be banned, since CED refers to the syntactic configuration (i.e., non-governed positions or non-complement positions). Thus, the exceptions to the adjunct island suggest that a uniform treatment of adjuncts seems to be untenable.

This point is not a new argument. For example, Stepanov (2007) already points out that CED does not hold for some languages which allow subject island violations, such as Russian, Hungarian and Japanese. Borgonovo & Neeleman (2000); Truswell (2007, 2011) observe that extraction out of adjuncts is transparent for some cases in English and it has been attested that other languages also do not show robust adjunct island effects.

In what follows, what I would like to do is to show that recent developments of the theory of structure-building operations can capture the exception to CED phenomena, suggesting that the CED effect might not be a robust generalization in the narrow syntax.

As we discussed, MERGE only generates EM or IM (the simplest cases) and the extension of Merge (e.g., parallel Merge, see (76) in §2.7.1) is illegitimate. In the case of EM (160), a new accessible element is {a,b}.

(160) External Merge (EM)

a. WS_1 = { a, b }

b. MERGE(a, b, WS_1) = [{a,b}] = WS_2

In the case of IM (161), there are apparently two new accessible elements; *b*, and {b,{a,b}}.

(161) Internal Merge (IM)

a. WS_1 = { {a,b} }

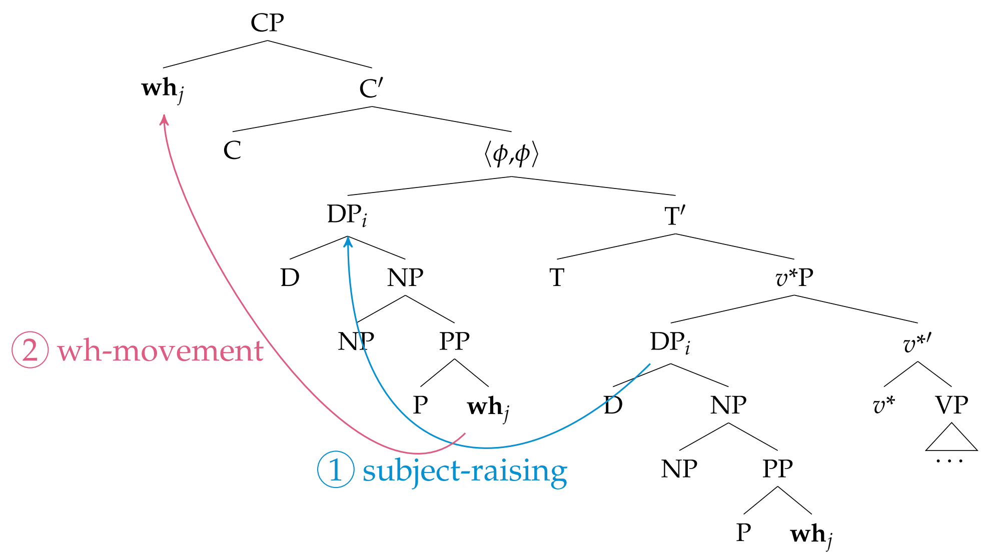
b. MERGE(b,{a,b}, WS_1) = { {b,{a,b}} } = WS_2

Within the core system of language (i.e., narrow syntax), the computation limits the domain that can be searched. One such limit is the phase impenetrability condition (PIC, Chomsky 2000). The other one is minimal search. In the case of IM, the lower copy is not used for further operations (cf. strict cyclicity in Chomsky 1973). Minimal search can only find the higher copy, and the lower copy becomes unavailable for MERGE. As a result, after IM, there is only one new accessible element {b,{a,b}}. Notice that the higher copy of *b* is still available for MERGE (see (70) in §2.7.1 for RR and minimal search.).

2.8.1 Subject Islands

An immediate consequence under MERGE is that the subject island effect, an instance of the freezing effect, cannot be derived since the subject island configuration is formed by internal Merge, which is legitimately formed by MERGE. This means that as long as there is an independent extrasyntactic principle to rule out the subject island, MERGE can create the configuration, hence it would be grammatical if it were only due to MERGE. Let us consider the following schematic tree representation that shows the subject island configuration.

(162)



The subject DP moves from [spec,*v**P] to [spec,TP], which is an instance of internal Merge. The subject in [spec,*v**P] is inaccessible due to minimal search (more precisely, it is a lower copy (*pace* Goto & Ishii, 2019)). Subextraction from the subject in [spec,TP] is fine since it is accessible for MERGE unless there are extra restrictions on subextraction from the subject.

In Chomsky (2008), the subject island effect is captured by minimal search, which is explicitly formulated by Gallego & Uriagereka (2007) as the edge condition.

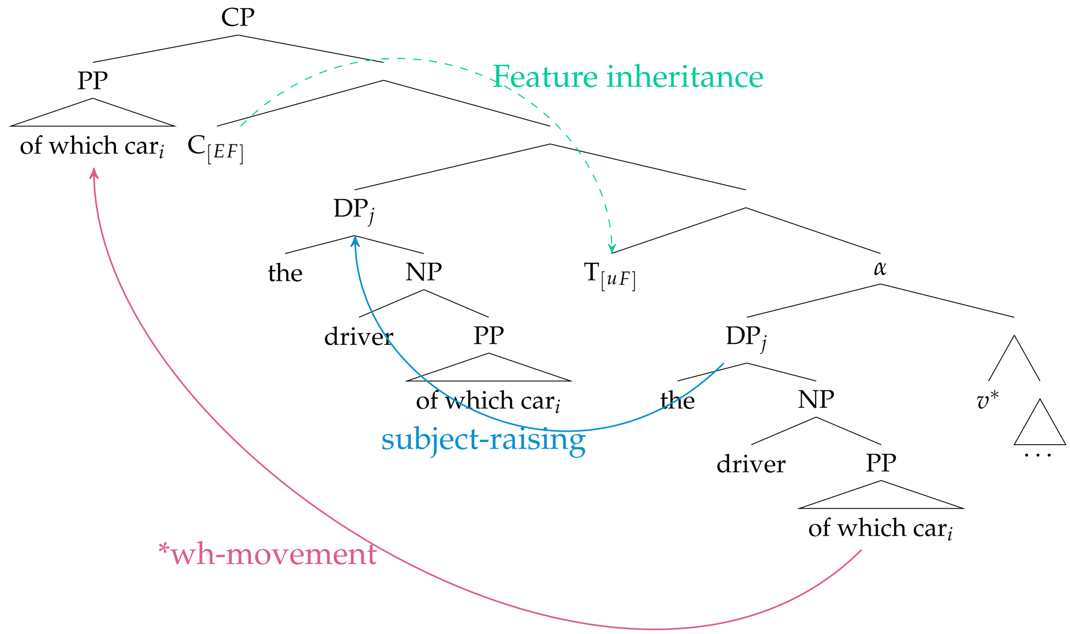
(163) Edge condition

Syntactic Objects in phase edges are internally frozen. (Gallego & Uriagereka, 2007, 55)

In Chomsky (2008), he assumes that the operations (e.g., feature inheritance, Agree, movement, etc.) apply at the phase level simultaneously.[45]

[45] I will skip the mechanical explanation for this derivation. But see Chomsky (2008) and §3.2.3 for the details. See also Broekhuis 2006; Boeckx 2008; Fortuny 2008 for the analysis of the subject island in terms of Chomsky (2008). See also Rizzi (2006); Lohndal (2011).

(164) A subject island effect configuration in Chomsky (2008)



Subextraction from the subject takes place in the subject at [spec,$v*$P], not [spec,TP], and moves the *wh*-phrase from the lower copy to [spec,CP]; the subject itself moves to [spec,TP] simultaneously. However, MERGE only allows a strictly cyclic derivation. Thus, Chomsky's (2008) analysis cannot be sustained under the strict cyclicity of MERGE. The subject-raising is followed by introducing C, which is a counter-cyclic movement.

Assuming that subextraction from the subject takes place in [spec,TP], one might think that the [spec,TP] position is a strong position/case position/EPP-position, which usually forces an element in it to be frozen (cf. Boeckx, 2003; Lohndal, 2011; Haegeman et al., 2015). Under the strong minimalist thesis, it's not clear how to derive the notion of *freezing* in the narrow syntax. We need to stipulate, for example, that the case position is frozen. However, the activity condition (Chomsky, 2001) is abandoned (Nevins, 2005; Chomsky, 2007). The EPP is derived from labeling theory (Chomsky, 2013, 2015) and subextraction from the subject does not cause any problem.[46] Furthermore, as Stepanov (2007) points out, linguistic variation cannot not be captured if we have uniform grammatical principles to rule out the subextraction from the subject.

Although how to deal with the subject island within narrow syntax remains unclear, I take this as an advantage since there are several exceptions to the subject island.

(165) a. Of which car$_i$ did [the driver t$_i$]$_j$ collapse t$_j$? (Zyman, 2021, 517,(15))

b. Of which books$_i$ did [the authors t$_i$] receive the prize? (Chaves & Putnam, 2020, 141,(29a))

c. Which doctors$_i$ have [patients of t$_i$] filed malpractice suits in the last year? (Chaves, 2013, 301,(32c))

Again, there is plenty of evidence from other languages that do not show the subject island (e.g. Stepanov, 2007).[47]

(166) Japanese (cf. Kikuchi, 1987; Takahashi, 1994; Omaki et al., 2020)

[Op$_i$ [Mary-ga t$_i$ yonda no]-ga akirakana yorimo John-wa
[Op$_i$ [Mary-NOM t$_i$ read that]-NOM is-obvious than John-TOP
takusan-no hon-o yonda.
many-GEN book-ACC read

'(*)John read more books than [that Mary read t] is obvious.'

(167) Turkish (cf. Kural, 1993)

[Op$_i$ [Ahmet-in t$_i$ git-me-si]-nin ben-i üz-dü-ğ-ü]
[Op$_i$ [Ahmet-GEN t$_i$ go-INF-AGR]-GEN I-ACC sadden-PAST-COMP-AGR
ev.
house

Lit.'The house [which [that Ahmet went to _] saddened me].'

[46]But see Bošković (2016).
[47]The following set of data comes from Stepanov (2007).

(168) Russian (cf. Polinsky et al., 2013)

S kem by ty xotel čtoby govorit' bylo by odno
with whom SUBJ you wanted that-SUBJ to-speak were SUBJ one
udovol'stvie?
pleasure

Lit.'With whom would you want that [to speak _] were sheer pleasure?'

This approach, at least, gives us room for other factors outside of the narrow syntax (such as sentence processing difficulty) to determine the (un)acceptability of these derivations. I am not trying to argue that the subject island effect is not syntactic at all, but the evidence suggests that MERGE is able to generate this syntactic configuration, at least.[48]

2.8.2 Adjunct Islands

In this subsection, I argue that the adjunct island, another subcase of the CED effect, is generable in the MERGE framework, using *Form Set*, which is a general operation that combines *n* elements (n>2) and forms a set. In Chomsky (2021a,b), the operation *Form Sequence* was proposed to capture phenomena involving unbounded unstructured sequences. Since *Form Sequence* is a departure from SMT, I argue that this operation is not necessary to explain the adjunct island. The empirical evidence suggests that adjunct islands are transparent when the adjuncts are regarded as an event of the matrix clause (Truswell, 2007, 2011). Building on this, I propose that *Form Set* combines *v*P and the adjuncts into a set. In this case, extraction from adjuncts is, in principle, possible, but extrasyntactic principles rule out some cases of extraction out of adjuncts.

[48]Of course, this approach leaves open the question how to deal with the typical subject island effect.

2.8.2.1 Form Sequence

Chomsky (2021a,b) assumes the structure-building operation *Form Sequence* (FSQ). According to him, this is necessary to capture *unbounded unstructured sequences*.

(169) a. the man was old, tired, tall, . . . , but friendly. (Chomsky & Miller, 1963, 298)

b. John, Mary, her mother, . . . and the detective are waiting for the decision.

Intuitively speaking, we want to have an operation that assembles *n* elements ($n \geq 1$) and forms a sequence without structures. If we combine these elements step by step, the structure becomes very complicated (Lasnik, 2011). The rule that we want is something like (170).

(170) $XP \rightarrow XP^n$ and XP ($n \geq 1$)

However, Chomsky & Miller (1963) do not adopt this rule since there are "many difficulties involved in formulating this notion so that descriptive adequacy may be maintained. . . " (Chomsky & Miller, 1963, 298).

Instead of this, Chomsky (2021a,b) proposes the operation *Form Sequence*. Suppose we select *n* elements ($n \geq 1$) in the workspace and we form the set $\{x_1, \ldots, x_n\}$. Then turn this set into a sequence $<x_1, \ldots, x_n>$.

(171) Form Sequence

a. select *n* elements in workspace

b. form sequence: $\langle(\&), x_1, \ldots x_n\rangle$

In cases of coordination, "&" is overtly realized. Chomsky (2021a,b) assume that "&" is also an element to be included in the sequence. When overt "&" is realized and *wh*-extraction takes place from only one conjunct, the sentence would usually become ungrammatical as a violation of the Coordinate Structure Constraint (cf. Ross, 1967).[49]

(172) a. which farm did John live on with his family?

b. * which farm did John live on and with his family?

(Chomsky, 2021b, (34)-(35))

Chomsky (2021a,b) assumes that the ungrammaticality of (172b) is due to a *matching condition*, according to which some sort of parallelism among the conjuncts needs to be assumed (cf. Bošković, 2020).[50]

Let us adopt this Form Sequence analysis for adjunct islands. The idea of adopting Form Sequence for adjuncts seems to be plausible. Adjuncts have been treated as being on a "separate plane" from "the primary plane" (Chomsky, 2004, 117-118). In Chomsky (2004), pair-Merge was proposed to introduce adjuncts, which renders adjunct phrases opaque for operations.

The derivation for the relevant representation is in (173). The first step is to set-Merge of *v**P and the adjunct. Then set-Merge "&" to the resultant structure (i.e., {(&), {*v**P, adjunct}}), as shown in (173a).

(173) Extraction out of adjuncts

a. Form Set (*v**P and adjunct), set-Merge & and IM of *wh*

WS_i = { {wh_i,{C, {subject, {T, {(&),{ {$_{v*P}$ ... }, {$_{adjunct}$ ... wh_i ... }}}}}}}}

b. Form Sequence: (form an ordered set)

WS_{i+1} = { {wh_i,{C, {subject, {T, ⟨(&), {$_{v*P}$ ... }, {$_{adjunct}$ ... wh_i ... }⟩ }}}}}}

[49] I wonder how Chomsky's (2021a; 2021b) analysis deals with the following example.

i. What did he go to the store, buy, load in his car, drive home, and unload? (Lakoff, 1986)

Some conjuncts do not include gaps, and the matching condition should rule out this example.

[50] Though it is not entirely clear what the exact condition this required is. See also Kasai (2004); Bruening (2010); Citko & Gračanin-Yuksek (2021) for a parallel condition on across-the-board movement. These are all referring to the syntactic positions.

Then the structure is extended up to the C phase by MERGE. The *wh*-phrase in the adjunct moves to the [spec,CP]. The next step is to apply Form Sequence, turning the set into an ordered set, which yields the sequence <&, *v**P, adjunct>. The matching condition should be applied at this point (or at the phase level). But what would such a condition be?

Here I adopt Truswell's (2007; 2011) *single event condition* in (174) as a matching condition on the adjunct island configuration.

(174) An instance of *wh*-movement is legitimate only if the minimal constituent containing the head and the foot of the chain can be construed as describing a single event. (Truswell, 2011, 112, (1))

As for the *single event*, Ernst (2022) summarizes the different kinds of single event (or macro-event) in Truswell (2011) and extends his analysis to island repair.

(175) Macro-Event Typology:[51]

Macro-events are groupings of core or non-core events mapped to at most one independent time, of the following types:

a. Core events: events having the structure of a lexically-encoded event, with a maximum of an atelic event (i.e. an activity, a preparatory process) and a culmination.

b. Extended Events (Truswell, 2011, 96)

   i. e_1 occurred and is agentive

   ii. The agent of e_1 intends e_n to occur;

   iii. For every e_k, $1 \leqslant k \leqslant n$, the agent of e_1 believes either that e_k CAUSE e_{k+1} or that e_k ENABLE′ e_{k+1}.[52]

[51] One might wonder whether this condition is universal or not. I leave this issue for my future research.

[52] See Truswell (2011 Ch3) for discussion of the relationship between ENABLE and ENABLE′. For him, causation and enablement are 'contingent relations.'

c. Event Groupings

An event grouping E is a set of two core or extended events $\{e_1, \ldots e_n\}$ such that:

i. every two events $e_1, e_2 \in E$ overlap spatiotemporally;

ii. no more than one (maximal) event $\in E$ is agentive.

(Ernst, 2022, 8,(26))

(174) and (175) explain the contrast between (176a) and (176b), which apparently cannot be accounted for by syntactic analyses.

(176) a. What$_i$ did John arrive [whistling t$_i$]?

b. * What$_i$ did John work [whistling t$_i$]?

(Truswell, 2007, 1369)

For (176a), (175c) applies; (i) the event of arriving could overlap with *John* whistling and (ii) the verb *arrive* does not have an agent, but the verb *whistle* does. Thus, there is only one agentive event in this extended event. However, in (176b), none of the conditions in (175) apply. The core event *work* does not lexically encode in which manner the event takes place. Neither the event *work* or the event *whistle* could have a causal relation. Furthermore, each event is agentive. Thus, (176b) cannot be interpreted as a single group event, which results in the unacceptability of the sentence.

Once the matching condition/the single event condition is met, extraction of a *wh*-phrase out of the adjunct phrase is licensed.

The advantage of adopting this Form Sequence analysis is that it allows extraction of *wh*-phrases from adjuncts in principle. This analysis is, thus, compatible with the exceptions, as Truswell (2007, 2011) discusses. Furthermore, there are pieces of evidence that adjunct islands are transparent in some languages (at least it has been attested in the following languages: Italian (di Ricerca, 2020), Japanese

(Ishii, 1997; Yoshida, 2006), Norwegian (Bondevik et al., 2020), Russian (Tiskin, 2017), and Swedish (Müller, 2017)). Thus, the CED effect is not observed in these languages.

Notice that the single event condition does not apply to subject islands.

(177) a. * Who$_i$ did [close friend of t$_i$] laugh?

(Chaves & Putnam, 2020, 77, (77b))

b. Which$_i$ politician$_i$ did [opponents of t$_i$] organize a protest?

(Chaves & Dery, 2014)

The subject in both sentences is the agent of the event (a part of the event). If the condition also applies to the subject, both sentences should be fine. Thus, the single event condition does not work for all dependencies.

2.8.2.2 Form Set

One might wonder whether *Form Sequence* is another operation distinct from MERGE, which needs to be justified, considering the evolvability criterion that is imposed on UG. In this context, Oseki (2015) suggests that the simplest Merge (i.e, Merge(α, β) = {α, β}) is enough to capture the absence/presence of an adjunct island effect. Based on Epstein et al. (2012), Oseki (2015) proposes that the adjunct is introduced by set-Merge rather than pair-Merge, using Epstein et al.'s (2012) two-peak structure. Suppose that YP is the adjunct phrase that is set-Merged with XP, which is *v*P, which yields an unlabeled {XP,YP} structure. Oseki (2015) assumes Hornstein's (2009) Label Accessibility Condition (178).

(178) Hornstein's (2009) Label accessibility condition:
Only the label of a syntactic object is accessible to Merge

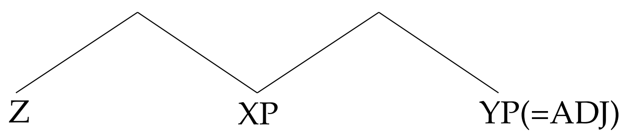
The basic idea is that labeling of syntactic objects is for further operations; if a computation finds an unlabeled structure, it won't continue the computation (see also Chomsky 2000; Epstein et al. 2012). Oseki (2015), then, suggests that Merge cannot combine Z, a head, with {XP,YP}, but it can combine with XP itself. Here the two-peak structure is created. In set notation, the intersection of the members is difficult to illustrate; however, two sets will be generated.

(179) a. {XP,YP}

b. {Z, XP}

The informal tree representation of the two-peak structure is as follows.

(180)



Based on Epstein et al. (2012), Transfer has to apply to the set {XP,YP}, so there is only one root node. Epstein et al.'s (2012) assumption is a different version of label accessibility condition.

(181) Epstein et al.'s (2012) Label accessibility condition:

Only the label of an entire syntactic object, the root, is accessible to narrow syntax. (Epstein et al., 2012, 254)

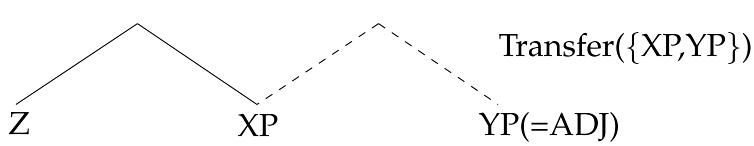
After transferring {XP,YP}, the resulting structure is {Z,XP}. Since YP, the adjunct phrase, is transferred, extraction is not allowed.

(182) a. {XP,YP} $\rightarrow$ Transfer

b. {Z, XP}

The informal tree representation of the two-peak structure is as follows.

(183)



However, Oseki (2015) argues that it is possible to extract from YP, based on Chomsky's (2013) labeling theory. Assuming that XP and YP share features, the labeling algorithm finds two relevant features. Here, Oseki (2015) assumes that $v*$ and the adjunct enter into an Agree relation for accusative case (den Dikken, 2012), or for Aspect-features (Miyamoto, 2012). The generalization is the following.

(184) Oseki's descriptive generalization

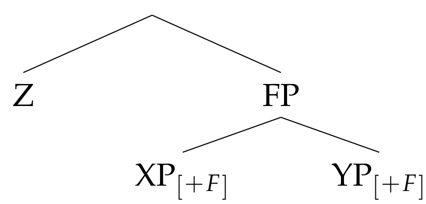
Adjuncts entering Agree are visible to extraction and binding.

(Oseki, 2015, 310,(24))

If XP and YP share a feature F, then Z, a head, is able to merge with {XP,YP} since it has a label. The structure becomes a one-peak structure.

(185) a. $\{Z,\{XP_{[+F]},YP_{[+F]}\}\}$

b.



Since the structure is a one-peak structure, extraction from the adjunct is possible.

There is a problem with this analysis. As the generalization implies, the place where the adjunct is merged is crucial. Brown (2017) argues that if the adjunct is within v^*'s c-command domain, extraction is possible (see also Borgonovo & Neeleman 2000; Narita 2014; Bode 2020 for relevant discussion). However, Boeckx (2012) points that this approach does not apply to the following example.

(186) a. John didn't talk [after any of our meetings]

b. * What meetings$_i$ didn't John talk [after any of t$_i$]?

(Boeckx, 2012, 146,fn14)

The example shows that the low adjunct includes a NPI element, which indicates that it is in the c-command domain of $v*$. Nevertheless, *wh*-extraction from it is impossible.

One problem with Oseki (2015) would be that the MERGE framework, a two-peak structure is a counter cyclic movement, which is not even generable by MERGE.[53] The schematic derivation of counter-cyclic movement is illustrated below. In (187), *c* in {a,{b, {c, d}}} is internally merged to {b, {c, d}} counter-cyclically.

(187) WS = [{a, {b, {c, d}}}]

a. Merge(c, {b, {c, d}}, WS)

b. The output of MERGE → WS′ = [{c, {b, {c, d}}}, {a, {b, {c, d}}}]

This yields (187a): {c, {b, {c, d}}} and {a, {b, {c, d}}}. Before MERGE, the accessible terms are seven: *a*, *b*, *c*, *d*, {c,d}, {b,{c,d}}, and {a, {b, {c, d}}}. After MERGE, the number of accessible terms increases by more than one. As for the set {c, {b, {c, d}}}, *a*, *b*, *c*, *d*, {c,d}, {b,{c,d}}, and {c, {b, {c, d}}} are accessible. Thus, the accessible terms are six. There is also a set {a, {b, {c, d}}} that is in WS′. Within this set, *a*, *b*, *c*, *d*, {c,d}, {b,{c,d}}, and {a, {b, {c, d}}} are accessible. That is, there are seven accessible terms in this set. As a result, after MERGE, the number of accessible terms is increased by six, which is a violation of MY (72). Hence the counter-cyclic operation is not an option.

However, in the Form Sequence (FSQ) analysis, extraction from adjuncts is in principle possible with a qualification: it needs to satisfy the single event condition. If we assume this, there is another way of deriving Oseki's analysis.

Assume that MERGE has to satisfy theta theory for EM, otherwise it is IM, which yields discourse properties (interrogative, focus, etc.). In the cases of merging adjuncts/adverbials, let us assume it is an operation Form Set (FST) where *n* elements can form a set, but not a sequence.

(188) $\text{FST}(X_1, \ldots, X_n, \text{WS}) = \text{WS}' = [\ \{ X_1, \ldots, X_n \}\ ]$

[53] See §2.7.1 and see Epstein et al. 2018a, 2021; Kitahara & Seely 2021 for details.

Chomsky (2021b) suggests that this is a general possibility for the structure-building operation (see also Goto & Ishii 2021 for relevant discussion.). However, natural languages are restricted to binary MERGE because RR has to be met. Nevertheless, FST seems to be the precursor of MERGE, hence it is not a departure from SMT; on the other hand, FSQ is a departure from SMT. In the following, I argue that FST is sufficient to capture derivations, in which adjuncts are introduced, and FSQ is unnecessary for adjuncts.

Let assume the derivation in question is as follows.

(189) a. FST of v^*P and adjunct phrase (ADJ): $\{_\alpha\ v^*\text{P, ADJ}\}$

b. Merge of T to α: $\{\text{T},\{_\alpha\ v^*\text{P, ADJ}\}\}$

c. subject-raising and introducing C: $\{\text{C},\{\text{subject},\{\text{T},\{_\alpha\ v^*\text{P, ADJ}\}\}\}\}$

d. extraction of a *wh*-phrase in the adjunct phrase:

$\{\text{wh}, \{\text{C},\{\text{subject},\{\text{T},\{_\alpha\ v^*\text{P, ADJ}\}\}\}\}\}$

In this example, FST applies to v^*P and ADJ, which does not satisfy the theta theory nor does it create discourse properties. The assumption is that MERGE creates argument structure and informational structure, whereas FST is only available when neither of these are satisfied. In this approach, extraction from the adjunct is possible, but the single event condition (Truswell, 2011) has to be satisfied. This way, we still captures Oseki's insight by FST, without adding the operation FSQ.

In this section, we started from the CED effect, which does not seem to be supported by the empirical evidence. I argued that under the MERGE framework (Chomsky, 2021a,b), the subject island configuration is possible to generable since the subject moves to the [spec,TP] position from [spec,v^*P] and the higher copy, the subject in the [spec,TP], is accessible. Nothing prevents extraction out of the subject. This will give us an account for the languages where subextraction from the subject is possible. The adjunct island, another effect of CED, was also examined in this section. Based on Oseki's (2015) insight and Chomsky's (2021b) analysis of FSQ, I argued that FST, which is a primitive operation of MERGE, can be used to introduce the adjunct phrase. I adopt Truswell's (2011) single event condition as extrasyntactic principle to license the extraction Thus, the proposed model in this section can capture both acceptable and unacceptable sentences in terms of adjunct islands.

2.9 Conclusion

This chapter proposes the restrictions on what derivations are possible. Specifically, MERGE framework does not allow all conceivable ways of applying Merge. The RR principle restricts the accessible elements in the workspace. The interaction between MY and phase theory provides an account of remnant movement. Another constraint is a language-specific condition, namely theta theory. Theta theory has to be satisfied, which explains why improper movement is impossible. As for the CED effect, namely the subject and adjunct islands, MERGE can generate the island configurations, and some extrasyntactic principles can rule out some of the derivations, but not others. This explains the variations in the acceptability of these islands discussed in the literature. More broadly, I argued in this chapter that the MERGE framework is motivated as a linguistical computational system that determines the possible derivations.

In this chapter, we started by reviewing labeling analyses of Japanese (Saito, 2016, 2018; Miyagawa et al., 2019) and we observed a couple of potential technical problems. For example, there are issues with the strength of heads and the label of the DP structure, CP scrambling and root clause labeling in Japanese. We also observed that the labeling analysis does not seems to explain some types of movement, such as improper scrambling, remnant movement, and the freezing effect. The labeling analysis of Japanese predicts that movement of the case-marked elements does not cause unlabeled {XP,YP} problems.[54]

Instead of adopting labeling theory to explain constraints on movement, I introduced Chomsky (2021a,b) to give some principled ways of explaining the different types of movement in Japanese. Under the MERGE system (§2.7.1), remnant movement yields a lethal ambiguity, like parallel Merge. However, in proper cases of remnant movement, PIC blocks one of the copies, so the further operations do not run into any problems. This analysis derives PBC, so there is no need to stipulate this condition as a narrow syntax principle. It also explains the asymmetry between raising-to-subject and raising-to-object in terms of remnant movement in Japanese, which is a natural consequence of MERGE and phase theory (§2.7.4).

As for improper movement (§2.7.5), assuming that the interpretive system cannot see the history of the derivation, Chomsky (2021a,b) argues that two types of IM-configuration exist and this yields obligatory control. Extending this idea to improper movement, I argued that although INT can assign copies freely, when all occurrences of the hyper-raised copies are regarded as copies, the A- and A′-positions are mixed. Thus, I adopt Chomsky's (2019b) idea of the A-/A′-distinction in terms of c-command; when a theta position c-commands a non-theta position, they are repetitions.

As a consequence of MERGE, I also built upon previous analyses to suggest that the subject and adjunct islands are not uniformly robust and the same mechanism cannot explain both of them. Rather, a non-uniform analysis seems to be plausible. For subject islands, MERGE can generate the subject island configuration (§2.8.1). For adjunct islands, *Form Set* makes room for extraction out of adjuncts and the matching condition (i.e., the single event condition) licenses the structure after *wh*-phrase is extracted (§2.8.2). That is, the matching condition applies to the representation of the structure in adjunct island cases, but not in subject island cases.

[54] The remaining question is how to capture Saito's (2016; 2018) insight about labeling in Japanese without causing any technical problems. I wuill come back to this issue in Chapter 3 (see §3.8).

This approach also opens the possibility of considering non-syntactic approaches. For subject islands, sentence processing difficulty (Culicover & Winkler, 2022), frequency effects (Chaves, 2013; Chaves & Putnam, 2020), and other kinds of factors might affect acceptability (Haegeman et al., 2015). As for adjunct islands, the single event condition (Truswell, 2011; Ernst, 2022) does not exclusively refer to the syntactic configuration. However, I do not deny that grammatical principles may make the sentences ungrammatical. Motivations for such principles are, however, very restricted under the strong minimalist thesis. At least, the CED effect does not seems to be a principle in the narrow syntax anymore. We need more fine-grained approaches that explain the violations of this effect, and more empirical data might illuminate different ways of dealing with this generalization.

CHAPTER III

Underspecification of Rule Ordering in Narrow Syntax

3.1 Introduction

As we discussed in the previous chapter, the structure-building operation MERGE restricts the possible derivations. In this chapter, we discuss the linguistic variation emerging from the narrow syntax, appealing to the third factor underspecification of rule ordering, which suggests that even though MERGE restricts the possible derivations, MERGE itself does not control how other operations such as Agree, feature inheritance, and Transfer work in narrow syntax. That is, each operation has a restriction, but, the order of rule application itself is not governed by MERGE, or by any other operations within narrow syntax.[1] The schematic patterns of rule ordering below illustrate this point (see the next section for the definition and details of each operation).

[1] See Chomsky (1955/1975, 1957), for example, for an early discussion of rule ordering where the interaction between transformation is discussed.See also Kiparsky (1968, 1971, 1976); Georgi (2013); Müller (2013) for relevant discussions.

(1) Pattern 1

i. Merge
ii. feature inheritance
iii. Agree
iv. Move
v. Transfer

(2) Pattern 2

i. Merge
ii. Move
iii. Agree
iv. feature inheritance
v. Transfer

In the first pattern (1), feature inheritance applies right after Merge and before Agree, whereas in pattern (2), feature inheritance applies after Agree, and these two derivations should be possible unless there are additional restrictions on rule ordering. And these are not the only possible derivations, in principle.

This approach is complaint with third factor principles in that the derivations in narrow syntax are optimized not by syntactic principles, but by non-language specific principles such as "computational efficiency" (cf. Chomsky, 2000, 2005). As a result, third factor-compliant operations have restrictions. For example, Agree, feature inheritance, Move, and Transfer cannot apply before the structure is built. Agree has to have a probe and a goal with unvalued features, the goal has to be the closest candidate for the probe, etc. As long as the probe is valued, different timing of movement would not make the derivation fatally crash. For example, Move can take place before Agree. Thus, the Move-Agree order and the Agree-Move order are possible, wand these orderings are not rule out by the structure-building operation or by Agree or Move. The order of rule application is underspecified in a restricted way. Obata et al. (2015) argue that this is one way of capturing linguistic variation. In this chapter, we extensively discuss and explore the possible rule orderings in narrow syntax, including the possible combinations of functional categories. In previous studies, the rule ordering options Agree-Move and Move-Agree were attested in constructions in particular languages. In section §3.8, I propose additional patterns to derive parametric variation from the underspecification of rule ordering, whether a head is introduced as a free-standing head or combined with another head or more before being introduced into narrow syntax.

In what follows, §3.2 frames the issues regarding linguistic variation under a restrictive grammatical theory and introduces the operations in narrow syntax under phase theory in detail. This section introduces the locus of linguistic variation in the historical context, given that the narrow syntax is uniform under the minimalist program (Chomsky, 2001). §3.3 has two concrete examples regarding the Agree-Move and Move-Agree orderings. The first set of examples involves raising constructions in Icelandic and the second set of examples is participle agreement in Swedish. §3.4 introduces the derivational steps based on Chomsky's (2013) labeling theory. §3.5 introduces Obata et al.'s (2015) analysis of T-subject agreement and T-object agreement in English and demonstrates how different derivational rule orderings derive different optimal derivations. This section also introduces the strictly cyclic derivation steps proposed in Chomsky (2015). §3.6 discusses and explores some previous ideas regarding showing derivations step by step in recent frameworks (Epstein et al., 2018b; Chomsky, 2020). I also explore two constructions in English: the *tough*-construction and the hyper-raising construction. Then §3.7 will show how the underspecification approach works for complementizer agreement in Haitian Creole, Cabo Verdean Creole and Brazilian Portuguese. I will reanalyze the complementizer agreement in Haitian Creole and Cabo Verdean Creole dicussed in Obata et al. (2015) based on a strictly cyclic derivational model with labeling theory. §3.8 will explore a new domain of underspecification of rule ordering, namely, the underspecification of the application of Merge to the heads. Comparing English, Germanand Japanese, I argue that whether or not Merge applies to functional heads derives a cluster effect without referring to macro-parameters. §3.9 concludes this chapter.

3.2 Theories of Linguistic Variation under the Minimalist Program

3.2.1 Universality and Diversity under the Minimalist Program

The advent of the minimalist program has given us a new challenge by asking how much can we minimize UG, which only includes the structure-building operation Merge (according to the strong minimalist thesis) and asking us how to motivate UG from the point of view of evolution (beyond explanatory adequacy, Chomsky 2004[2]). In this sense, the notion of parameters, which were the locus of parametric variation in the previous literature, is not conceptually sustainable under the minimalist program. It does not seem plausible to assume, for example, that fifty or more parameters have evolved in our genetic endowment. The question is, then, what is the source of linguistic variation?

The notion of 'parameters' was one of the legacies of the principles and parameters approach (Chomsky, 1981), which allowed researchers to maintain principles, but also capture the variation among languages. Parameters that were proposed by scholars, mainly in 1980s, were macro-parameters that captured 'cluster effect' in variation that cannot be deduced from the properties of lexical items. For example, the head parameter is not just on the verbal head, but also on the P, T, and C heads. In Japanese, every head is on the right side of its complements in the X′-schema, whereas in English, heads are on left side of their their complements. Another type of macro-parameter is the null subject parameter (or the pro-drop parameter). In null subject languages, the following properties were predicted by this parameter (cf. Chomsky, 1981)

[2] See also Fujita 2007, 2009; Narita 2010; Narita & Fujita 2010; Boeckx & Uriagereka 2007 for relevant discussion.

(3) a. missing subjects

b. free inversion in simple sentences

c. long *wh*-movement of subjects across *wh*-islands

d. empty resumptive pronouns in embedded clauses

e. apparent violations of the that-trace filter

(Newmeyer, 2005, 187-188,(12))

However, parameters have been criticized both conceptually and empirically under the minimalist program; they are conceptually criticized from the minimalist perspective as I noted above, and they are empirically criticized because it turns out that the cluster effect does not seem to be robust (cf. Newmeyer, 2005, 2017; Haspelmath, 2008; Boeckx, 2011; Boeckx & Leivada, 2013; Duguine et al., 2018).

One way of minimizing UG can lead to the conclusion that parameters cannot exist in UG. That is, "macro-parameters, at least many of them, do not exist in UG" (Chomsky, 2017a). Then, how do we capture the language variation without any parameters in the minimalist framework? There are two main views.

(4) The locus of linguistic variation

a. Chomsky-Borer Conjuncture: The locus of the linguistic variation might come from the inflectional features (cf. Baker, 2008)

b. The Berwick-Chomsky Conjecture: Linguistic variation comes from sources after the narrow syntax (i.e., externalization) (Berwick & Chomsky, 2011, 2016)

One candidate for the locus of linguistic variation is the lexicon. Borer (1984) argues that the elements that make linguistic variation are found in the lexicon (e.g., inflectional elements). Another approach is that surface order is a property of externalization (cf. Chomsky, 2013; Berwick & Chomsky, 2011, 2016). In this view, the head parameter effect (involving linear order), for example, is not a property of the narrow syntax. Rather, it might be a property of the sensory-motor system (e.g., Fukui 1995b for a head parameter and Richards 2008 for linearization at the phonological component under the symmetric merge approach.).

In a more explicit and comprehensive way, Fukui (1995a) proposes the following.

(5) a. Parametric variation outside of the lexicon must be limited to ordering restrictions ("linearity").

b. Inside the lexicon, only [+F] elements ("functional elements") are subject to parametric variation ("functional parametrization hypothesis")

(Fukui, 2006, 112, (20a),(20b))

The spirit of Fukui's approach is to restrict the theory of parametric variation under the principles and parameters approach, though his proposal fits the Borer-Chomsky conjuncture and the Berwick-Chomsky conjecture.

The combination of the Borer-Chomsky and Berwick-Chomsky conjectures suggests that the narrow syntax, the core system of computation that yields an infinite array of hierarchical structures, is uniform. Chomsky (2001) calls this the *Uniformity Thesis*. Later, Miyagawa (2010, 2017) and Boeckx (2011) came up with stronger versions of the Uniformity Thesis, as shown below.

(6) a. Uniformity Thesis (Chomsky, 2001, 2,(1))

In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances.

b. Strong Uniformity (Miyagawa, 2010, 12,(15)) (see also Miyagawa 2017)[3]
   All languages share the same set of grammatical features, and every language overtly manifests these features.

c. Strong Uniformity Thesis (Boeckx, 2011, 210,(2))
   Principles of narrow syntax are not subject to parameterization; nor are they affected by lexical parameters.

As indicated above, there are two alternatives; one is from Miyagawa (2010, 2017). This version says the set of grammatical features is uniform, but the realization of those features can differ from language to language (e.g., whether the ϕ-features are realized on C or T head.). Another version of Uniformity from Boeckx (2011) is a stricter version. This version is very strong in that it does not allow for lexical parameters. In this thesis, I indirectly adopt Miyagawa's (2010) version (see §3.8 for the details.)

3.2.2 A Third Factor Principle Approach to Parametric Variation

There is another possibility for deriving a part of linguistic variation in the narrow syntax. In a broader sense, this approach could be called *the third factor principle-driven approach*. Assuming that the faculty of language is implemented in our biological endowment, Chomsky (2005) suggests there are three factors that are involved in the language growth.

(7) a. Genetic endowment

   b. Experience

   c. Principles not specific to the faculty of language (Chomsky, 2005, 6)

[3]Miyagawa (2010: 12) mentions that "This strong interpretation of the Uniformity Principle cannot be right for all features of a language. After all, languages do vary." He suggests that, at least, the realization of ϕ-features can vary among languages.

(7a) is supposed to be uniform for the human species. For Chomsky (2005), it determines the general course of the development of the faculty of language. (7b) refers to the input throughout the development of the language faculty. (7c) refers to principles that are not language-specific, such as principles of data analysis and principles of efficient computation (Chomsky, 2005). In the principles and parameters approach, the first factor was rich enough to capture language acquisition, and the second factor (7b) was necessary to fix the parameter settings. However, as we discussed, since the principles and parameters approach is not tenable in the minimalist program, the goal is the deduction of the first factor (7a) by appeal to the third factor (7c).

Under the third factor principle-driven approach, Biberauer & Roberts (2015) and their subsequent work (Roberts et al., 2014; Biberauer et al., 2014; Roberts, 2019) suggest that this approach can derive parameters. Assuming that the functional elements are underspecified in UG, they suggest the following principles.

(8) The third factor effect/learning process (Biberauer & Roberts, 2015, 7,(6))

   a. Feature Economy (FE): Postulate as few formal features as possible to account for the input

   b. Input Generalisation (IG): If a functional head F sets parameter P_j to value v_i then there is a preference for all functional heads to set P_j to value v_i (cf. Boeckx's (2011) Superset bias)

(8a) minimizes the possible formal features (FFs) given the data, and (8b) maximizes the value sharing of the available FFs. For example, language aquirers, by default, do not have FFs on heads by (8a). Once they detect FFs, they maximize the effect, namely, all relevant heads will have FFs. Later, they might detect some heads without FFs. FE and IG are implemented in Biberauer's (2019) model with the principle *Maximise Minimal Means*, which regulates the acquisition of formal features (Biberauer, 2017, 2019). The schema of this model is represented as follows.[4]

(9) UG + input + Maximise Minimal Means (MMM) → Adult Grammar

(Biberauer, 2019, 213, (3))

Given this approach, the FFs are acquired, rather than specified as part of UG. Based on this principle, Biberauer & Roberts (2015) argue that the parameter hierarchy and different types of parameters are deduced from third factor principles, instead of specified as first factors.

(10) Types of parameters

For a given value v_i of a parametrically variant feature F:

a. Macroparameters: all heads of the relevant type, e.g. all probes, all phase heads, etc, share v_i;

b. Mesoparameters: all heads of a given natural class, e.g., [+V] or a core functional category, share v_i;

c. Microparameters: a small, lexically definable subclass of functional heads (e.g. modal auxiliaries, subject clitics) share v_i;

d. Nanoparameters: one or more individual lexical items is/are specified for v_i

(Biberauer & Roberts, 2015, 9,(8))

The typical example of a macro-parameter (10a) would be the head-final order of Japanese since this language only has the head-final property across all syntactic categories. According to Biberauer & Roberts (2015), an example of (10b) is the null-subject parameter, where some Romance languages have FFs on T and pronominal Ds (see Rizzi, 1982). (10c) is Borer-Chomsky conjecture-type variation, where a functional category in a given language has or does not have a given

[4] See also Westergaard 2014 for a related approach invoking micro-cues in L1 acquisition.

value (case-feature, ϕ-features, etc.). The last one (10d) is realized when some lexical items, or only one, show FFs. Thus, their approach deduces every type of parameter from the underspecification of formal features, elaborated with the two principles (8a) and (8b). In this way, Biberauer & Roberts (2015) derive parametric variation: the underspecification of formal features is the key.

3.2.3 Syntactic Operations and Underspecification of Rule Ordering under Phase Theory

A similar, but not identical approach is proposed by Obata et al. (2015) who suggest that linguistic variation could come from the underspecification of the rule ordering in the narrow syntax. They show that different orders of the application of syntactic operations derive linguistic variation. This means that the orders in which rules are applied is not fixed in the narrow syntax at the beginning.

There are several operations in the narrow syntax: (i) Merge, (ii) Move (which is an instance of Merge), (iii) feature inheritance, (iv) Agree, (v) labeling, and (vi) Transfer (cf. Chomsky 2000, 2001, 2004, 2005, 2007, 2008, 2013, 2015).

(11) Syntactic operations

a. Structure-building operation Merge

b. Agree/Labeling: Minimal search (top-down search)

   i. Top down (probe-goal) Agree: (cf. Chomsky, 2000)

      i. the probe has an unvalued feature

      ii. the probe finds the relevant feature in its domain (c-command)

      iii. the goal has a valued feature and it is not frozen (Chomsky, 2001)

c. Feature Inheritance (Richards, 2007; Chomsky, 2007, 2008, 2020)

   i. unvalued features on the phase head are transmitted to the non-phase head (e.g., from C/$v*$ to T/V).

   ii. the valuation process takes place in the domain of the phasal complement, since it is transferred (cf. Richards, 2007)

d. Transfer (cf. Chomsky, 2004)

   The phase complement is sent off to the interfaces cyclically

Merge forms the structure, to which other operations apply. The operations Agree and labeling, which are based on minimal search, find the relevant features. The search algorithm is not language-specific, but rather is based on a third factor principle: it finds the relevant element in a certain domain.[5] Feature inheritance transmits the features on phase heads to non-phase heads. This is connected with the interpretability of the features and the domain of Transfer, which sends off the information to the interfaces. The domain of Transfer corresponds to the Phase Impenetrability Condition (PIC) domain.

(12) Phase impenetrability condition

Given the structure HP = [α [H β]] where H is a phase head, β is a domain of H, and α is H's edge, in phase α with head H, the domain of H is not accessible to operations outside α; only H and its edge are accessible to such operations (cf. Chomsky, 2000, 108)

This condition bans long-distance movement across phases in one fell swoop and makes derivations cyclic.

The idea is that in the minimal computational unit or, more precisely, within a phase, the rule ordering is underspecified. Once the derivation has reached the phase level, all operations are available, but the order is not fixed. The following

[5] See Ke (2019) for a more formal definition of minimal search.

is a demonstration of a phase-level derivation based on Chomsky (2008). First the assumptions are as follows.

(13) Assumptions regarding phasal derivations in Chomsky (2008)

a. The loci of unvalued features are phase heads (v* and C)

b. Phase heads have edge-features and Agree/ϕ-features

c. Edge-features are on lexical items that permit merge to apply[6]

d. Operations apply at the phase level except for external Merge (cf. Epstein & Shim, 2015)

e. Operations are applied simultaneously once the phase head is introduced

In a phase-theoretical model, we assume that unvalued features (such as ϕ-features or what Chomsky (2008) calls Agree-features) are on the phase heads, which also defines the domain of the phases (cf. Chomsky, 2015). Lexical items have their own features to be merged, which are called edge-features. As a result, the phase head has unvalued features and edge-features. In Chomsky (2008), the syntactic operations are assumed to be applied simultaneously at the phase level, when the phase head is introduced. See the following derivation that shows the C phase unit (the v* domain also works in a similar way; see Chomsky 2007, 2008):

[6]Edge-features were proposed in Chomsky (2007, 2008). The idea was that this feature guarantees that elements merge recursively. Chomsky (2008) states that

> "For an LI[lexical item] to be able to enter into a computation, merging with some SO, it must have some property permitting this operation. A property of an LI is called a *feature*, so an LI has a feature that permits it to be merged. Call this the *edge-feature* (EF) of the LI. If an LI lacks EF, it can only be a full expression in itself; an interjection. When merged with a syntactic object SO, LI forms {LI,SO}; SO is its complement. The fact that Merge iterates without limit is a property at least of LIs - and optimally, only of LIs, as I will assume." (Chomsky, 2008, 139)

The edge-feature is assumed here for expository reasons only. Since I update this derivational model in the following sections, I won't specifically assume edge-features in the following discussion.

(14) A derivation of "who saw John"

a. Merge C: {C, {T, {who, {v*, {see, John}}}}}

b. At phase level: {who$_i$ {C, {who$_j$ {T, {who$_k$, {v*, {see, John}}}}}}}

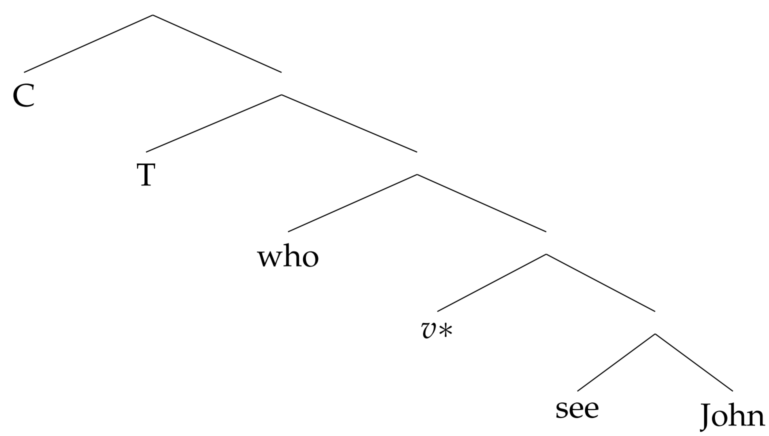
i. Agree(C,who)

ii. Feature inheritance

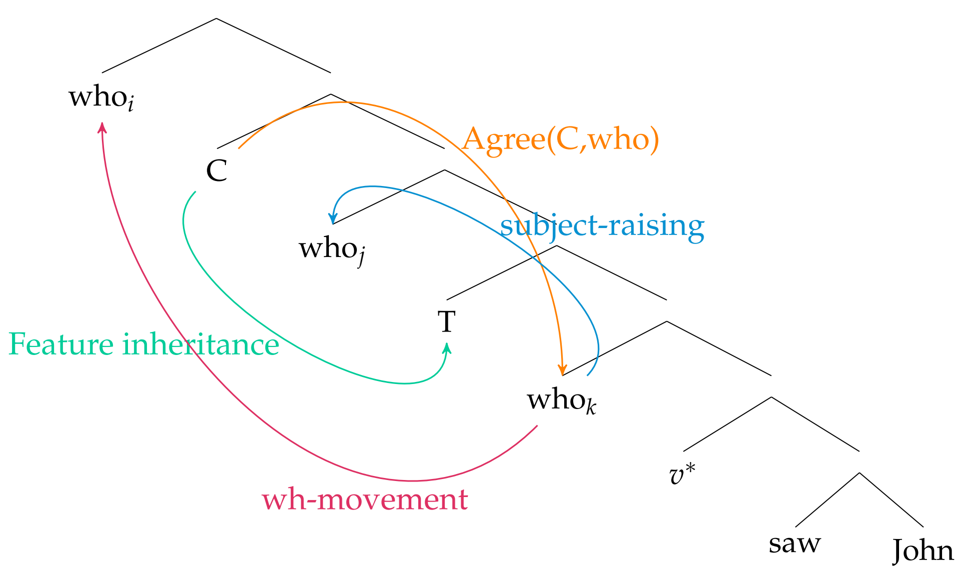
iii. IM of *who* to [spec,CP] (via edge-feature)

iv. IM of *who* to [spec,TP] (via Agree-feature)

(15) a.



b.



Once C is introduced (15a), C initiates the Agree operation and finds the goal (*who*). The Agree-feature, which is inherited by T from C, raises the goal to its specifier position (subject-raising). The edge-feature on C also raises the goal to its specifier position (*wh*-movement). Then the complement of the phase head, namely TP, is transferred (15b).

There has been debate over whether the operation Transfer literally gets rid of the structure from narrow syntax, so it is not accessible for further operations because its structure is gone from the narrow syntax. The idea of Transfer goes back to Spell-out in Chomsky (1995), where the complete derivation is sent off to the phonological component. In Uriagereka (1999), Chomsky (2000), and Chomsky (2001), Spell-out actually "strips away the [...] phonological features, so that the derivation can converge at LF" (Chomsky, 2000, 118). Chomsky (2004) replaces Spell-out with Transfer, where the derivation is sent off to the phonological component and the semantic component at the same time. Some literature adopts the idea that this Transfer is stripping the structure away to send it to the interfaces (Ott, 2011; Narita, 2011; Epstein et al., 2012; Narita, 2014). On the other hand, Obata (2010), Obata (2017), and Chomsky (2013); Chomsky et al. (2019) consider that the transferred structure cannot be removed completely. Consider the following example:

(16) Whose claim that John bought the book did Mary believe?

a. Step 1: [$_{DP}$ whose claim [$_{CP}$ that [$_{TP}$ John bought the book]]]
$\rightarrow$ TP is transferred

b. Step 2: Mary (did) believe [whose claim that [$_{TP}$ ___]]

c. Step 3: [$_{DP}$ whose claim that [$_{TP}$ ___]]$_i$ did Mary believe t$_i$

d. The output: *Whose claim that did Mary believe John bought the book?

(adapted from Obata 2017: 121 with slight modification)

Assume that Transfer gets rid of the structure. In (16a), the TP is transferred, while the DP that includes the transferred TP will be fronted at Step 3 (16c). If Transfer is strong enough to strip the structure TP, then the output will be (16d). Obata (2017) concludes that the transferred structure should remain in the derivation, while the transferred domain becomes inaccessible (e.g., due to the PIC). Moreover, Bošković (2007) shows that the PIC domain is penetrable, which also supports Obata's idea that the transferred domain remains in the narrow syntax. Another issue is how to transfer the root clause, which connects to Richards' (2007) feature inheritance argument based on Chomsky (2008). He argues that feature inheritance is necessary for the valuation process by phase, assuming that uninterpretable/unvalued features must be deleted before being sent off to the interfaces. These features become indistinguishable from the interpretable/valued features through Agree. Thus, the transfer has to apply when Agree takes place (see also Chomsky, 2000, 2001).[7] Richards (2007), then, assumes the following.

(17) Value and Transfer of uFs must happen together. (Richards, 2007, 556,(1))

"uF" means uninterpretable/unvalued feature in this context.

Another point is the cyclic nature of the derivation. If the whole derivation is transferred, it becomes inaccessible. Thus, the edge of the phase (a phase head and its specifier) needs to be activated in the further computations as an 'escape hatch', while the complement of the phase has completed its computation.

(18) The edge and nonedge (complement) of a phase are transferred separately. (Richards, 2007, 568,(2))

Since the phase head is on the edge and edge elements are not transferred (17), when Agree takes place, C, for example, cannot be transferred because of (18). This leads to fatal crash; no derivation will converge at the interfaces. Richards (2007) proposes the following.

[7] See also Epstein & Seely (2002) for relevant discussion.

(19) uF must spread from edge to nonedge (i.e., from C to T, v* to V, etc.)

(Richards, 2007, 569,(3))

This is the argument for feature inheritance. The question, which goes back to our question, is how to transfer the root clause. Assume that the following structure is the root clause (CP).

(20) a. {wh, {C, {$_{TP}$ subj, {T, . . .

→ TP is transferred.

b. {wh, {C }}

After TP is transferred, the computation does not proceed further since it is a root clause. Obata (2010) and Goto (2011a) simply propose that at the root clause, the edge and the non-edge can be transferred. I will assume this unless there is a serious problem regarding this issue. One could think about it in this way: Transfer makes the interior domain of the phase (the complement of the phase head) inaccessible for further operations due to PIC, whereas the interfaces can access the derivation at any point. Thus, Transfer is not for the interfaces, but rather for further operations (e.g., for computational efficiency). At the root clause, the interfaces can access the derivation and interpret it. At the root clause, there are no further operations to apply, so the derivation terminates. Interfaces simply access the edge of the root clause and Transfer has nothing to do with it. Since we will discuss phasal derivations in the rest of this chapter, I will use the notion of Transfer as a component of the phase theory in the narrow syntax.

To sum up, this section introduces and demonstrates the syntactic derivations that are constructed by syntactic operations that apply at the phase level. The phase-theoretical model limits the computational domain via Transfer. The phasal edge will remain for further computation, whereas the complement of the phase head will be transferred. Nonetheless, the Transfer domain will remain in the narrow syntax. This domain is not accessible by Transfer/PIC. Thus, movement has to apply at the phase level, otherwise it will be blocked by Transfer, and movement from an original position which has been transferred cannot take place in one fell swoop to the next phase due to the PIC (i.e., a strong version of subjacency, see Chomsky 2000).

This leaves open the possibility that the phase-theoretical model allows multiple optimal derivations. As long as the derivation does not crash (e.g., leaving unvalued features, merging elements that do not satisfy selection restrictions, etc.), the derivations are legitimate.

Given phase theory and these operations under the more recent minimalist frameworks which I will introduce, we will focus on the underspecification of the rule ordering in the rest of the chapter.

The next section demonstrates that different rule orderings allow different derivations, both of which are grammatical. I will first show raising constructions in Icelandic and participle agreement in Swedish, where two syntactic operations interact, namely Move and Agree.

3.3 Move-Agree and Agree-Move order

3.3.1 Raising Constructions in Icelandic

The immediate consequence of the underspecification of rule ordering involves ordering the options of Move (, which is an instance of Merge), and Agree. Based on Chomsky (2008), Goto (2011b) already proposes such an analysis. Goto's assumptions are as follows.

(21) Goto's (2011b) approach under Chomsky's (2008) framework

a. Phase heads (C,v*) have edge-features (EPP-featurea) and uϕ-features

b. C-to-T inheritance does not take place in the V2 environment, though it does in the non-V2 environment

c. In the V2 environment, the ϕ-probe can interact with the EF-probe on C

d. There are two possible derivations interacting with Agree and Move

i. Agree-Move order

ii. Move-Agree order

Goto (2011b) shows that in Icelandic Move applies before Agree takes place. In Icelandic raising constructions with the verb 'seem', ϕ-feature agreement is optional.

(22) Mér$_i$ virðist/virðast t$_i$ hestarnir vera seinir
me-DAT seems$_{SG/PL}$ t$_i$ the-horse$_{PL\text{-}NOM}$ be slow$_{NOM}$
'It seems to me that the horses are slow.'

(Holmberg & Hróarsdóttir, 2003, 147)

The first possible derivation is as follows. In this derivation, Agree takes place first and the closest goal is found, namely *me*.

(23) Derivation 1: Agree-Move order → "seems" (singular)

a. {me, {C, {T, {seem, {me$_i$,{horses, {. . . }}}}}}

b. EM of T: {T, {seem, {me,{horse{. . . }}}}}

c. EM of C: {C, {T, {seem, {me,{horse{. . . }}}}}}

d. Agree (C, me): C has singular:

{C$_{[uF]}$,{T, {seem, {me$_{[vF]}$,{horses$_{[vF]}$, {. . . }}}}}}

IM of 'me': {me$_i$,{C, {T, {seem, {me$_i$,{horses, {. . . }}}}}}

In this derivation, since C probes its c-command domain, the closest element is 'me'. Agree takes place (Agree(C,me)). As Goto (2011b) assumes, the raising of 'me' locates *'me'* at [spec,CP] since feature inheritance does not take place in the V2 environment (21b).

In the next derivation, movement of 'me' takes place before Agree takes place. Once 'me' is moved, the probe can find 'horses'.

(24) Derivation 2: Move-Agree order → "seem" (plural)

a. {me, {T, {seem, {me$_i$,{horses, {...}}}}}}

b. IM of 'me': {me, {C, {T,{seem, {me$_i$,{horses, {...}}}}}}}

c. Agree(C,horses):

{me, {C$_{[uF]}$, {T,{seem, {me$_i$,{horses$_{[vF]}$, {...}}}}}}}

Notice that if there is an intervenor, the second derivation still gets singular inflection on the verb.

(25) Mér$_i$ virðist/?*virðast t$_i$ Jóni lika hestarnir.
me-DAT seems$_{SG/PL}$ t$_i$ John-DAT like horse-PL-NOM
'It seems to me that John likes horses.' (Boeckx, 2000, 359,(19))

In this example, even though 'me' moves first there is still an intervenor, thus the only option is to agree with *Jóni* and the number-feature can only be a singular feature.

3.3.2 Participle Agreement in Swedish

Extending the idea of underspecification of rule ordering discussed with respect to Goto's (2011b) analysis in the previous subsection, I also found that participle agreement in Scandinavian languages shows a similar pattern. First, I adopt Holmberg's (2002) assumptions as shown below.

(26) Assumptions (Holmberg, 2002)

a. Participle head (Prt) has an EPP-feature and ϕ-features, and it is a phase head.

b. The basic TP structure: [$_{TP}$ T [$_{PrtP}$ [$_{Prt'}$ VP]]]

c. Participle agreement requires a spec-head relation.

d. The expletive in (27) is base-generated in [spec,PrtP]

The following set of examples shows that when the object moves to the [spec,PrtP], agreement takes place in a spec-head configuration, whereas there is no such an agreement when the object stays *in situ* (cf. Richards, 2012).

(27) Swedish

a. Det har blivit [$_{PrtP}$ skrivet/*skrivna tre böcker om detta].
EX have been [$_{PrtP}$ written-N.SG/written-PL three books about this]
'There have been written three books about this.'

b. Det har blivit [$_{PrtP}$ tre böcker *skrivet/skrivna om detta].
EX have been [$_{PrtP}$ three books written-N.SG/written-PL about this]
'There have been written three books about this.'

(Holmberg, 2002, 86,(3), with a slight modification)

Again, we can appeal to the underspecification of rule ordering. The first possible derivation (27a) shows that the expletive is externally merged to [spec, PrtP] and agrees with the participle in a spec-head relation.

(28) Derivation 1: Merge-expletive-Agree → PrtP: SG (27a)

a. {Prt, {$_{vP}$ ... three book ...

b. EM of EX: {EX, {Prt, {$_{vP}$ ... three book ...

c. spec-head Agree(EX,Prt): {EX, {Prt, {$_{vP}$ ... three book ...

d. DP ('three books') receives default case.

Another possible derivation is (27b). The object moves to [spec,PrtP], and the object and the participle agree.

(29) Derivation 2: Move-Agree → PrtP: PL (27b)

a. EM of EX: {EX, {Prt, {$_{vP}$ … three book …

b. IM of 'three books': {three books$_i$,{EX,{Prt,{$_{vP}$ … three book$_i$ …

c. EM of T: {T, {three books$_i$,{EX,{Prt,{$_{vP}$ … three book$_i$ …

d. IM of EX: {EX$_j$, {T, {three books$_i$,{EX$_j$,{Prt,{$_{vP}$ … three book$_i$ …

e. spec-head Agree(three books, Prt):

{EX$_j$, {T, {three books$_i$,{EX$_j$ {Prt,{$_{vP}$ … three book …

Thus, the participle agreement can also be explained by the underspecification of rule ordering.

In this section, I briefly showed how different rule orderings yields a different agreement behaviors. The next section will introduce Chomsky's (2013) derivation, considering labeling theory.

3.4 Phase Level Derivations

The following sections discuss another specific model of derivations and its issues. In §3.2.3, the derivational model in Chomsky (2008) was discussed. Now, we introduce a slight modification of this model, which is proposed in Chomsky (2013). We first discuss examples of A/A′-movement in v* and C phases.

3.4.1 Subject Raising

Assume with Chomsky (2013) that (i) Merge is the simplest form: Merge (α,β) = {α,β} and Merge applies freely; (ii) unvalued features are marked by phase heads (C and v*), and feature inheritance transmits the unvalued features (ϕ-features, etc.) on phase heads to non-phase heads; (iii) the labeling algorithm, which is an instance of minimal search, finds the most prominent features on the elements in the sets to label syntactic objects (e.g., lexical items or features on LIs, see §2.2 for details.). Also assuming the predicate-internal subject hypothesis, C phases are formed by Merge in Chomsky (2013) to derive the sentence "John likes Mary" as below.

(30) a. External Merge(EM) of *John*: $\{_\alpha$ John,$\{v*, \ldots\}\}$

b. EM of T: $\{$T, $\{_\alpha$ John,$\{v*, \ldots\}\}\}$

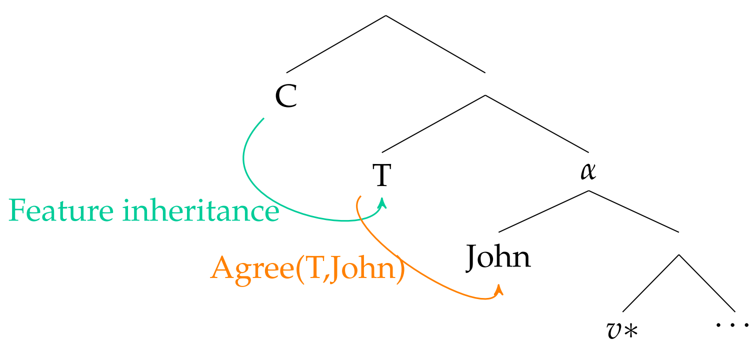
c. EM of C: $\{$C,$\{$T, $\{_\alpha$ John,$\{v*, \ldots\}\}\}\}$

d. Feature Inheritance (C to T)

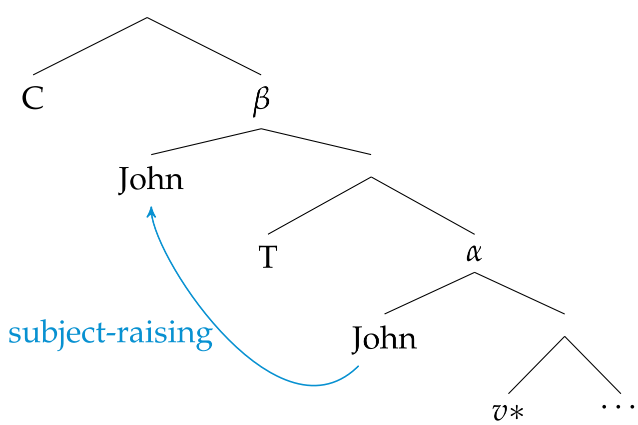
e. Agree(T,John) (the Probe=T, the goal=John)

f. Internal Merge (IM) of *John*: $\{$C,$\{_\beta$ John, $\{$T, $\{_\alpha$ John,$\{v*, \ldots\}\}\}\}\}$

(31) a.



b.



After C is introduced, unvalued features are transmitted to T, which probes the DP, namely *John*. *John* moves to the [spec, TP] position. Given the probe-goal system (Chomsky, 2000, 2001), the moved subject in [spec, TP] position is "frozen" in place. The unvalued case feature on *John* is valued as a reflex of agreement (cf. Chomsky, 2000). According to Chomsky (2007), Chomsky (2008) and Richards (2007), the valuation and transfer of the complement of the phase heads must take place at the same time (which justifies the feature inheritance). Thus, the structure labeled by β is transferred.[8]

As I reviewed in Chapter 2, section 2.2, one of the issues in this derivation in Chomsky (2013) is labeling. As a consequence of bare phrase structure (Chomsky, 1995), Chomsky (2013) proposes that the simplest formulation of the structure-building operation is Merge (Merge (α, β) = $\{\alpha, \beta\}$), which combines syntactic objects (SOs) into a set. This structure building creates hierarchical structure, but no linear order or label. Thus, the structure-building itself does not form projections (cf. Chomsky, 1995). Chomsky (2013) argues that there is a fixed labeling algorithm that searches through the whole SO to find the most prominent item (e.g., in {V, NP}, the most prominent item is V). The non-trivial cases are the structures involving two phrases (i.e., {XP, YP}). The case of (30a) shows that the whole structure labeled by α[9] includes *John* and v*P. Chomsky (2013) proposes that when one of the two phrases moves out of the structure, the label of the structure is determined by the remaining phrase. *John* in [spec, $v*$P] in (30a) moves to [spec, TP] in (30f), so that the label α is determined as v*, which is now the prominent element in the v*P. However, the structure β results is an {XP, YP} structure again. Another solution proposed by Chomsky (2013) is feature sharing. That is, when the two phrases share a prominent feature such as an agreement feature, the structure is labeled by $\langle\phi, \phi\rangle$, which shows an agreement relation.[10] Thus, the label β in (30f) becomes $\langle\phi$,

[8]One might wonder if the derivation does not follow the Markovian approach. However, this is the evaluation process that takes place close to the interfaces.

[9]This notational label is just for convenience. Again, Merge and the labeling algorithm do not attach label to the structure. If they did, it would be a violation of the Inclusiveness condition (Chomsky, 1995).

[10]Chomsky (2013) argues that just matching the feature (e.g., categorial feature) is not strong

$\phi\rangle$.

Note that operations such as feature inheritance and Agree are applied at the phase level (i.e., as part of the transfer). The rationale of the EPP can be deduced from this labeling analysis, namely, EPP is just a solution for a labeling problem, not a pure configurational requirement ("I need a spec"), given that the lack of a label causes the violation of interface conditions (e.g., the Full Interpretation principle (cf. Chomsky 1986)).

3.4.2 *Wh*-Movement

Another example is A′-movement. The example is the derivation of "Who did John like?"

(32) a. IM of wh: $\{_\alpha \text{wh}, \{ \text{John}, \{v*, \ldots, \text{wh}, \ldots\}\}\}$

b. EM of T: $\{\text{T}, \{_\alpha \text{wh}, \{\text{John}, \{v*, \ldots\}\}\}\}$

c. EM of C: $\{\text{C}, \{\text{T}, \{_\alpha \text{wh}, \{\text{John}, \{v*, \ldots\}\}\}\}\}$

d. FI

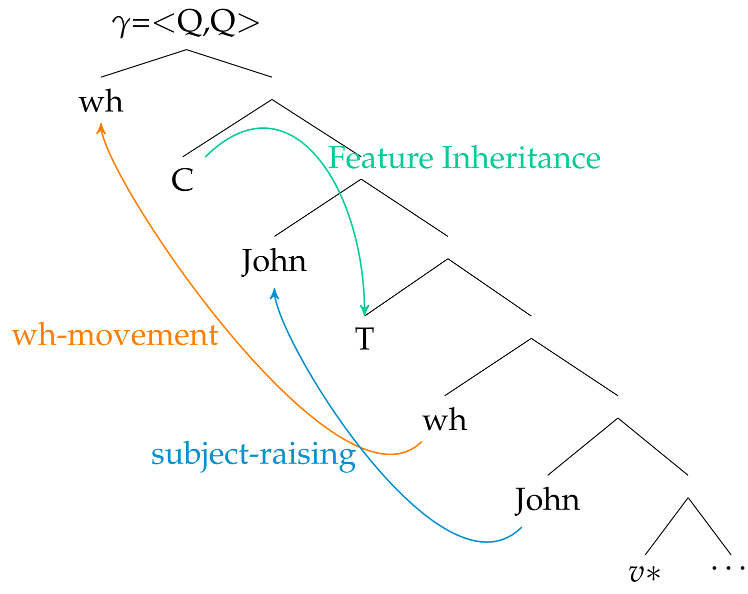
e. IM of John: $\{_\gamma \text{wh}, \{\text{C}, \{_\beta \text{John}, \{\text{T}, \{_\alpha \text{wh}, \{\text{John}, \{v*, \ldots\}\}\}\}\}\}\}$

f. IM of wh: $\{_\gamma \text{wh}, \{\text{C}, \{\text{T}, \{_\alpha \text{wh}, \{\text{John}, \{v*, \ldots\}\}\}\}\}\}$

g. Labeling, transfer

enough to share and label the structure. Agreement relations are strong enough to label the structure, which satisfies the Full Interpretation principle.

(33)



We have already discussed subject-raising, but how about *wh*-movement? Assuming with Chomsky (2008) that C probes the goal *wh*-phrase, then it attracts the *wh*-phrase to [spec,CP] via an edge-feature.[11] Then, the label γ becomes $\langle Q, Q \rangle$ since the *wh*-phrase and C share the interrogative feature Q (Chomsky, 2013, 45).

In next subsection, we will go through the different possible derivations in Obata et al. (2015), where the derivational model is based on Chomsky (2013).

3.5 Rule Ordering (I)

As we saw in the previous section, we assume with Chomsky (2013) some operations in the narrow syntax: (i) Merge, (ii) Move (which is an instance of Merge), (iii) feature inheritance, (iv) Agree, (v) labeling, and (vi) Transfer. Let's start from the derivations of two examples (Obata et al., 2015, 13) regarding T-related agreement.

[11] Although Chomsky (2013) didn't assume edge-features in his paper, I use them just for expository reasons. Rules are not strictly ordered, but we will see more details of rule ordering in the following sections. See footnote 6 for edge-features.

(34) a. Which dogs am I seeing?

b. * Which dogs are I seeing? (*in standard English)

These examples show that in standard English, T cannot ϕ-agree with the fronted *wh*-phrase and always agrees with the subject.

One of the possible derivation patterns will be (35). The example is in (36).

(35) Derivation 1: feature inheritance PRECEDES *wh*-movement to [spec,CP]

(Obata et al., 2015, 13,(32))

(36) a. *Which dogs are I seeing?

b. IM of *wh*: $\{$which dogs$_i$, $\{$I, $\{v*$, $\{$V, which dogs$_i\}\}\}\}$

c. EM of T: $\{$T, $\{$which dogs$_i$, $\{$I, $\{v*$, $\{$V, which dogs$_i\}\}\}\}\}$

d. EM of C: $\{$C, $\{$T, $\{$which dogs$_i$, $\{$I, $\{v*$, $\{$V, which dogs$_i\}\}\}\}\}\}$

e. Feature inheritance (C to T)

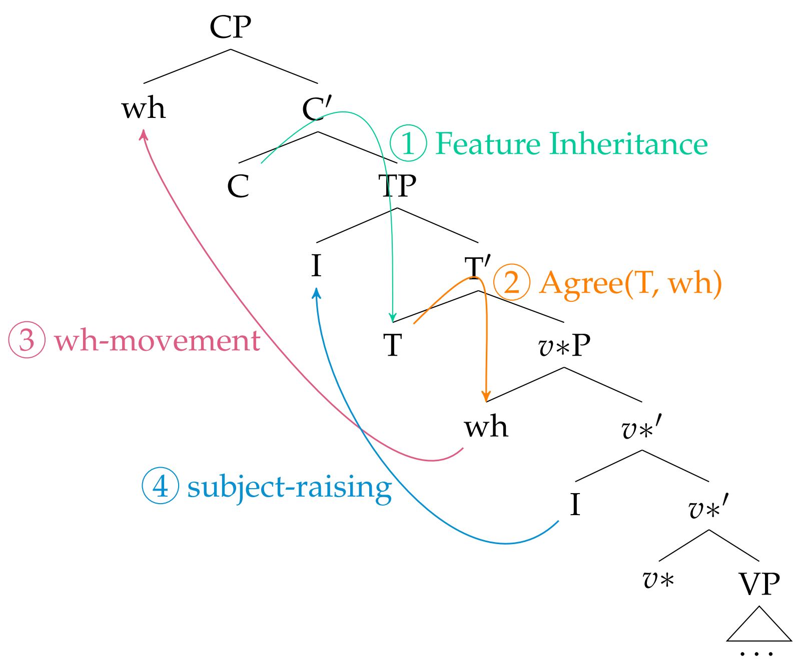
f. Agree(T, *which dogs*) $\rightarrow$ T = "are"

g. IM of *which dogs* and IM of 'I':

$\{$which dogs$_i$, $\{$C, $\{$I, $\{$T, $\{$which dogs$_i$, $\{$I, $\{v*$, $\{$V, which dogs$_i\}\}\}\}\}\}\}\}$

h. Transfer

i.



The steps are the following: (i) IM of *which dogs* to [spec,$v*$P], (ii) EM of T, (iii) EM of C, (iv) FI, (v) T agrees with *which dogs*, (vi) IM of *which dogs* to [spec,CP], (vii) IM of subject to [spec,TP]. In this derivation, the subject does not agree with T, hence the subject does not get a nominative case, and the labeling problem arises since the subject and TP structure do not share features, but the raised *wh*-phrase agrees with T. Thus, in English, this derivation becomes illegitimate.

The next derivation is T-subject agreement, which is grammatical in standard English.

(37) Derivation 2: *wh*-movement to [spec,CP] PRECEDES feature inheritance

(Obata et al., 2015, 13)

(38) a. Which dogs am I seeing?

b. IM of wh: {which dogs$_i$, {I, {v∗, {V, which dogs$_i$}}}}

c. EM of T: {T, {which dogs$_i$, {I, {v∗, {V, which dogs$_i$}}}}}

d. EM of C: {C, {T, {which dogs$_i$, {I, {v∗, {V, which dogs$_i$}}}}}}

e. IM of which dogs:

{which dogs$_i$, {C, {T, {which dogs$_i$, {I, {v∗, {V, which dogs$_i$}}}}}}}

f. Feature inheritance (C to T)

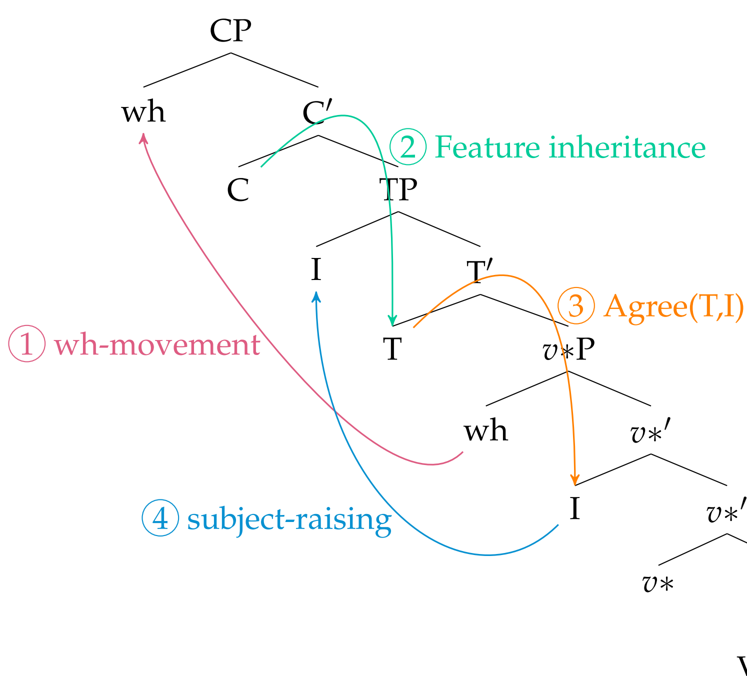
g. Agree(T,I) → T = "is"

h. IM of 'I':

{which dogs$_i$, {C, {I, {T, {which dogs$_i$, {I, {v∗, {V, which dogs$_i$}}}}}}}}

i. Transfer

j.



The derivational steps are the following: (i) EM of *which dogs* to [spec,v∗P], (ii) EM of T, (iii) EM of C, (iv) IM of IM of *which dogs* to [spec,CP], (v) FI, (vi) T agrees with subject. When T, which has unvalued ϕ-features, finds the goal, the closest one is the subject since *which dogs* is already moved up, hence invisible. Then, (vii) IM of subject to [spec,TP] takes place.

Derivation 1 (T-object agreement) is ungrammatical, while derivation 2 is grammatical in Standard English. The subject in derivation 1 does not receive case in the derivation and the labeling problem arises there. However, in the case of *wh*-constructions in Kilega, derivation 1 is grammatical as shown in Obata et al. (2015).[12]

(39) Kilega[13]

a. Bábo bíkulu b-á-kás-íl-é mámí bíkí mu-mwílo?
2that 2woman 2SA-A-give-PERF-FV 1chief 8what 18-3village
'what did those women give the chief in the village'

b. Bíkí bi-á-kás-íl-é bábo bíkulu mámí mu-mwílo?
8what 8CA-A-givePERF-FV 2that 2woman 1chief 18-3village
'What did those women give the chief in the village'

(Carstens, 2005, 220)

Kilega can have an *in situ wh*-phrase as shown in (39a), where T agrees with the subject. When the *wh*-phrase is fronted, T agrees with the *wh*-phrase. This is the pattern in (35) where T agrees with *wh*-phrase in [spec,*v**P], followed by *wh*-movement.

In the following two subsections, we discuss the derivational steps in Chomsky (2015), which abandons the counter-cylic movement of the subject and allows application of internal merge before introducing the phase heads. After these subsections, we will come back to the derivations where T-subject or T-object agreement takes place and show that Chomsky's (2015) analysis is compatible with the current approach.

[12]Note that in Kilega the case system is independent from ϕ-feature agreement (Carstens, 2005).

[13]The abbreviations in the gloss is as follows. SA: subject agreement, CA: complementizer Agreement, A: Kilega vowel /a/, FA: final vowel of Bantu verbs, PERF: perfect tense, Arabic numerals: noun classes.

3.5.1 Strictly Cyclic Derivation 1

This subsection briefly reviews Chomsky (2015), the updated version of Chomsky (2013). Assuming that labeling applies at the phase level along with other operations, the first issue is what is called the halting problem (i.e., when the movement should stop in a 'criterial position' in Rizzi's (2007) approach[14]). Consider the following example where *which dog* successive cyclically moves from the embedded object position to the matrix [spec,CP] via the embedded [spec,CP].

(40) *{$_\beta$which dog do you wonder {$_\alpha$t {$_\gamma$C$_Q$ John likes t'}}} (Chomsky, 2015, 8)

The issue causing the halting problem is the timing of the labeling. If the labeling is done for each application of Merge, α should be labeled by the Q-feature, though the *wh*-phrase moves up to the matrix [spec,CP]. The edge of the phase is not subject to labeling at the C phase in the embedded clause, but at the next phase. This is desirable, since the phrase on the edge needs to move up to the next phase. If the labeling took place for the phasal edge, every phasal edge would be a criterial position, which is not the case. When we apply the labeling algorithm later, the label of α becomes Q. But the interpretation becomes "gibberish," as Chomsky (2015) discusses.

Chomsky (2015) proposes that labeling theory requires raising-to-object for ECM constructions to get a label for the structures. An example of a derivation is below.

[14] See Rizzi (2015) for the criterial freezing approach under labeling theory.

(41) they [$_{v*P}$ expected [$_{TP}$ John to win]] (ECM)

a. {v^* {$_\alpha$DP$_i$ {R(expect) {$_\beta$ t$_i$...}}}}

b. form R-β by EM

c. IM of DP in α (EPP)

d. Merge v^*, reaching the phase level

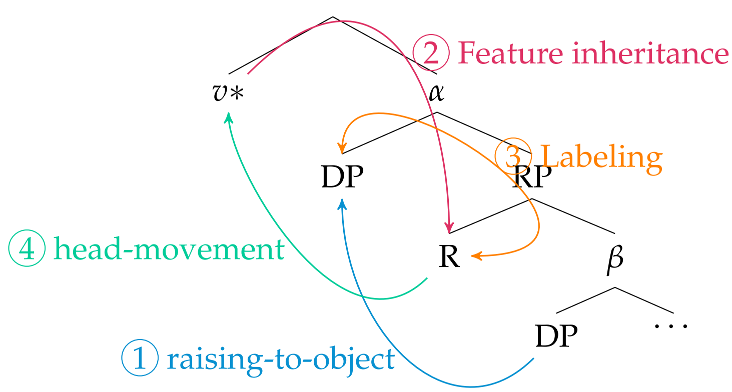
e. Inheritance

f. Labeling[15]: α is labeled $\langle\phi,\phi\rangle$

g. R raised to v^* forming R with v^* affixed, hence invisible, so phasehood is activated on the copy of R, and DP(which can be a *wh*-phrase) remains in situ, at the edge.

h. transfer of β (Chomsky, 2015, 12)

i.



A couple of clarifications are necessary. R is a root that needs a categorizer v^* to become a verb (which is an assumption in Distributed Morphology; e.g., Embick & Noyer 2007). In this derivation, feature inheritance transmits the features, including phasehood of v^*, to R. However, this phasehood is activated only when v^* becomes invisible via head movement. Chomsky (2015) assumes that head move-

[15]Although Chomsky (2015) does not mention Agree in his paper, it is natural to assume that minimal search does Agree and labeling together. See also chapter 7 in Epstein et al. (2021) for relevant discussion.

ment is done by pair-Merge, another structure-building operation that makes v^*, by definition. Pair-Merge was proposed in Chomsky (2004) to capture adjunct islands where adjuncts are opaque for extraction. Extending pair-Merge to head movement, Chomsky (2015) assumes that when R moves up to v^*, v^* is adjoined to R, which makes v^* invisible for narrow syntax (Chomsky, 2015, 12). Based on these assumptions, the example above also shows that raising-to-object needs to take place; otherwise, labeling failure occurs. DP, the object, moves to [spec,RP] (i.e., raising-to-object, R=*expect*) before $v*$ is introduced. After raising-to-object, R has ϕ-features due to feature inheritance and minimal search finds the raised object *John* in [spec,RP] and R, and Agree takes place. At the same time, since the raised object *John* and R share ϕ-features, α gets a $<\phi, \phi>$ label. After head movement of R to v^*, phasehood is activated at the lower copy of R, then β is transferred, according to Chomsky (2015) (see below for the consequences of the transfer domain in C phases).

In this derivation, the assumption in Chomsky (2008) that operations (except for external Merge) apply at the phase level is dropped. After introducing $v*$, feature inheritance takes place, where R and DP are in a spec-head relation where labeling takes place to find the ϕ-related agreement.

3.5.2 Strict Cyclic Derivation 2

There is a case where raising-to-object is assumed to be optional.

(42) a. who do you think that read the book

b. $\{_\gamma$who$_i$ do you $v*$ $\{_\epsilon$think $\{_\delta$C $\{_\alpha$t$_i$ T $\beta\}\}\}\}$

Suppose raising-to-object is obligatory in (42a). Assuming that the embedded CP moves to the matrix domain [spec,RP](R = *think*, (43b)), there would be no relevant ϕ-features, thus the derivation would be ruled out.

(43) a. $\{_{RP}$ R(think), $\{_{CP}$ C $\{_{\alpha} t_i$ T $\beta\}\}_j$ $\}\}\}$

b. $\{_{?}$ $\{_{CP}$ C $\{_{\alpha} t_i$ T $\beta\}\}_j$, $\{_{RP}$ $\{$R(think), $\{_{CP}$ C $\{_{\alpha} t_i$ T $\beta\}\}_j$ $\}\}$

Chomsky (2015) concludes that when there is no unvalued feature and movement does not provide a label, then movement does not apply.

In the CP domain, the derivational steps are similar to in the $v*$ domain, as shown below, where the derivation for the embedded CP is represented.

(44) who do you think read the book

a. $\{_{\gamma}$who do you $v*$ $\{_{\epsilon}$think $\{_{\delta}$C $\{_{\alpha}$t T $\beta\}\}\}$

b. IM of *who* from β in α (EPP)

c. Merge C, reaching the phase level

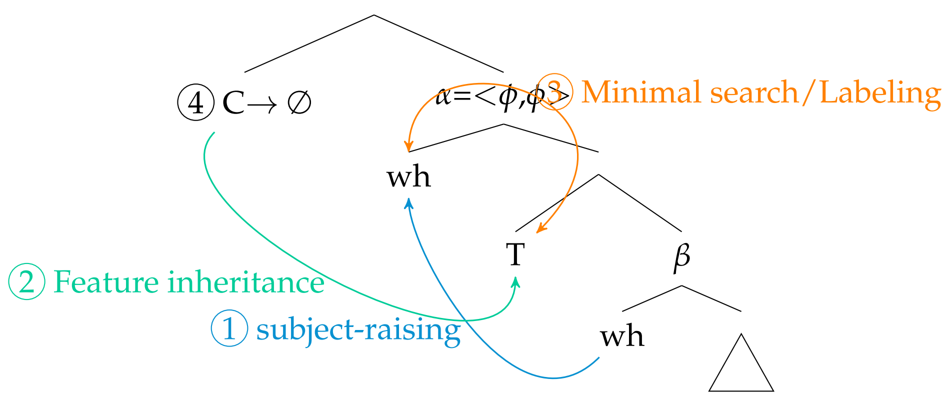
d. Inheritance

e. Labeling: α is labeled $\langle\phi,\phi\rangle$

f. deleting C, so that who can remain in situ and still be accessible to IM in the next phase

g. Transfer of the complement of T, namely, β.

h.



The *wh*-subject moves to [spec,TP] before C is introduced. After C is externally set-merged to α, feature inheritance takes place. (44f) is an idiosyncratic operation that is required for English, which activates the phasehood of T after feature inheritance. This is analogous to head movement in the *v** phase (41g), where feature inheritance takes place and phasehood is transmitted to a non-phase head (in (44), which is T). That is, in the C phase, C deletion makes C invisible. As a consequence, T is activated as a phase head, which determines the phase domain, therefore, the complement of T (i.e., β) is transferred (44g). This makes the *wh*-phrase in [spec,TP] available for further computation. If the transfer domain is TP, the *wh*-phrase never moves out of it since it is transferred. Thus, transmission of the phasehood from C to T is necessary in Chomsky (2015).

Putting aside the technicality and the validity of the steps of derivations, the main idea in Chomsky (2015) is that the strictly cyclic derivation is assumed. The subject-raising was assumed to take place counter-cyclically after the phase head C is introduced in Chomsky (2008, 2013). However, it is pointed out by Epstein et al. (2012) and Chomsky (2015) that this subject-raising is more complicated than what internal Merge does. In other words, it is a counter-cyclic movement, hence, it has to destroy the relation between C and T and merge the subject to TP, and then remerge with C. Chomsky (2015) avoids this counter-cyclic subject-raising; nevertheless, Chomsky's overall proposal motivates the subject-raising as a consequence of labeling theory, rather than just postulating an EPP-feature.

In these subsections, Chomsky's (2015) analysis was discussed. The strictly cyclic derivation requires the DP subject to move into [spec,TP]/[spec,RP] before the phase head is introduced. Moreover, minimal search finds the relevant features under a spec-head relation. As a result, the structure is labeled as $<\phi,\phi>$.

In the next subsection, I clarify some points about labeling in Chomsky (2015), and the relation between Move and Agree.

3.5.3 A Clarification on the Labeling Algorithm

Chomsky (2013) and Chomsky (2015) suggest that the timing of the labeling is at the phase level, but which SOs are labeled within a phase domain? It seems that the phasal complement is labeled. Let us start from the case of strictly cyclic raising-to-object in the framework of Chomsky (2015). (45) shows the *v** phase in general, where an external argument (EA) is merged with a predicate α (which includes *v**, R, and an internal argument (IA)).

(45) $\{_{\beta}$EA,$\{_{\alpha}$[R-v*], $\{_{\gamma}$IA$_i$,$\{_{\delta}$R, IA$_i\}\}\}\}$

a. R is activated as phase head as a consequence of head movement.

b. labeling of γ and δ

c. Transfer of the complement of R, namely the lower copy of IA.

Labeling applies at the phase level. γ is labeled as $\langle\phi,\phi\rangle$, and δ is labeled by R only after the head movement since it is strengthened by ϕ-agreement with a DP in its specifier (Chomsky, 2015, 10).

Assume that γ is not labeled at $v*$ phase level, but only the transfer domain is, namely the complement of the lower copy of R. IA (e.g., a *wh*-phrase) in [spec,RP] moves up successive cyclically and the lower copies become invisible.[16] Hence, at the next transfer domain, the label γ is the label δ. However, the label δ becomes a label after it agrees with the higher copy of IA. When IA is moving up, it is invisible at the time of the labeling at the *v** phase in (45). Hence it cannot be involved in the labeling. Thus, the labeling domain at the phase level has to be the complement of $v*$ and C. In Chomsky (2015:10), γ is labeled as $\langle\phi,\phi\rangle$ even when the *wh*-phrase moves up.

Let's look in the detail at Chomsky (2015) again. The first example is a cyclic derivation in C-phase of "John likes Mary."

[16]See §2.2 for relevant discussion.

(46) a. EM of John: $\{John, \{v*, \{R(V), Mary\}\}\}$

b. EM of T: $\{T, \{John, \{v*, \{R(V), Mary\}\}\}\}$

c. IM of John: $\{_{\alpha}John, \{T, \{John, \{v*, \{R(V), Mary\}\}\}\}\}$

d. EM of C: $\{C, \{_{\alpha}John, \{T, \{John, \{v*, \{R(V), Mary\}\}\}\}\}\}$

e. Feature inheritance (C to T)

f. Labeling Algorithm/minimal search(T,John)

g. Transfer

There is no counter-cyclic IM of the subject and the labeling of α is done by minimal search, not by the probe-goal relation. This cyclic derivation suggests that a raising-to-object (IM of Mary) analysis should be adopted, as shown in (47), which is the derivation of the v^*-phase of "John likes Mary."

(47) a. EM of R(V) and Mary: $\{R(V), Mary\}$

b. IM of Mary: $\{Mary, \{R(V), Mary\}\}$

c. EM of $v*$: $\{v*, \{Mary, \{R(V), Mary\}\}\}$

d. Feature Inheritance

e. Labeling algorithm/minimal search (R,Mary)

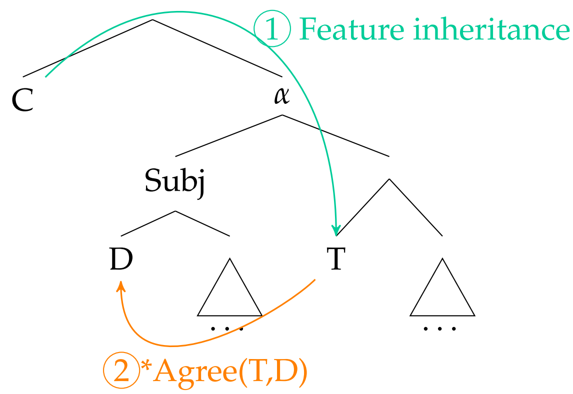
f. head movement

g. Transfer of the complement of R(V)

As we examined above, IA moves to [spec,RP] (i.e., raising-to-object) and after the head movement of R to $v*$, $v*$ becomes invisible, which triggers the activation of phasehood at the R in the lower position. As a consequence, the transfer domain becomes the complement of the lower copy of R, and IA in [spec,RP] is on the edge of the derived phase head R. In this derivation, IA does not have to move up to the [spec,v^*P] since it is already on the phase edge. We don't want to label the phasal edges, since the edge always make an unlabeled {XP,YP} structure (e.g., successive cyclic movement).

Another issue is Movement and Agree. In Chomsky (2020) and Epstein et al. (2018b), it is pointed out that if the subject cyclically moves up to [spec,TP] and feature inheritance of C to T applies, T cannot find the higher copy of the subject, since it is already moved into [spec,TP].

(48)



In order for Agree to operate properly, they independently propose that C can directly agree with the subject in [spec,TP], then feature inheritance can apply to label TP.[17]

Considering this possibility, the following is the refined version of a derivation.

(49)
a. EM of John: {John,{$v*$,{R(V), Mary}}
b. EM of T: {T, {John, {$v*$, {R(V), Mary}}}}
c. IM of John: {John, {T, {John, {$v*$, {R(V), Mary}}}}}
d. EM of C: {C,{John, {T, {John, {$v*$, {R(V), Mary }}}}}}
e. Agree(C,John)
f. FI

[17]Goto (2013b: 68, (19)) explicitly proposes the following regarding the feature inheritance and labeling.

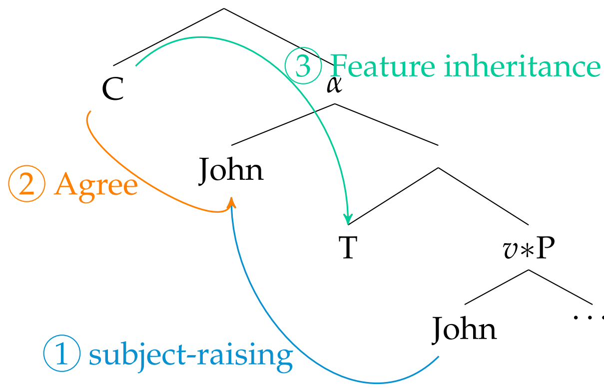
i. Feature Inheritance is required for labeling for interpretation at the interfaces.

Alternatively, assume that labels are necessary for interpretation, the role of feature inheritance is to label the TP phrase. Assume feature inheritance does not take place. Then there would be no valuation problem, but T would remain a weak head, which results in a violation of Full interpretation.

g. Labeling

h. Transfer

i.



In this derivation, Chomsky (2020) and Epstein et al. (2018b) suggest that IM does not have a prerequisite Agree. Thus, the order in which IM is followed by Agree becomes possible.[18]

In the next section, we discuss how these strictly cyclic derivations are compatible with Obata et al.'s (2015) analysis of the derivations where T-subject agreement and T-object agreement are both generable without any fatal crash at interfaces.

3.6 Rule Ordering (II)

3.6.1 The Timing of Agree/Minimal Search

Based on Chomsky (2015), Chomsky (2020), and Epstein et al.'s (2018b) approaches, which require derivation to be strictly cyclic, the derivation patterns for (50) can be derived as shown below. The examples are, again, from Obata et al.

[18]It seems it would be inconsistent to use minimal search as labeling and Agree. In Chomsky (2015), Labeling is done by minimal search. In the case of subject-raising, minimal search finds the subject in [spec,TP] and T, but if we assume probe-goal based on Agree, it seems redundant, since we need to have (i)Agree, (ii) subject raising, and (iii)labeling. If labeling and Agree are done by minimal search, why do we need these redundant steps? Chomsky (2020) and Epstein et al. (2018b) seem to assume Agree before feature inheritance, but the problem they point out would be avoided if we just assume minimal search can detect the relevant features in the configuration of spec-head. See Ke (2019) and Epstein et al. (2020) for discussion of minimal search.

(2015).

(50) a. Which dogs am I seeing?

b. *Which dogs are I seeing? (*in the standard English)

The first pattern is T-subject agreement (Derivation 1).

(51) A derivation of "Which dogs am I seeing?"

a. v^*P phase: {$v*$, {which dogs$_i$, {R, which dogs$_i$}}}

b. feature inheritance

c. labeling (R, which dogs)

d. head movement

e. transfer of the complement of R

f. EM of 'I': {I, {$v*$, {which dogs$_i$, {R, which dogs$_i$}}}}

g. EM of T: {T,{I, {$v*$, {which dogs$_i$, {R, which dogs$_i$}}}}}

h. IM of 'I': {I$_j${T,{I$_j$, {$v*$, {which dogs$_i$, {R, which dogs$_i$}}}}}}

i. EM of C: {C, {I$_j${T,{I$_j$, {$v*$, {which dogs$_i$, {R, which dogs$_i$}}}}}}}

j. IM of which dogs:

{which dogs$_i$, {C,{I$_j${T,{I$_j$,{$v*$,{which dogs$_i$, {R,which dogs$_i$}}}}}}}}

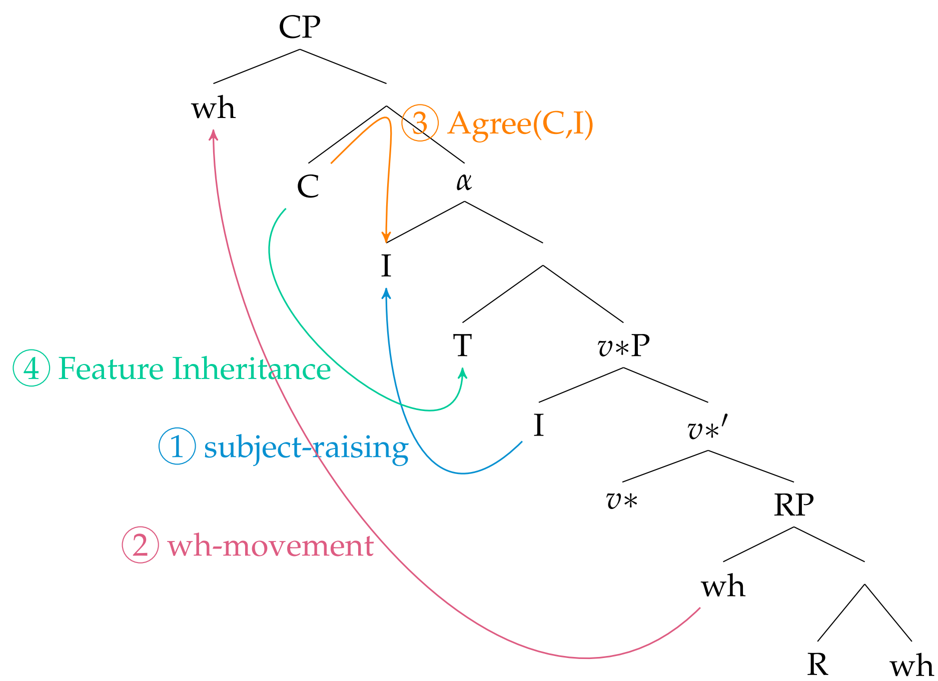
k. Agree(C,I)

l. Feature inheritance

m. Labeling for CP=Q-feature, TP=ϕ-features

n. Transfer

o.



This derivation is grammatical in English. The *wh*-phrase is not in [spec,*v*∗P] since [spec,RP] becomes the phase edge, as we discussed in the previous sections. After *which dogs* moves to [spec, CP], C agrees directly with the subject 'I', then feature inheritance applies, so there is no labeling problem.

The next derivation is the T-object agreement pattern, which is grammatical in *wh*-constructions in Kilega (Derivation 2).

(52) a. *Which dogs are I seeing?

b. *v*∗ phase: {*v*∗, {which dogs$_i$, {R, which dogs$_i$}}}

c. feature inheritance

d. labeling (R, which dogs)

e. head movement

f. transfer of the complement of R

g. EM of 'I': {I, {*v*∗, {which dogs$_i$, {R, which dogs$_i$}}}}

h. EM of T: {T,{I, {*v*∗, {which dogs$_i$, {R, which dogs$_i$}}}}}

i. IM of 'I': $\{I_j\{T,\{I_j, \{v*, \{\text{which dogs}_i, \{R, \text{which dogs}_i\}\}\}\}\}\}$

j. EM of C: $\{C, \{I_j\{T,\{I_j, \{v^*, \{\text{which dogs}_i, \{R, \text{which dogs}_i\}\}\}\}\}\}\}$

k. Feature inheritance

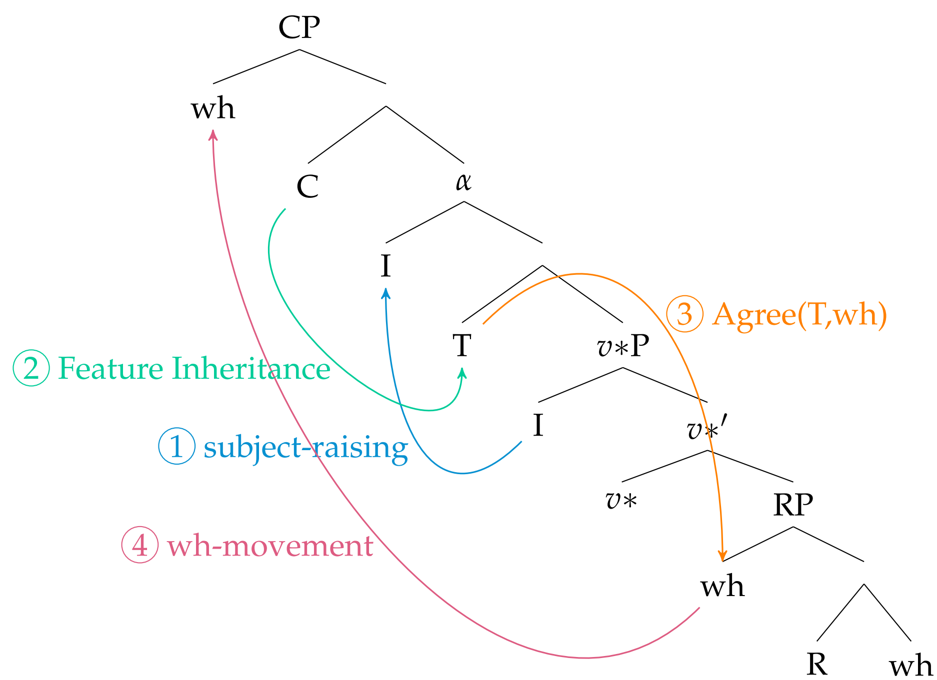
l. Agree(T,IA)

m. IM of which dogs: $\{\text{which dogs}_i, \{C,\{I_j\{T,\{I_j,\{v*,\{\text{which dogs}_i, \{R,\text{which dogs}_i\}\}\}\}\}\}\}\}$

n. Labeling for CP=Q-feature, TP=ϕ-features

o. Transfer

p.



After feature inheritance, T finds the closest goal, which is 'which dogs' since the subject 'I' is a lower copy, hence invisible. Notice that T and R agree with 'which dogs', and 'I' never agrees with T or R. Thus, 'I' is caseless in this derivation. If English is same as Kilega, there is no problem (see footnote 12).

3.6.2 Another Possible Derivation

Epstein, Obata & Seely (2018b) also propose that the *tough* construction in English is a T-object agreement derivation. That is, the derivation is the same as (52), assuming that the surface subject in the matrix clause originates in the embedded position and moves to [spec, TP] in the matrix clause.[19]

(53) a. Which dogs are tough (for me) to see

b. *v* phase: $\{v$, {which dogs$_i$, {R, which dogs$_i$}}}

c. feature inheritance

d. labeling (R, which dogs)

e. head movement

f. transfer of the complement of R

g. EM of T('to'): {T,$\{v$, {which dogs$_i$, {R, which dogs$_i$}}}}

h. EM of 'for(C) me': {for me,{T, $\{v$, {which dogs$_i$, {R, which dogs$_i$}}}}}

i. EM of 'tough':
{tough, {for me,{T, $\{v$, {which dogs$_i$, {R, which dogs$_i$}}}}}}

j. EM of T:
{T,{tough, {for me,{T, $\{v$, {which dogs$_i$, {R, which dogs$_i$}}}}}}}

k. IM of which dogs:
{{C, which dogs, {T,{tough, {for me,{T, $\{v$, {which dogs$_i$, {R, which dogs$_i$}}}}}}}}}

l. EM of C:
{C, {which dogs$_i$,{T,{tough, {for me,{T, $\{v$, {which dogs$_i$, {R, which dogs$_i$}}}}}}}}}

m. Agree(C,which dogs)

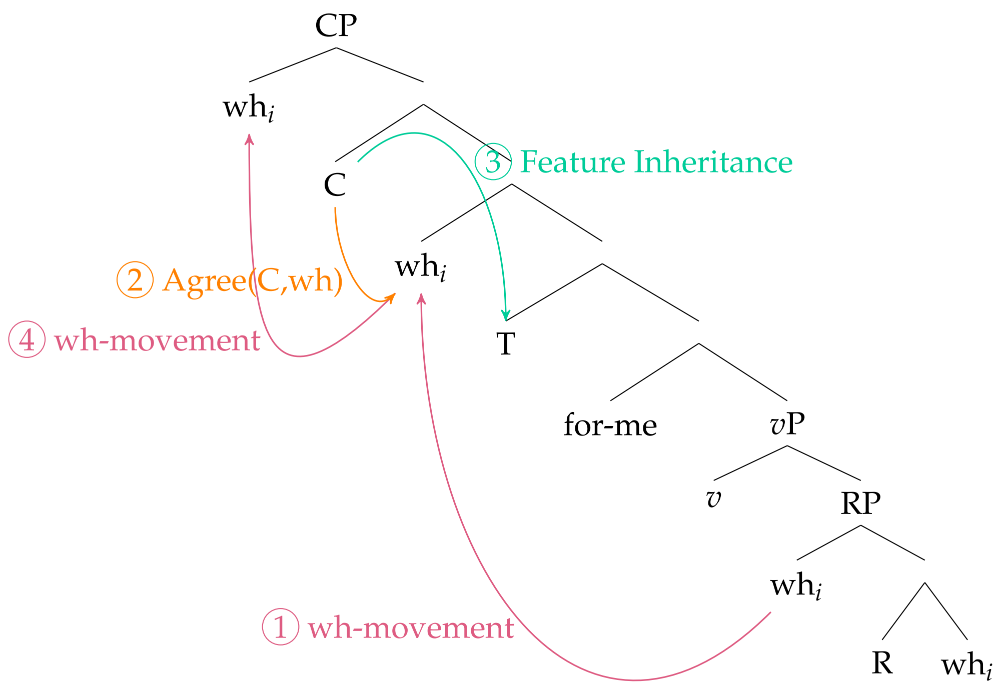
n. Feature inheritance

o. Labeling for CP=Q-feature, TP=ϕ-features

[19] But see also Chomsky (1977) for the null operator approach to the tough construction. For the recent discussion on this construction, see Hartman (2011); Fleisher (2013); Longenbaugh (2017) among many others.

p. Transfer

q.



I modify Epstein, Obata & Seely's (2018b) analysis of T-object agreement. Epstein, Obata & Seely (2018b) discuss the timing of the feature inheritance and Agree for T-object agreement, where feature inheritance has to take place before Agree since the lower copy of the subject is invisible. Thus, T agrees with IA and IA moves to [spec,CP]. In (53), however, the *wh*-phrase has to move to [spec,TP] to get nominative case since there is no EA. Instead, 'for me' is merged with *v*P, which does not move to [spec, TP]. Also, *wh*-movement to [spec,TP] has to apply before C is introduced otherwise it is counter-cyclic movement. Since *wh*-movement takes place before C is introduced, C can directly agree with the *wh*-phrase. As Chomsky (2020) and Epstein, Obata & Seely (2018b) propose, feature inheritance takes place after Agree.

If (52) and (53) are correct, then why is the following derivation (known as hyper-raising) ungrammatical? The model should predict that it is grammatical.

(54) a. *John seems is intelligent.

b. EM of 'John' and 'intelligent': {John, intelligent}

c. EM of T: {T, {John, intelligent}}

d. IM of 'John':{John$_i$,{T, {John$_i$, intelligent}}}

e. EM of C: {C, {John$_i$,{T, {John$_i$, intelligent}}}}

f. Feature Inheritance

g. deletion of C

h. labeling:

i. EM of 'seems': {seem,{John$_i$,{T, {John$_i$, intelligent}}}}

j. EM of T: {T, {seem,{John$_i$,{T, {John$_i$, intelligent}}}}}

k. EM of 'John'" {John$_i$,{T, {seem,{John$_i$,{T, {John$_i$, intelligent}}}}}}

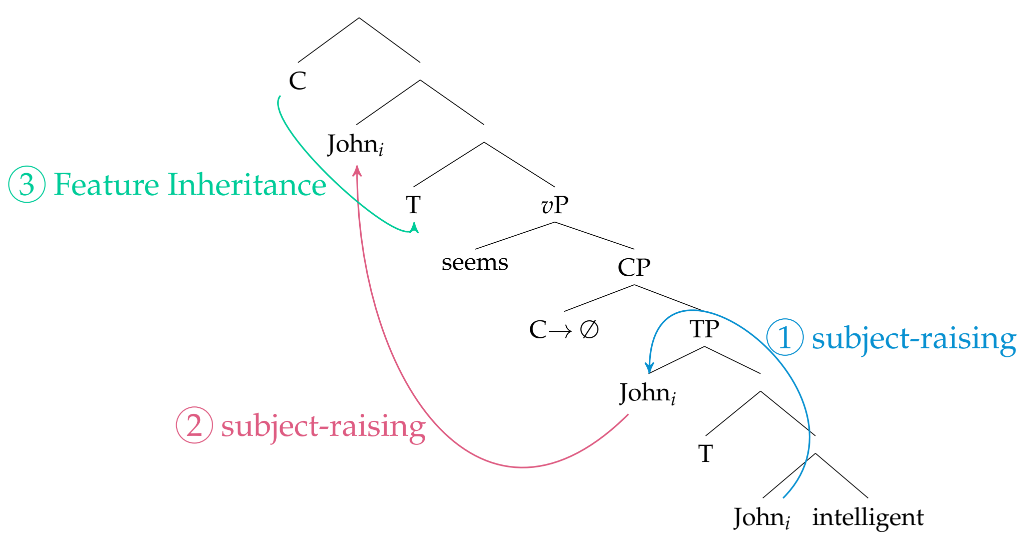
l. EM of C: {C,{John$_i$,{T, {seem,{John$_i$,{T, {John$_i$, intelligent}}}}}}}

m. Feature Inheritance:

n. labeling (minimal search)

o. Transfer

p.



In (54), at the embedded C-phase, deletion of C applies,[20] and [spec, TP] becomes the phase edge. Then T in the matrix clause finds the embedded subject, which moves to [spec,TP] in the matrix clause. This is based on Chomsky (2015), which makes the embedded [spec,TP] an escape hatch. As we discuss in chapter 2, the derivation of hyper-raising (e.g., (54)) is banned by theta theory under MERGE (see §2.7.5). Thus, I conclude here that the derivation for hyper-raising is generable by narrow syntax, but other principles such as theta theory and FC imply that this derivation cannot be interpreted properly.

In this section, we discussed how T-subject agreement and T-object agreement take place under the framework in Chomsky (2015), Chomsky (2020), and Epstein, Obata & Seely (2018b). T-subject agreement is derived by the direct agreement between C and the subject in [spec,TP], followed by feature inheritance, whereas T-object agreement is derived by the application of feature inheritance, followed by Agree.

3.7 Possible Complementizer Agreement

In the previous sections, we have discussed different derivational models that capture the underspecification of the rule ordering, including Chomsky (2013), Chomsky (2015), Chomsky (2020), and Epstein et al. (2018b).

In this section, we will discuss complementizer agreement with special reference to Haitian Creole (HC), Cabo Verdean Creole (CVC), and Brazilian Portuguese. The distribution of the morphological realization of the complementizer is different between these languages. I argue with Obata et al. (2015) that the difference comes from the different application of rules in narrow syntax. Moreover,

[20]Suppose C deletion does not takes place. *John* in the embedded [spec,TP] has to move to the phasal edge, namely [spec,CP]. This does not properly label the embedded TP structure, since at the phase level, it has already moved to [spec,CP] and the copy which is located in [spec,TP] is invisible for labeling.

I extend Obata et al.'s (2015) proposal to the current derivational model where the derivations proceed in a strictly cyclic way. As we discussed, Obata et al.'s (2015) analysis was based on Chomsky (2008, 2013), where subject raising takes place after the phase head C is introduced, which is a counter-cyclic movement. The same derivation is adopted in the derivations of HC and CVC in their analysis. Thus, we are going to show that the underspecification approach is still applicable in the strictly cyclic derivational model.

3.7.1 Complementizer in HC

This and the next subsection discuss the timing of movement and agreement in Haitian Creole and Cabo Verdean Creole. The first examples show that the complementizer *ki* in Haitian Creole is realized when the *wh*-phase is a subject, while when the *wh*-object is in [spec,CP], there is no realization of *ki*.

(55) a. Kilés ki te wé Mari?
who COMP ANT see Mari
'Who saw Mari?'

b. *Kilés te wé Mari?
who ANT see Mari
'Who saw Mari?'

(56) a. Kilés Mari te wé?
what Mari ANT see
'Who did Mari see?'

b. *Kilés ki Mari te wé?
who COMP Mari ANT see
'Who did Mari see?'

Takahashi & Gračanin-Yuksek (2008) propose that when this complementizer and the *wh*-phrase agree in both Q and ϕ-features, the complementizer is morphologically realized.

(57) "C is spelled out as *ki* only if both *u*wh and *u*ϕ-features on C are checked off by a single goal."

(Takahashi & Gračanin-Yuksek, 2008, 229)

The assumption here, based on Germanic languages (Carstens 2003; Haegeman & van Koppen 2012) and Bantu languages (Carstens, 2005), is that the C^0 head possesses unvalued ϕ-features that are valued under Agree.

(58) Kilega

Bikí	bi-á-kás-íl-é	bábo	bíkulu	mwámí	mu-mwílo?
8what	8CA-A-give-PERF-FV	2that	2woman	1chief	18-3village

'what did those women give the chief in the village' (Carstens, 2005, 220)

Carstens (2005) argues that (58) shows the complementizer agreement with the *wh*-phrase.

Building on Takahashi & Gračanin-Yuksek's assumption (57), Obata et al. (2015) argue that the steps of the derivation for subject *wh*-extraction in HC are the following (see §3.4 for details of the derivational model, based on Chomsky (2013)).

(59) a. {C, {wh_i, {T, {wh_i, {$v*$, {V,Obj}}}}}}

b. EM of C

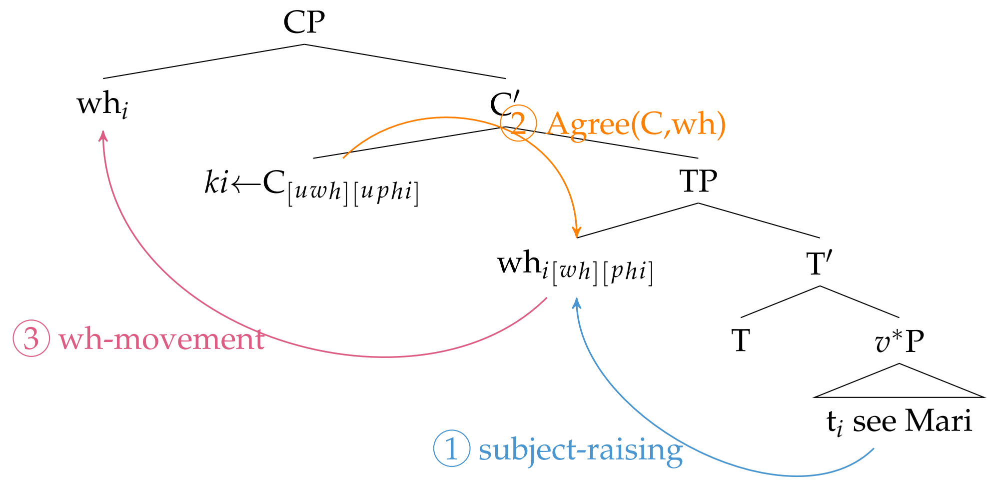
c. IM of the *wh*-subject to [spec,TP][21]

d. Agree(C,wh)

e. IM of the *wh*-phrase to [spec,CP]

[21]One might wonder why the subject raising takes place at this point, as Obata et al. (2015) assume. Suppose the subject does not move to [spec,TP] before C agrees with it. Since the subject in [spec,*v**P] is hierarchically higher than the object, C agrees with the subject anyway. The result is the same as what is shown in (59). Again, once we adopt the underspecification of rule ordering, these two options should be fine. In this case, the different rule orderings do not result in different representations.

f.



At the phase level, C agrees with the closest element, namely the *wh*-phrase, and the *wh*-phrase moves to [spec, CP]. Since C agrees with the *wh*-phrase, which has *wh*- and ϕ-features, the condition (57) is satisfied, so *ki* is morphologically realized.

As for object *wh*-extraction in HC (Obata et al., 2015, 8), the steps are shown below.

(60) a. {C, {subj$_i$, {T, {wh$_j$, {subj$_i$, {v^*,{V,wh$_j$}}}}}}}

b. EM of C

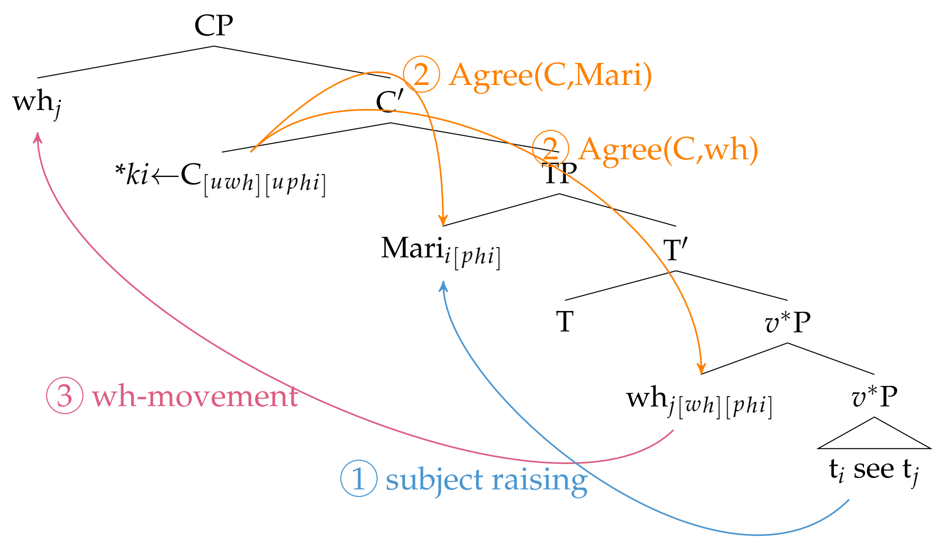
c. IM of the subject to [spec,TP]

d. Agree(C,subj)

e. Agree(C,wh)

f. IM of the *wh*-phrase to [spec,CP]

g.



Assume that the *wh*-phrase moves to [spec,v*P], otherwise it gets transferred. At the phase level of C, the subject moves to [spec,TP] and C agrees with the subject *Mari* in terms of ϕ-features. C still has an unvalued *wh*-feature, which finds the object *wh*-phrase in [spec,v^*P] to agree with. Then the object *wh*-phrase moves to [spec,CP]. In this case, *ki* is not morphologically realized based on the generalization in (57). That is, C agrees with different elements separately in this derivation.

Notice that in this derivation, the subject in [spec,v^*P] has to move to [spec,TP] before the C agreement. Then, C separately agrees with the subject for ϕ-features and with the *wh*-phrase for the Q-feature.

In this subsection, I reviewed Obata et al.'s (2015) derivations of both subject and object *wh*-extraction in HC. In both cases, the subject raising takes place before C agreement takes place. In the next subsection, Obata et al.'s (2015) derivations of both subject and object *wh*-extraction for CVC are discussed.

3.7.2 Complementizer in CVC

Cabo Verdean Creole also has the complementizer *ki* and it is realized in both subject and object *wh*-phrases (see also Baptista & Obata 2015).

(61) a. Kenhi ki odja João?
who COMP see João
'Who saw João?'

b. *Kenhi odia João?
who see João
'Who saw João?'

(62) a. Kuze ki nhos odja?
what COMP you see
'What did you see?'

b. *Kuze nhos odja?
what you see
'What did you see?'

Obata et al. (2015) argue that the condition on the realization of *ki* is same in HC (57). That is, *ki* is spelled out when C agrees with a single element that has both *wh*- and ϕ-features. Assuming that the properties of the complementizer are the same as in HC, Obata et al. (2015: 8) argue that the steps of the derivation for subject *wh*-extraction in CVC are as follows.

(63) a. $\{C, \{T, \{wh_i, \{v^*, \{V, Obj\}\}\}\}\}$

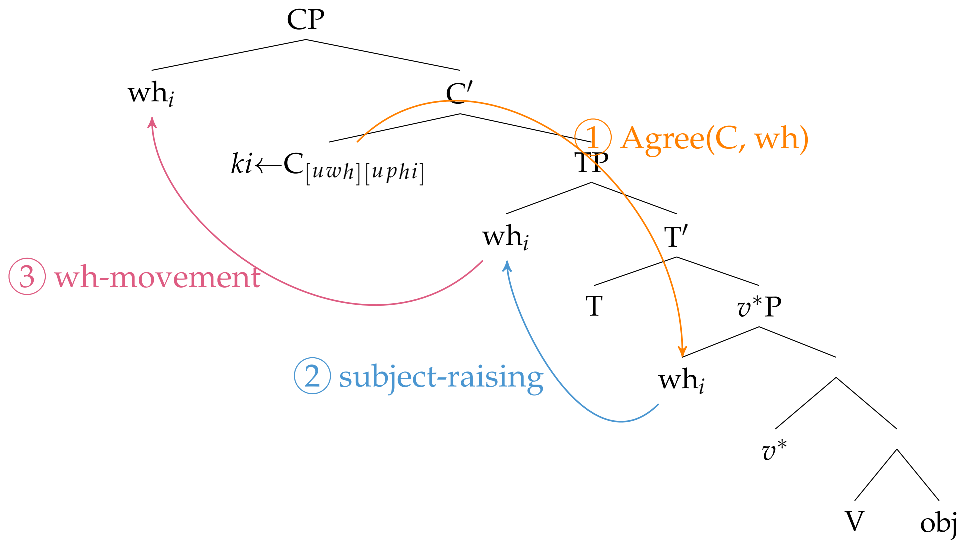
b. EM of C

c. Agree(C,wh)

d. IM of the *wh*-subject to [spec,TP]

e. IM of the *wh*-subject to [spec,CP]

f.



C agrees with the *wh*-subject in [spec,v^*P] in both the *wh*-feature and ϕ-features. This again triggers the morphological realization of *ki*. After Agree takes place between C and the *wh*-phrase, the *wh*-phrase moves up to the [spec, CP] via [spec,TP].

Next is the case of object *wh*-extraction in CVC (Obata et al., 2015, 8).

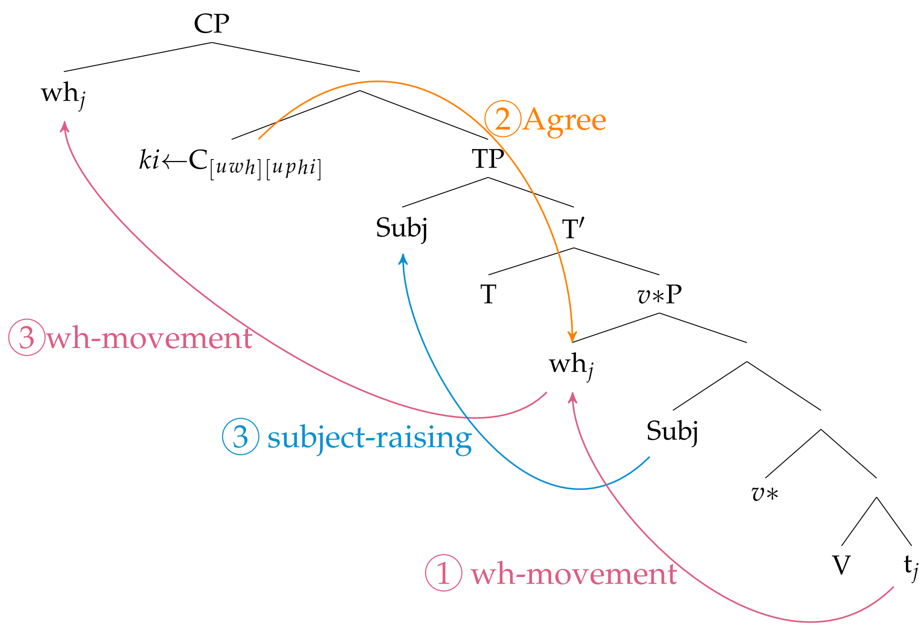
(64) a. {C, {T, {wh$_j$, {subj$_i$, {$v*$,{V,wh$_j$}}}}}}?

b. Agree(C,wh)

c. IM of the *wh*-phrase to [spec,CP]

d. IM of the subject to [spec,TP]

e.



Assume that at the v^* phase level, the *wh*-phrase moves to the edge of [spec,v^*P], which is higher than the subject in [spec,v^*P]. After C is introduced, Agree takes place first. C finds the higher element, namely the *wh*-phrase. Notice that Agree takes place before subject raising from [spec, v^*P] to [spec,TP]. C agrees with a single element, the *wh*-phrase, in terms of both the *wh*-feature and ϕ-features, which triggers the morphological realization of *ki*.

To summarize Obata et al. (2015), in both HC and CVC, the morphological realization of the complementizer *ki* depends on how C agrees with the *wh*-phrase; the complementizer *ki* is spelled out when C agrees with a single element (i.e., a *wh*-phrase) in both the *wh*-feature and ϕ-features (Takahashi, 2008) (see (57)).

	wh-subject	*wh*-object	the order of rules
HC	*ki*		subject-raising precedes C-Agree
CVC	*ki*	*ki*	C-Agree precedes subject-raising

Table 3.1: A summary of the complementizer *ki* in HC and CVC (cf. Obata et al., 2015, 10)

The difference between HC and CVC is due to the different rule orderings, that is, whether subject-raising takes place before or after C-Agree. In *wh*-object constructions (60) in HC, subject raising takes place before Agree takes place, which prevents C from agreeing with the *wh*-object in both *wh*- and ϕ-features. Thus, this derivation does not show the morphological realization of *ki*. On the other hand, in (64) in CVC, subject raising takes place after Agree. Then C Agrees with the *wh*-object in both *wh*- and ϕ-features, which explains the morphological realization of *ki* in object *wh*-questions as well.

One conceptual question concerns the derivations for CVC (63) and (64). In both derivations, C agrees with the *wh*-object before subject raising, but this implies a counter-cyclic movement, which is banned in Chomsky (2015) (see also §2.8.2.2 for the discussion of why MERGE cannot generate counter-cyclic movement.).

Let us assume a cyclic derivation which was not explored in Obata et al. (2015) and Epstein, Obata & Seely (2018b). In §3.5.2, we discussed Chomsky's (2015) derivations where the subject raising takes place before reaching the phase level (e.g., introducing a phase head C). Thus, subject raising has to apply before C is introduced for any derivations.

Considering the conceptual argument for strictly cyclic derivations, I suggest the following derivation which shows a strictly cyclic version of *wh*-subject extraction in CVC. That is, the subject is raised to [spec,TP] before C is introduced.

(65) a. {C, {wh$_i${T,{v^*,{wh$_i$,{V,Obj}}}}}}

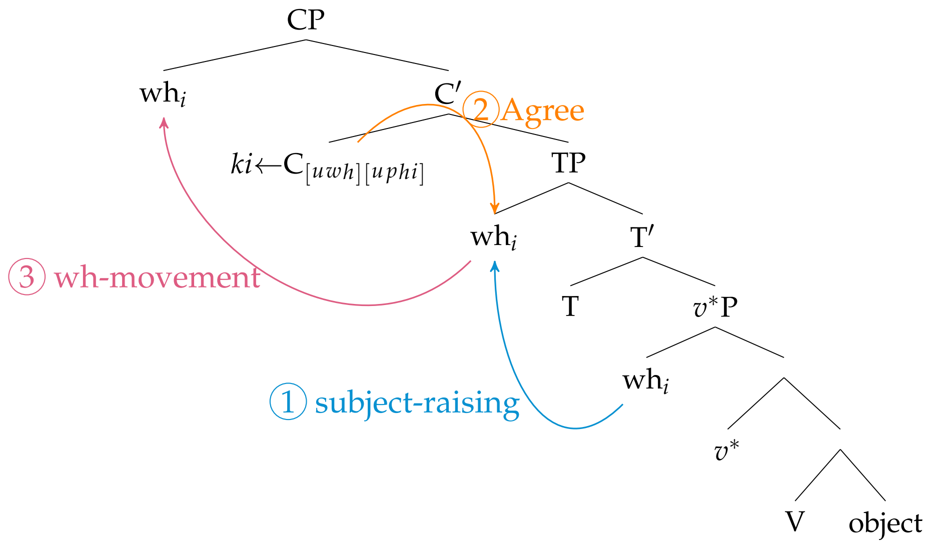
b. IM of the subject to [spec,TP]

c. EM of C

d. Agree(C,wh)

e. IM of the *wh*-phrase to [spec,CP]

f.



Since the *wh*-subject moves to [spec,TP] before C is introduced, when C finds the closest element, the *wh*-subject is found and C agrees with it in terms of both the *wh*-feature and ϕ-features. Thus, *ki* is morphologically realized. This derivation is representationally similar to (63). In (63), subject raising precedes C-Agree, but counter-cyclic movement is involved in Obata et al. (2015). That is, in Obata et al.'s derivation in (63), after C is introduced in the narrow syntax, subject raising takes place and then C-agreement takes place. I suggest here that (65) is the legitimate derivation for *wh*-subject extraction in CVC to capture the morphologically realized *ki*.

The next derivation is *wh*-object extraction. Since the strictly cyclic derivational model in Chomsky (2015) does not allow counter-cyclic subject movement, we need to find another way to yield a derivation where C agrees with a single element, namely, the *wh*-object. In the following I will show two possible derivations, though I will show that the latter derivation is the one that is allowed.

The first possible derivation for *wh*-object extraction in CVC is shown below.

(66) a. {wh, {C, {subj, {T,{subj$_i$, {$v*$, {wh$_j$, {V,wh$_j$}}}}}}}}?

b. IM of Subj to [spec,TP]

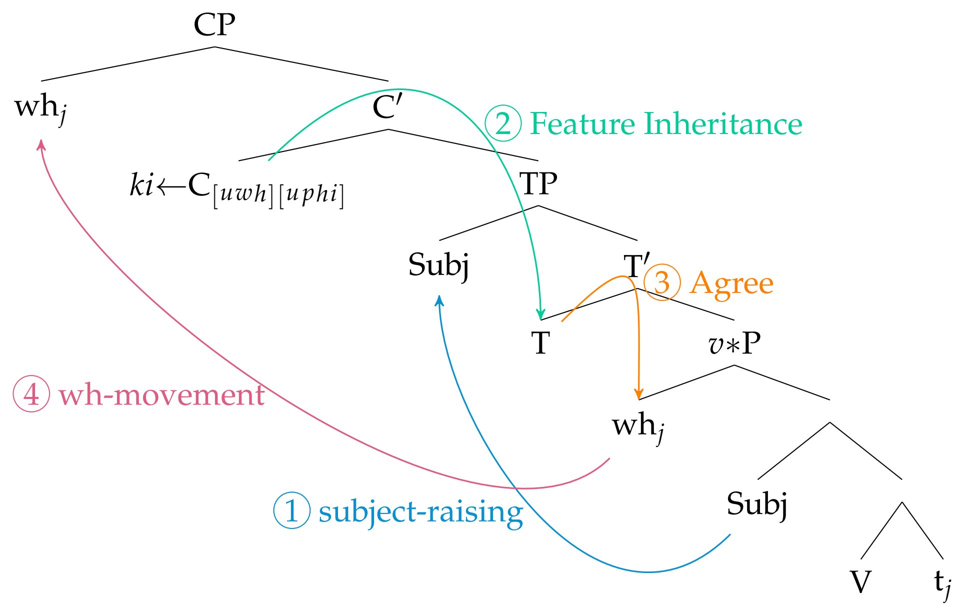
c. EM of C

d. Feature inheritance

e. Agree(T,wh)

f. IM of the *wh*-phrase to [spec,CP]

g.



In order to make the derivation strictly cyclic, the subject has to move to [spec,TP] before C is introduced. From C, the closest goal is the raised subject. However, in *wh*-object extraction, this is not predicted since C has to agree with the *wh*-object in both *wh*- and ϕ-features. Suppose feature inheritance applies and the relevant features are transmitted to T before Agree takes place. Then T agrees with the *wh*-object in both *wh*- and ϕ-features, and then the *wh*-object moves to [spec,CP]. Unfortunately, this derivation cannot correctly predict that the complementizer *ki* is morphologically realized since what agrees with the *wh*-object is T, not C.

Alternatively, I suggest that the following derivation is possible.

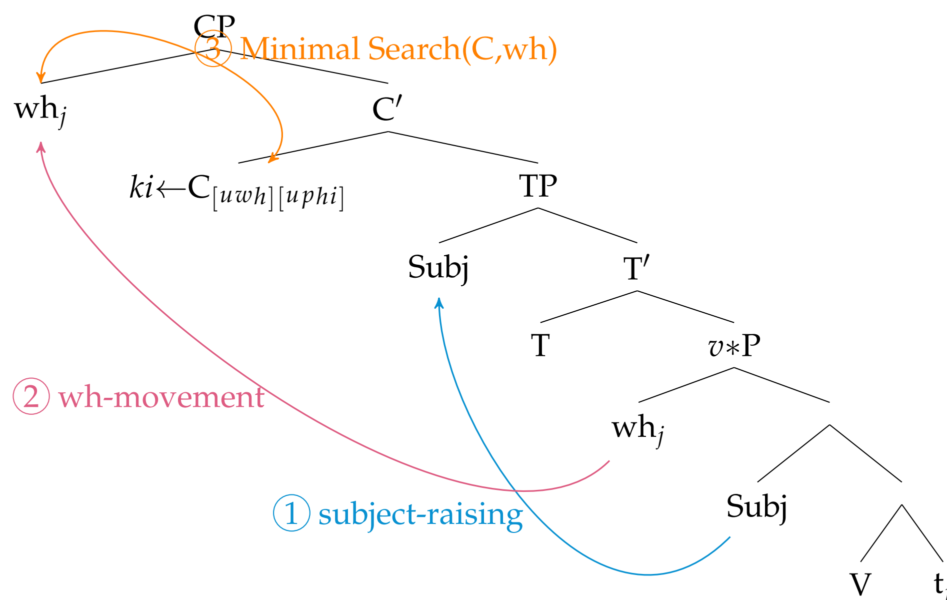
(67) a. {wh, {C, {subj, {T,{subj$_i$, {$v*$, {wh$_j$, {V,wh$_j$}}}}}}}}?

b. IM of Subj to [spec,TP]

c. EM of C

d. IM of the *wh*-phrase to [spec,CP]

e. Minimal Search (C,wh)

f. 

First, subject raising takes place and C is introduced, then *wh*-movement takes place. C directly agrees with the *wh*-object in [spec,CP] in a spec-head relation via minimal search. In fact, Baptista & Obata (2015) argue that the complementizer *ki* in CVC is realized only in a spec-head configuration. Since the *wh*-feature and ϕ-features on C are valued by a single element (i.e., *wh*-object), the complementizer *ki* is morphologically realized.

To summarize, the previous subsection and this subsection discussed the possible derivations for complementizer agreement in HC and CVC in terms of different derivational models (Chomsky 2013 and Chomsky 2015). Obata et al. (2015) explore two different rule orderings in HC and CVC that make predictions about the realization of the complementizer *ki* based on Takahashi & Gračanin-Yuksek (2008). Our particular interest was how to capture *wh*-object extraction in CVC, since in order to yield the derivation properly, counter-cyclic movement was assumed in Obata et al. (2015). However, Chomsky (2015) and the MERGE framework (§2.8.2.2) do not allow counter-cyclic movement, since it is an illegitimate operation (not allowed by the Merge operation). I suggested that we can avoid counter-cyclic movement, but capture the derivation where the *wh*-object in CVC can agree with C, so the complementizer *ki* is morphologically realized, as shown in (67).

In the next subsection, I discuss an unexplored pattern of complementizer agreement in Brazilian Portuguese where complementizer agreement is optional in both *wh*-subject extraction and *wh*-object extraction.

3.7.3 On Optional Complementizers

Brazilian Portuguese (BP) shows another relevant pattern. The complementizer is optionally realized in both local subject *wh*-questions and local object *wh*-questions as *que*.

(68) Brazilian Portuguese

a. Quem (que) viu o João?
who (C) saw the João?
'who saw João'

b. Quem (que) você viu?
what (C) you saw
'what did you see'

This is different from Standard French where the complementizer *qui* cannot occur with the *wh*-subject/object in the matrix clause.

(69) French

a. *Quel garçon qui est venu?
which boy C has come
'which boy has come' (Takahashi & Gračanin-Yuksek, 2008, 234,fn12)

b. *Qui que tu as vu?
who that you have seen
'Who did you see' (Kayne 1976, adopt from Torrence 2013, 245, (37a))

In BP, the optional *wh-in-situ* pattern is also allowed, but the complementizer is not morphologically realized.

(70) Você viu quem?
you saw who
'Who did you see' (Kato, 2013, 178,(7b))

In the *wh-in-situ* question, it could be interpreted as ordinary question or an echo question.[22] Thus, in BP, when the *wh*-phrase is fronted, the morphological realization of the complementizer is optional, whereas the complementizer cannot be realized when the *wh*-phrase stays *in situ*.

In what follows, I argue that the absence/presence of the complementizer in BP is due to different rule orderings as well, while in the *wh-in-situ* cases, I argue that there is no way of realizing the complementizer since there is a closer candidate for Agree with C (that is, there is an intervention effect).

The first pattern is the presence of the complementizer with the *wh*-object. This is the same as CVC, not the same as HC.

(71) Quem que você viu?

a. $\{C, \{subj, \{T,\{subj_i, \{v*, \{wh_j, \{V,wh_j\}\}\}\}\}\}$

b. IM of the subject to [Spec,TP]

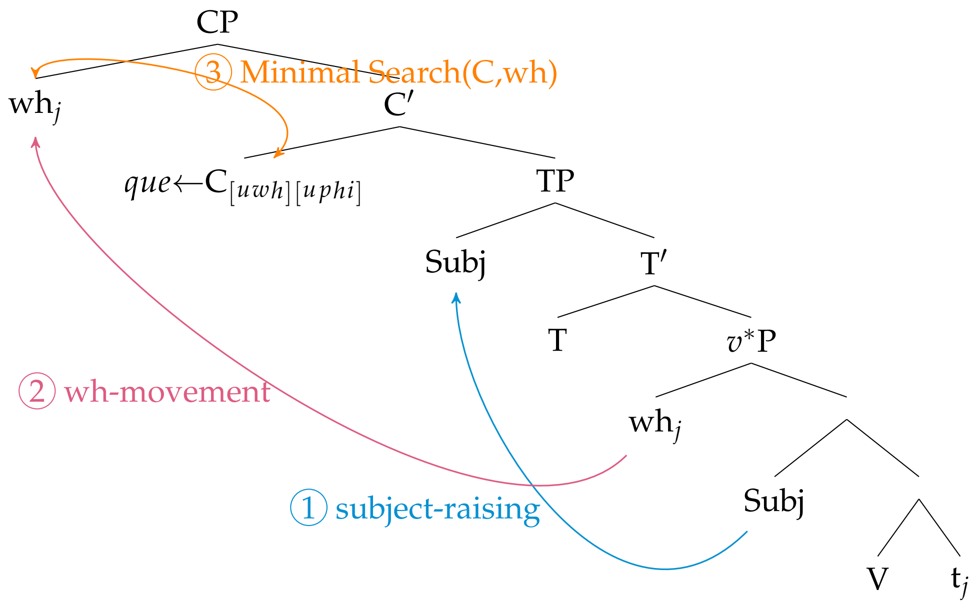
c. EM of C

d. IM of the *wh*-phrase to [spec,CP]

e. Agree(C,wh)

[22]It also depends on the prosody. See Pires & Taylor (2007); Kato (2013) for details.

f.



In this derivation, as we discussed with *wh*-object extraction in CVC, the subject raises to [Spec,TP], then the *wh*-phrase moves to [spec,CP]. Then C agrees with the *wh*-object in both the *wh*-feature and ϕ-features in a spec-head relation via minimal search, which results in the morphological realization of *que*.

Next is the absence of the complementizer in the *wh*-object extraction case, which is the same pattern as HC.

(72) Quem você viu?

a. {C, {subj$_i$, {T, {wh$_j$, {subj$_i$, {$v*$,{V,wh$_j$}}}}}}}

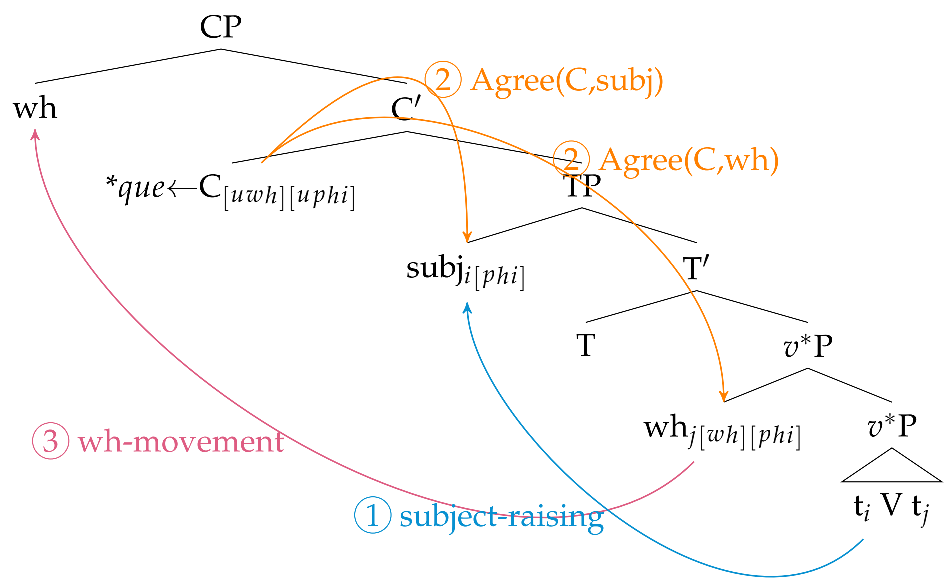
b. IM of the subject to [spec,TP]

c. Agree(C,subj)

d. Agree(C,wh)

e. IM of the *wh*-phrase to [spec,CP]

f.



In this derivation, C agrees with the subject in terms of ϕ-features since the subject raising takes place before Agree. At the same time, C also agrees with the object in terms of the *wh*-feature. Since C does not agree with the *wh*-object in terms of both ϕ-features and the *wh*-feature, the complementizer is not morphologically realized.

Next is the presence of the complementizer *que* with the *wh*-subject. This is the same as the HC and CVC *wh*-subject cases.

(73) Quem que viu o João?

a. {C, {T, {wh$_i$, {v*,{V,Obj}}}}}

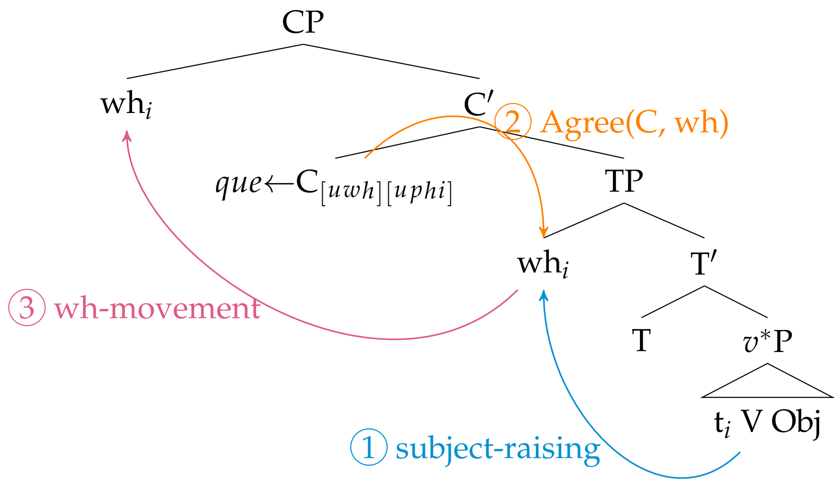
b. IM of the subject to [spec,TP]

c. EM of C

d. Agree(C,wh)

e. IM of the *wh*-phrase to [spec,CP]

f.



The *wh*-subject moves to [spec,TP], and C agrees with the subject. Then *wh*-movement takes place. In this derivation, C agrees with the *wh*-subject in both *wh*- and ϕ-features, thus *que* is morphologically realized.

Finally, the last pattern is the absence of the complementizer with the *wh*-subject.

(74) Quem viu o João?

a. {C, {wh$_i$,{T, {wh$_i$, {v*,{V,obj}}}}}}

b. IM of the subject to [spec,TP]

c. EM of C

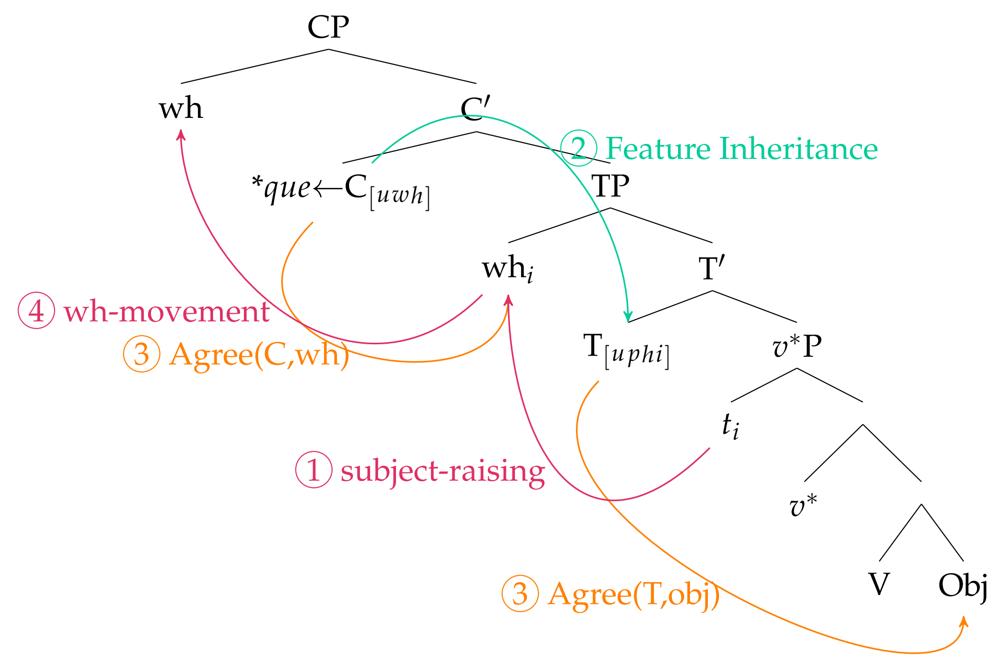
d. Feature inheritance (C-T)

e. Agree(T,obj)

f. Agree(C,wh)

g. IM of the *wh*-phrase to [spec,CP]

h.



This is a T-object agreement pattern where the *wh*-subject moves to the [spec,CP] without ϕ-agreement with C or T. Interestingly, both HC and CVC lack this pattern.[23]

To summarize, we discussed the different distributions of the complementizer *ki*/*que* in HC, CVC, and BP (See Table 3.2).

	wh-subject	*wh*-object
HC	*ki*	∅
CVC	*ki*	*ki*
BP	(*que*)	(*que*)

Table 3.2: The distribution of the complementizer in three languages

The different distributions suggest that the ordering of the application of Agree, feature inheritance, and movement derives different patterns of the realization of

[23]Another possible derivation is that the sentence involves a cleft formation, as suggested by Kato (2013)

i. (É) quem (que) chegou?
   is who that arrived
   'Who has arrived?'

As the example above shows the copula can be reduced and the complementizer has to be reduced/deleted.

the complementizer in HC, CVC, and BP.

3.8 Merge of Heads and Linguistic Variation

The last type of rule-ordering underspecification approach that I address in this chapter is the possible combination of functional categories.[24]

In the previous sections, the underspecification approach was illustrated to show some inter-/intra- I-language variation with a focus on more construction-specific phenomena. However, in the Principles & Parameters approach, macro-parameters such as the null subject parameter were argued to show cluster effects, which derived relevant phenomena (e.g., free subject inversion, the absence of a complementizer-trace effect, etc; see Rizzi 1982; Huang & Roberts 2017; Roberts 2019 among many others), and the idea was that once a macro-parameter is activated, these relevant phenomena follow. However, in the minimalist literature, the Borer-Chomsky conjecture was assumed. That is, the lexical item in the lexicon is the locus of variation. This approach cannot say much about cluster effects since the parametric variation is associated with each lexical item.

In this section, however, I argue that it is possible to derive such a cluster effect from the underspecification approach. The goal in this section is not to recapture the cluster effect in the Principles & Parameters approach (Chomsky, 1981), but to capture a certain type of cluster effect without the notion of (macro-)parameters. More specifically, we are interested in functional categories; how the functional categories are introduced to the narrow syntax derives the linguistic variation in a systematic way.

In what follows, Chomsky (2015) and his subsequent works will be discussed. In Chomsky (2015), when V-to-v^* movement takes place, V and v^* become an amalgam with v^* adjoined to the V. Extending this idea, Epstein, Kitahara & Seely

[24] This section is the revised and extended version of Blümel, Goto & Sugimoto (2021, 2022).

(2016) propose that V and v^* can form an amalgam before they are introduced to the narrow syntax. Based on the amalgamation theory, I will propose that this becomes the locus of the variation in §3.8.2. Based on that proposal, I will seek theoretical consequences in §3.8.3 and empirical consequences in §3.8.4. §3.8.5 captures the intra-linguistic variation in Basque based on the proposed analysis.

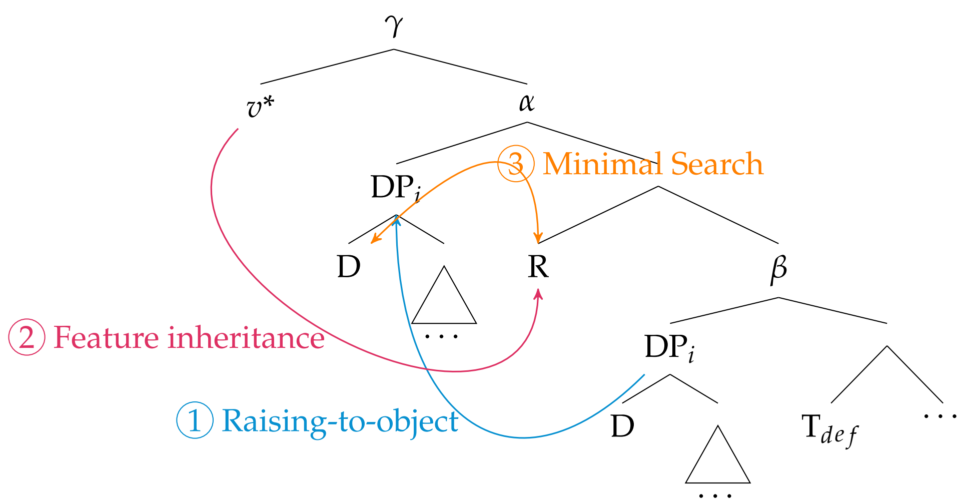
3.8.1 A Theory of Amalgamation

In his work on labeling theory, Chomsky (2015) introduces the possibility of capturing head movement via pair-Merge.

(75) a. They expected John to win.

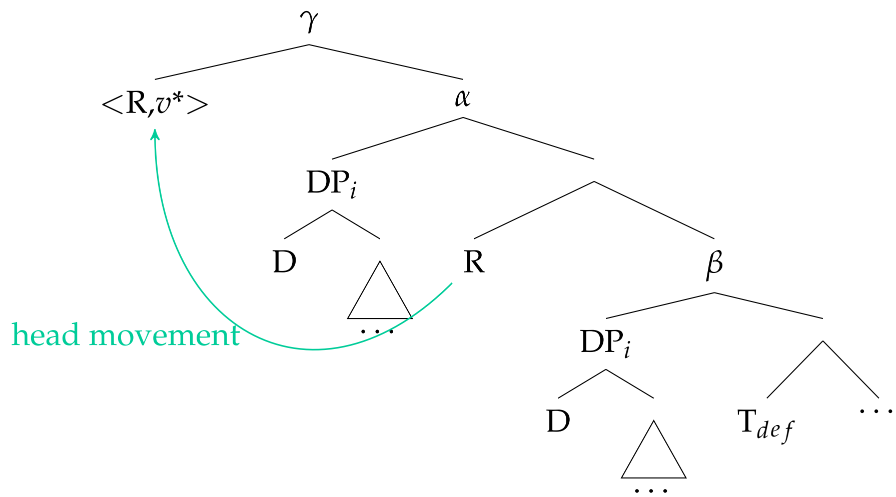
b. {they, $\{_\gamma$ <R,v^*>, $\{_{\alpha=<\phi,\phi>}$ DP$_i$, {R(=expect), $\{_\beta$ DP$_i$(=John)...}}}}}

c.



The steps of the derivation are as follows: (i) after the matrix Root(R) is introduced by EM, DP moves to [spec,RP] ("raising-to-object"), (ii) the phase head $v*$ is introduced, (iii) feature inheritance takes place, thus, the unvalued features are transmitted to R, and (iv) minimal search finds two relevant features in terms of ϕ-features, namely unvalued features on R and valued features on D, therefore α will be labeled as <ϕ, ϕ> by the labeling algorithm. Chomsky (2015) also assumes that phasehood could be transmitted, but $v*$ will lose its phasehood only when head movement via pair-Merge takes place.

(76)



Here Chomsky (2015) assumes that when head movement of R to v^* takes place (assuming that the root has to get its category from a categorizer in the Distributed Morphology framework that Chomsky (2015) adopts only for this aspect (cf. Marantz, 1997)), R becomes the host of the structure in $<$R,$v^*$$>$, rather than v^* being the host. In some languages v^* is realized as a suffix (cf. Embick, 2015), not a main part of the verb. Thus, v^* is adjoined to R, which makes v^* invisible. Then R is activated as a phase head and transfer takes place in the complement of RP (i.e., β). How do we label γ, then? Chomsky (2015) argues that the amalgam $<$R,$v^*$$>$ itself becomes the label, which is interpreted as a verbal element that denotes events.[25]

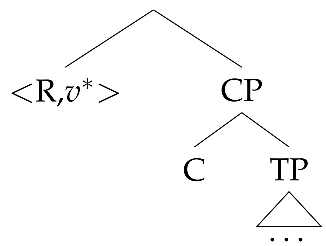
Extending this idea of amalgamation by pair-Merge, Epstein, Kitahara & Seely (2016) proposes that external pair-Merge of heads is possible. The relevant case is a bridge verb case.

[25] In his paper, Chomsky (2015) suggests that the amalgam [R,v^*] becomes a label.

> "...raising or R to v^* yields an amalgam with v^* adjoined to R, and the affix is invisible to the labeling algorithm. Note that although R cannot label, the amalgam [R-v^*] can (Chomsky, 2015, 12)."

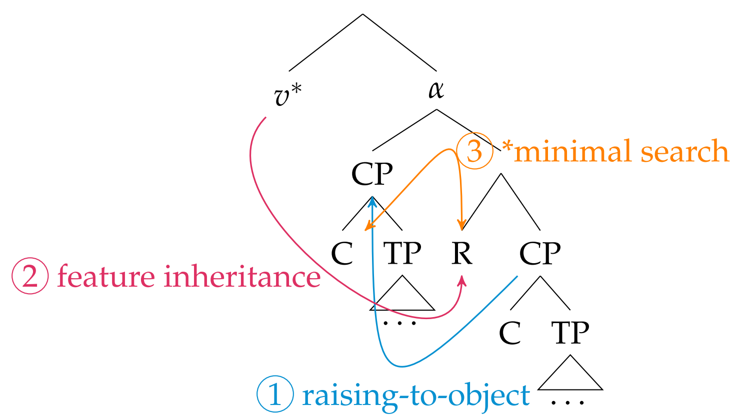
(77) a. John thinks that he will lose this game.

b.



In this derivation, the verb *think* is introduced as an amalgam <R,v^*>, not introduced as separate heads (R and v*). Why do we need to form an amalgam of R and v^*? Suppose we introduce R and v^* separately (78). CP moves up to the [spec,RP], but the CP clause usually lacks ϕ-features, at least in English, assuming with Chomsky (2015) that unvalued ϕ-features on v^* are transmitted to R.[26]

(78)



Since C does not have a relevant valued feature, α will remain unlabeled. Moreover, R still has unvalued ϕ-features.

Considering this, Epstein et al.'s (2016) analysis of external pair-Merge of R and $v*$ explains the bridge verb case. This external pair-Merge approach avoids the unvaluation problem. Notice that since the locus of the phasehood in Chomsky (2015) is unvalued features, the amalgamation inactivates the phasehood. Thus, in the bridge verb cases, the matrix verbal domain loses its phasehood.

[26]See Halpert (2016, 2018) for relevant discussion.

Putting aside the technicality of this analysis, the idea that amalgamation could yield a different construction/verb (a bridge verb) is a novel analysis, which can be extended to other functional categories.

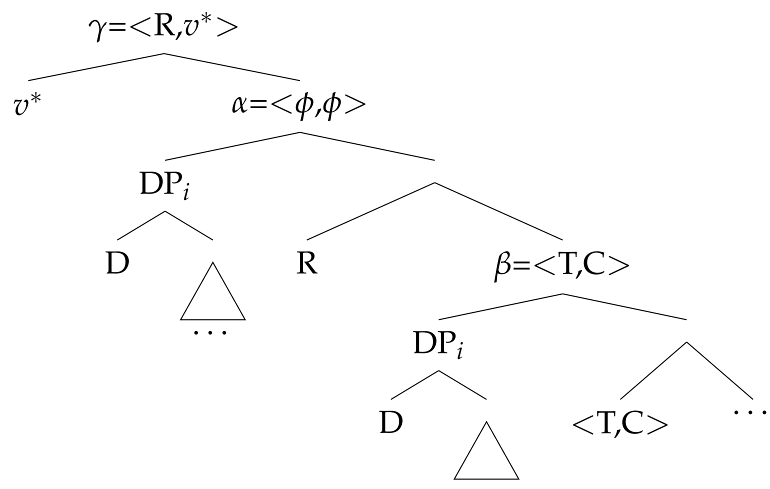
Extending this idea to the infinitival clause, Sugimoto (2021) proposes that what is called a defective T is an amalgam of T and C (i.e., <T,C>).

(79) (= (75))

a. They expected John to win.

b. $\{\text{they}, \{_{\gamma} \langle \text{R},v^* \rangle, \{_{\alpha=\langle \phi,\phi \rangle} \text{DP}_i, \{\text{R(=expect)}, \{_{\beta}\ \text{DP}_i\text{(=John)}\ldots\}\}\}\}\}$

c.



In (79c), the derivational steps are the same as (75) except that the infinitival T is introduced as an amalgam of T and C. Labeling of β was left unsolved in Chomsky (2015), but once we assume that the defective T is an amalgam <T,C> the labeling problem is solved based on the status of the amalgam.

3.8.2 A Proposal

Notice that all cases that apply the operation to create an amalgam in the previous section are construction-specific (applying to an ECM construction, a bridge verb case, and an infinitival clause), and there is no attempt to capture parametric variation by free application of Merge of heads in the literature (except for Epstein et al. 2018b, though they do not explore this possibility in detial). In this section, I propose that this amalgamation approach can capture aspects of parametric variation. The following schematic representations show the multiple possible derivations for a clause structure.

(80) a. $\{_{CP}\ \mathrm{C}\ \{_{TP}\ \mathrm{T}\ \{_{v*P}\ v^* \dots \}\}\}$ — the analytical type

b. $\{_{CP}\ \mathrm{C}\ \{_{<v*,T>}\ <v^*,\mathrm{T}> \dots \}\}$ — the mixed type

c. $\{_{<v*,T,C>}\ <v^*,\mathrm{T},\mathrm{C}> \dots \}$ — the agglutinative type

Supposing that the application of amalgamation is underspecified, just as Merge applies freely, the derivational steps of how to introduce functional categories such as *v**, T, and C are underspecified. For example, they can be introduced to the derivation step by step via Merge like (80a). Another possibility is that two functional heads could amalgamate before entering the narrow syntax, and Merge to the structure as an amalgam (80b). The last pattern that I will explore in this section is the amalgam that consists of three functional heads in (80c). These derivations are possible as long as there is no unvaluation/labeling problem. Thus, each derivation in (80) is an optimal derivation.[27,28]

[27]One might wonder what the possible combinations of the functional categories are. For example, do we have a language where *v** is a free-standing head, while T and C form an amalgam (e.g., <T,C> in a declarative sentence? In (90b), I showed that <T,C> captures the infinitival clause in English, but I am not aware of a language where <T,C> is used as a declarative sentence. This pattern is unattested, but in principle, it is possible. Another type of possible combination would be to amalgamate, for example, *v** and C, while T is a free-standing head (i.e., $\{\{\mathrm{T}, \dots\}, <v^*,\mathrm{C}>\}$). One stipulation might be that there is a certain functional category hierarchy (Ramchand & Svenonius, 2014; Rizzi & Cinque, 2016). I will leave this issue open.

Also, I suspect some possible combinations are banned by other principles, such as the Final-over-Final Condition (FOFC, Sheehan et al. 2017).

i. The Final-over-Final Condition
A head-final phrase αP cannot immediately dominate a head-initial phrase βP, if α and β are members of the same extended projection. (Sheehan et al., 2017, 1,(1))

Assuming that T is head-initial and the amalgam is head-final in the structure $\{\{\mathrm{T}, \dots\}, <v^*,\mathrm{C}>\}$, this is a violation of FOFC. Thus, the word order restriction could rule out some possible combination of functional categories.

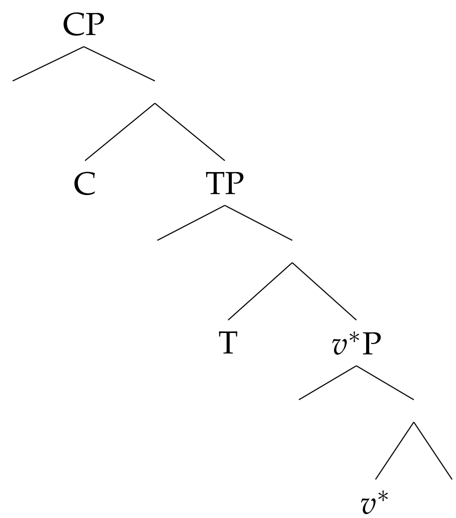
[28]One also might wonder what the connection between head movement and morpho-phonological components is (cf. Julien, 2002). Under the MERGE framework, Chomsky argues that head movement is unformulable by MERGE (i.e., it is basically counter-cyclic movement),

We propose (80a) for English type-languages where each functional category is introduced by set-Merge.

(81) English-type languages

a. $\{_{CP}\ \mathrm{C}\ \{_{TP}\ \mathrm{T}\ \{_{v*P}\ v^{*}\ \ldots\ \}\}\}$

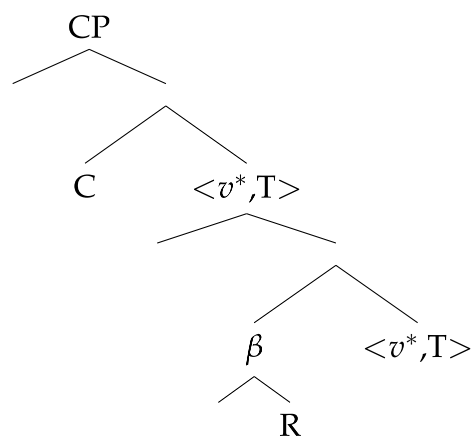
b.



(82) German-type languages

a. $\{_{CP}\ \mathrm{C}\ \{_{<v*,T>}\ <v^{*},\mathrm{T}>\ldots\}\}$

b.



Mixed-type languages include German (80b). The last type is Japanese where the

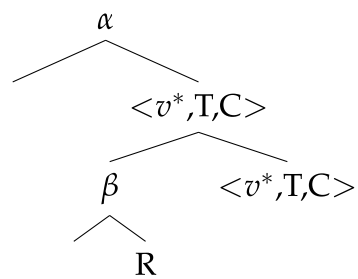
hence, not a syntactic operation. This implies that the amalgamation formed by head movement is not formed by a syntactic operation (i.e., head movement) and this amalgamation process has to do with the PF component (see Chomsky 2021a,b). Notice that our proposal does not involve syntactic head movement through the derivation, but the amalgamation takes place before introducing the amalgamated elements into the derivation in the narrow syntax.

v^*, T, and C are all amalgamated together (80c). The amalgam is formed before entering the narrow syntax.

(83) Japanese-type languages

a. $\{_{<v^*,T,C>} <v^*,\mathrm{T},\mathrm{C}> \ldots \}$

b.



There is a good reason to adopt these structures. As some initial evidence, the following set of examples suggests that the adjacent elements are sensitive to the insertion of adverbials.[29]

(84) a. English

John has (often) embraced Mary

b. German

dass Cindy das Buch gelesen (*wahrscheinlich) hat
that Cindy the book read (*probably) has

'that Cindy (probably) read the book'

c. Japanese

Cindy-wa Mary-ga sono hon-o yon (*tabun) da
Bill-TOP Mary-NOM that book-ACC read (*probably) PAST
(*tabun) to omotteiru.
(probably) C think-PROG

'Cindy thinks that Mary (probably) read the book'

In (84a), the adverbial phrase is inserted between the auxiliary and the past participle. The same applies in (84b), except that the sentence is degraded. In Japanese,

[29]My intention here is not argue for a typological generalization, or to provide a theory of language typology. In the following, I will show that this approach is a better theory of parametric variation than the principles and parameters approaches.

the pattern is the same as in German. The adverbial *tabun* 'probably' cannot intervene between the verb and T head or between T and C.

This set of data can be captured by the proposal that the multiple functional categories could be amalgamated before merging in the narrow syntax.

The structures that we propose for German and Japanese are not a completely new idea. For German, Haider (1988) proposes that there is a matching projection (cf. Bayer & Kornfilt, 1994), and he does not assume TP-projection. This view can be seen in the recent discussion of the German verbal cluster (Keine & Bhatt 2016, but see also Wurmbrand 2007; Salzmann 2013). Fukui (1995b) and Fukui & Sakai (2003) argue that Japanese has a defective T that only functions as a place holder. We reinterpret this as an amalgam $<v^*$,T,C$>$; essentially Japanese only has VP-projection, as Fukui (1995b) suggests.

In what follows, I explore theoretical and empirical consequences of the proposal. The theoretical consequences involve labeling theory and phase theory. Since the theory of amalgamation provides an amalgam as a label, this could provide a solution to the problem in Chomsky (2013) regarding the labeling issue with the subject-predicate structure (§3.8.3.1). Another theoretical consequence concerns the phasal edge condition (Gallego & Uriagereka, 2007; Chomsky, 2008). I will argue that the edge condition applies only when the phase head is introduced as a free-standing head, whereas the amalgam is de-phased (cf. Chomsky, 2015; Epstein et al., 2016), so there is no edge condition there (§3.8.3.2). As for empirical consequences, I will show that the theory of amalgamation derives a cluster effect in §3.8.4. §3.8.5 has to do with intra-linguistic variation within a particular I-language regarding the different applications of Merge to heads.

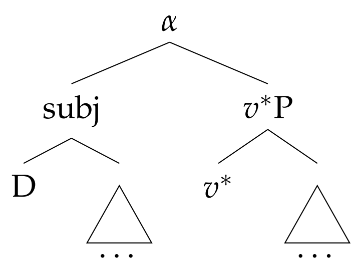
3.8.3 Theoretical Consequences

3.8.3.1 Labeling Theory

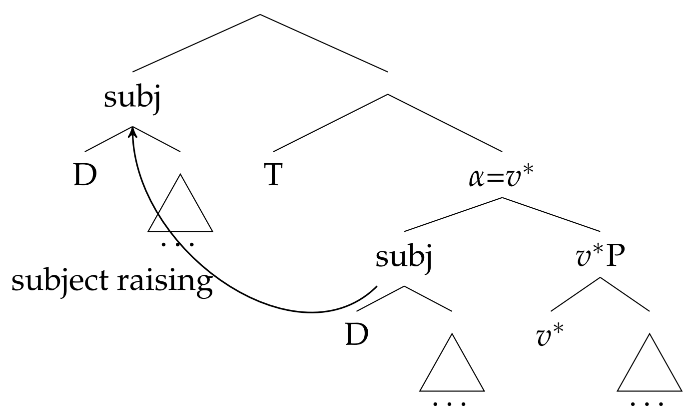
One consequence of this approach involves labeling theory (Chomsky, 2013, 2015). As we discussed, leaving the structure unlabeled (e.g., {XP,YP}) will result in the interfaces not being able to interpret it. The relevant structure is the subject-predicate structure.

(85) a. $\{_\alpha$ subj, v^*P$\}$

b.

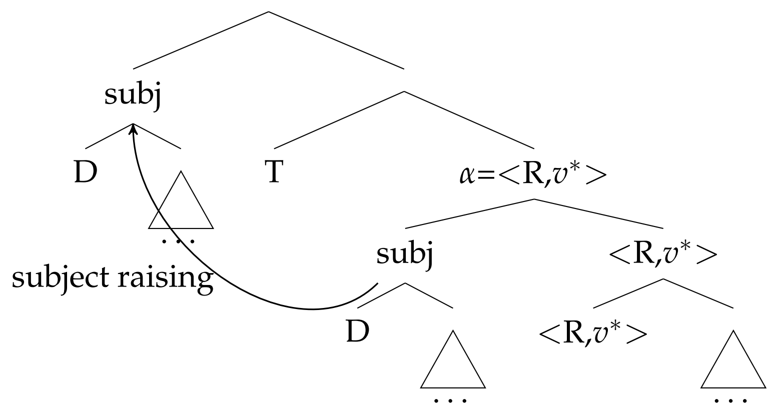


c.



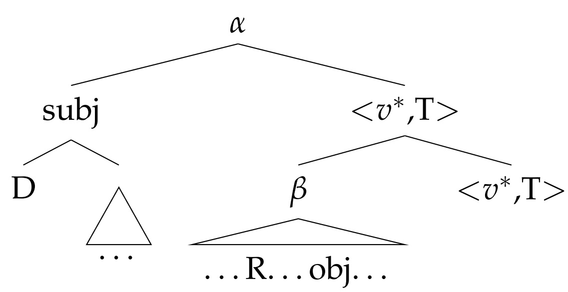
In English, the subject will raise to [spec,TP], which results in the subject in [spec,v^*P] being invisible. Thus, the label of α becomes v^*. Also notice that in the $v*$P phase, R raises to $v*$ and forms the amalgam <R,v^*> in Chomsky (2015). Hence the labeling of α becomes <R,v^*>.

(86)

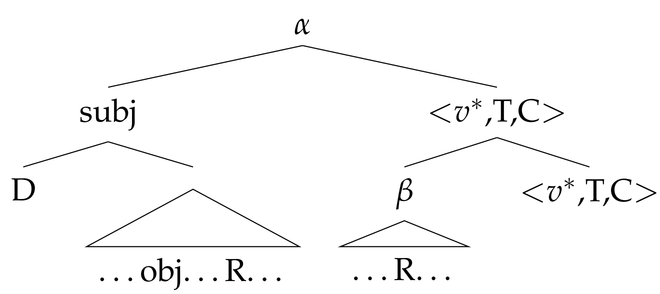


However, this movement analysis does not necessarily apply to German and Japanese. In both types of languages, the subject remains $in-situ$. In (87a), the subject is externally merged to [spec, $<v^*$,T$>$]. In (87b), the subject is externally merged to [spec,$<v^*$,T,C$>$].

(87) a. German



b. Japanese



The particularly important part is that the candidates for labeling are limited. For example, T and R are too weak to serve as a label, but the shared features can. Moreover, the amalgam becomes the label, and seems to be the only unit that can be a label by itself.

Chomsky (2013) also pointed out this labeling problem. He adopts Alexiadou,

& Anagnostopoulou's (2001) analysis where either the external or internal argument has to move out in the structure in (88a).

(88) a. $\{_\alpha$ Subj, $\{v^*$, {V, IA}}}

b. {IA$_i$, {. . . , $\{_\alpha$ Subj $\{v^*$, {V,IA$_i$}}}}}

After IA moves out of the v^*P domain, the labeling algorithm finds Subj and v^* after transferring the complement of the phase head. The label α becomes v^*. However, this analysis does not apply to Japanese, since both EA and IA stay *in situ*. Thus, the labeling problem for the structure α in (88a) remains.

Instead of assuming the analysis above, we propose here that the labeling algorithm prefers the "richer" amalgam. The amalgam is somehow a salient element for the labeling algorithm.

(89) When the Labeling Algorithm finds rich amalgams and a free-standing head in a structure such as a {XP,YP} structure where X is an amalgam and Y is a free-standing head, the amalgam becomes the label, namely X in {XP,YP}.[30]

In the structures (87a) and (87b), the labeling algorithm/minimal search finds two heads, namely D in the subject and the amalgam <v^*,T>/<v^*,T,C>. Then the label α becomes the amalgam, <v^*,T> and <v^*,T,C> respectively, by (89). This also suggests that in English, there is no labeling problem in a {subj,v^*P} structure (Chomsky, 2015). Again the relevant example is (75). (90c) shows that the subject is externally merged to the v^*P structure, followed by the internal pair-Merge of R to v^*.

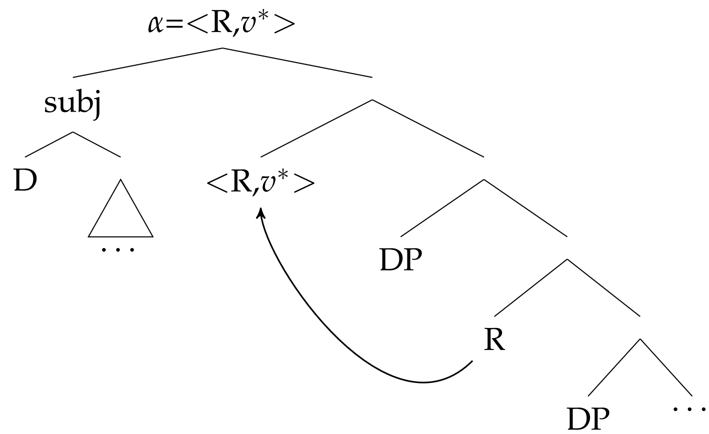
[30] I will take this as a working hypothesis.

(90) (= (75))

a. They expected John to win.

b. $\{\text{they}, \{_{\gamma} <\text{R},v^*>, \{_{\alpha=<\phi,\phi>} \text{DP}_i, \{\text{R(=expect)}, \{_{\beta} \text{DP}_i\text{(=John)}\ldots\}\}\}\}\}$

c.



This does not mean that the subject does not have to move. If it does not move to [spec,TP] for further derivation, the TP structure will not be labeled since subject raising to [spec,TP] needs to be done to label the TP structure via ϕ-related valuation (Chomsky, 2013, 2015). Thus, in English-type languages, the subject raising is still needed.

This amalgamation approach provides an alternative approach to Japanese in terms of labeling. As we discussed in chapter 2, Saito (2016) and Saito's (2016; 2018) labeling analysis has conceptual and empirical problems. Saito proposes that Case markers and inflectional elements in Japanese have λ-features.[31] First, if labels are for interface interpretation (Full Interpretation), why do case-markers play a role at the interfaces? Second, Saito's analysis does not capture the (un)grammaticality of improper scrambling and embedded CP scrambling (see §2.4).

Considering these issues in Saito's framework, I propose here that a labeling problem does not arise once we adopt (89). This approach also applies to English, and we do not need an extra mechanism to solve the labeling problem in Japanese, or in German where the subject stays *in situ*.

[31] See Narita & Fukui (2022) for criticism of Saito's (2016) labeling analysis.

(89) gives us an interesting consequence. The generalization is as follows.

(91) A generalization

a. If a functional category is free-standing (i.e., introduced in the narrow syntax by set-Merge), the specifier is available and the labeling algorithm forces valuation (spec-head relation), or the element in the specifier has to move out of it.

b. If multiple functional categories are amalgamated before being introduced in narrow syntax, the specifier is not available for valuation (i.e., for spec-head agreement). Due to (89), there would be no labeling problem.

This generalization is derived simply from the structures that we propose for English, German and Japanese. For example, in English, every functional category is introduced to the derivation as a free-standing head that can be in a relation with the specifier. This derives subject movement when the head is T, for example. If two heads are amalgamated, for example v^* and T, there is no specifier of v^*P/TP, therefore subject raising does not take place (see §3.8.4 for empirical discussions).

This generalization can be tested by observing scrambling phenomena. In English, scrambling is in general impossible.

(92) English

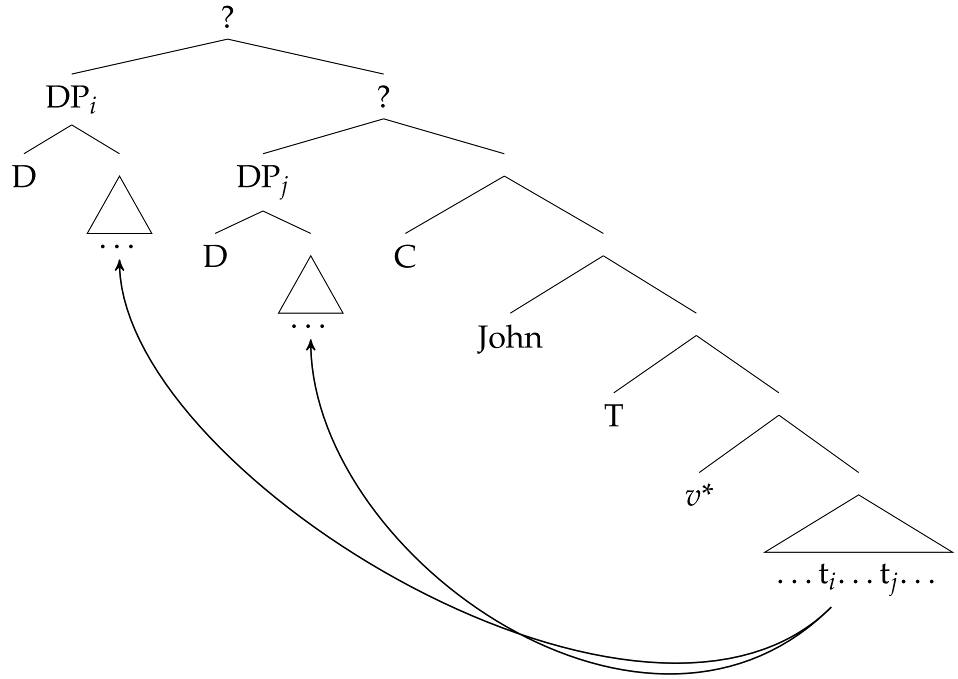
a. John put that book on the table

b. *On the table$_i$, that book$_j$, John put t$_i$, t$_j$. (Fukui, 1988, 257, (14))

c. *On the table$_i$, that book$_j$, Bill thinks that John put t$_i$ t$_j$

In English, the{XP, YP} problem cannot be solved in (92); in (92b), multiple DPs cannot be scrambled without having extra functional heads for each DP in feature sharing strategy (each category will end up moving to, e.g., a focus position, which is not, by definition, scrambling). The schematic structure for (92b) below illustrates the point.

(93)



Furthermore, long-distance scrambling in English is also not allowed in (92c).

On the other hand, German partially allows scrambling (short/clause-bounded scrambling), as shown in (95).

(95) German

a. dass das Objekt dem Subjekt den ersten Platz streitig macht
That the object the subject the initial place contested makes
'That the object competes with the subject for the initial place'

b. dass dem Subjekt den ersten Platz das Objekt streitig macht
That the subject the initial place the object contested makes

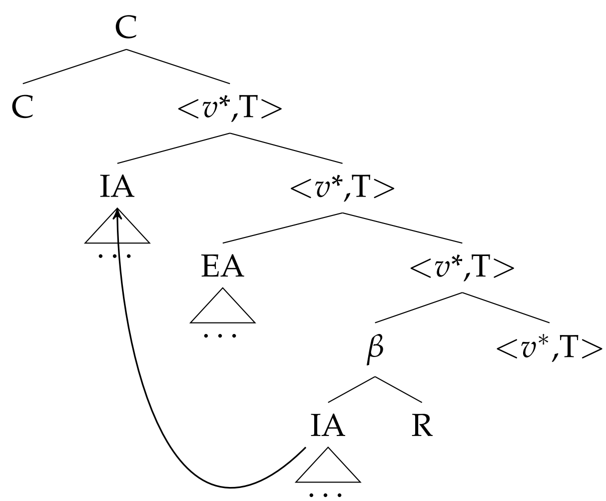
(Haider, 2006, 208,(6))

c. *dass dieses Buch$_i$ Hans dem Studenten gesagt hat [$_{CP}$ dass Maria t$_i$ besitzt]
that this book Hans the student told has [$_{CP}$ that Maria t$_i$ owns]
'Hans told the student that Mary owns this book'

(Grewendorf & Sabel, 1999, 10-11,(21))

In (95a) and (95b), the subject and the object position are switched, hence scrambling takes place, which suggests that clause-bound scrambling is possible by virtue of the amalgam, (i.e., $<v^*,\mathrm{T}>$) (95a).

(96) The structure for (95a)



But once scrambling goes beyond C, at the CP level, a labeling problem arises. That is, long-distance scrambling is also impossible (95c). On the other hand, Japanese does not have such a restriction. Scrambling multiple elements out of the embedded clause is possible due to the amalgam $<v^*,\mathrm{T,C}>$ (97b).[32]

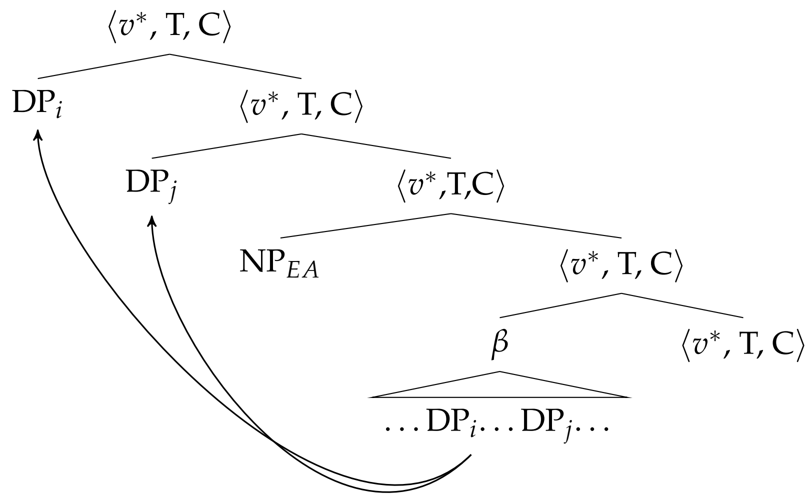
(97) Japanese

a. {Taroo-ga, Hanako-ni, sono hon-o} age-ta
{Taroo-NOM, Hanako-DAT, sono hon-ACC} give-PAST
'Taro gave that book to Hanako'

b. {Hanakoi-ni, sono hon$_j$-o} Jiro-wa [Taroo-ga t_i t_j age-ta
{Hanako$_i$-DAT, that book$_j$-ACC} Jiro-TOP [Taroo-NOM t_i t_j give-PAST
to] omot-te-iru.
C] think-te-PROG
'Jiro thinks that Taroo gave that book to Hanako'

[32] As for the multiple long-distance scrambling, see Koizumi (2000); Fukui & Sakai (2003); Agbayani et al. (2015) for relevant discussion.

(98)



(98) shows the schema of the tree representation for Japanese scrambling. Whenever the DP is scrambled, the amalgam becomes the label due to (89), thus, multiple scrambling is possible. The same applies to the multiple long-distance scrambling.

These patterns among English, German, and Japanese suggest that the proposed labeling analysis provides an account for the (un)availability of scrambling.

The next subsection explores other theoretical consequences of the theory of amalgamation, regarding phase theory.

3.8.3.2 Phase Theory

Another consequence of the theory of amalgamation involves phase theory. We assume the following:

(99) Phase Theory (cf. Chomsky, 2015)

a. C and $v*$ are phase heads, which have unvalued features

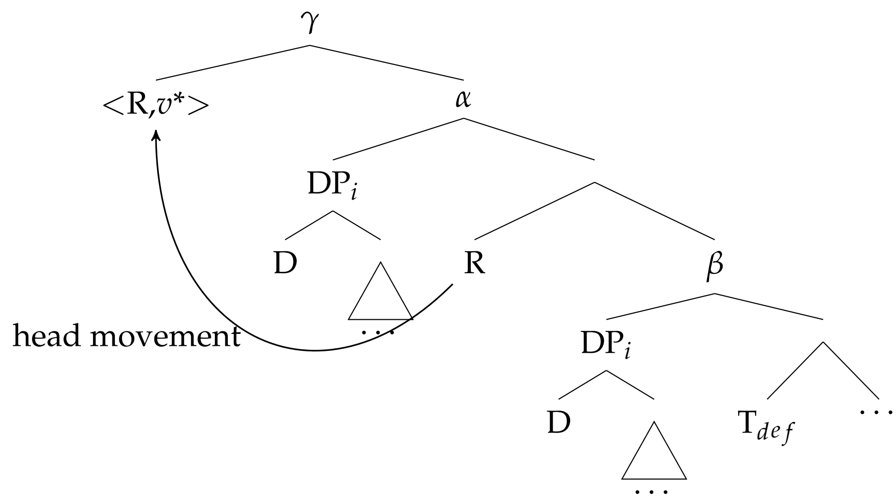
b. Phase heads have phasehood, which defines the Spell-out domain (i.e., the complement of the phase head)

c. Phasehood can be deactivated by amalgamation (e.g., head movement)

In Chomsky (2015), phase theory is connected to unvalued features (99a). Once the phase heads are introduced to the derivation, they define the domain of the phase, allowing the derivation to be carried out locally and cyclically (99b). The last point (99c) needs to be further discussed, following Chomsky (2015). As we discussed in §3.8.1, in the *v**P domain derivation in Chomsky (2015), head movement takes place to form an amalgam (see (75) and (76)).

(100) a. They expected John to win

b.



After feature inheritance from *v** to R, α is labeled by <ϕ,ϕ> and head movement takes place. Chomsky (2015) assumes that head movement makes the phase head *v** invisible, then R in the lower position is activated as a phase head. Now, the transfer domain becomes the complement of the derived phase head (R), namely β in (100b).

The main idea is that when the phase head becomes invisible because of the amalgamation, the phase status of the phase head is deactivated.

The immediate consequence is the absence of phasehood of v^* and C in Japanese. In the theory of amalgamation with phase theory, Japanese does not have phasehood because of the amalgam <v^*,T,C> (99c); it has been argued that syntactic island effects are, in general, weak in Japanese (cf. Fukui & Speas 1986; Fukui

1991; Ishii 1997 among others). Due to the lack of phasehood resulting from the amalgamation, Spell-out should be available anytime in Japanese.[33]

Based on this analysis, we assume here that Spell-out/Transfer is freely available in Japanese because (i) there are no strong phases in Japanese since the PHs v^* and C are in the amalgam <v^*,T,C> and (ii) uninterpretable/unvalued features on the amalgam (<v^*,T,C>) are not available. On the other hand, English shows at least two strong phases, v^* and C. In English-type languages, these phase heads are free-standing, therefore, they inherently have phasehood. As for German-type languages, the prediction is that C has strong phasehood, whereas v^* does not.

One piece of empirical evidence for the presence/absence of phasehood of C phases comes from Chomsky's (2008) phase edge condition (cf. Gallego & Uriagereka 2007).

(101) Extraction from phasal edge

a. *YP$_i$... [$_{CP}$ [$_{XP}$ ... t$_i$] C [$_{TP}$ ...]], where XP is on a phasal edge

b. Edge condition: Syntactic objects at phase edges are internally frozen

(Gallego & Uriagereka, 2007, 19)

The examples that follow show sub-extraction from the moved element that is located in the embedded CP.

In English, the sentences become degraded.

(102) English

?? Who$_i$ do you wonder [$_{CP}$ [which picture of t$_i$] Mary bought t$_i$]?

(Lasnik & Saito, 1992, 102)

German also shows that the sentence is degraded (103b).

[33] Fukui & Kasai (2004) argue that Japanese can spell out the verb arguments since Japanese does not have uninterpretable features (e.g., phi-features). See Fukui & Kasai (2004) for details.

(103) German (cf. Corver 2017; Müller 1998, 2010)

a. Ich denke [$_{CP}$ [$_{VP}$ das Buch gelesen]i [$_{C}$ hat$_k$ [keiner t$_i$ t$_k$]]]
I think [$_{CP}$ [$_{VP}$ that book read]$_i$ [$_{C}$ has$_k$ [no one t$_i$ t$_k$]]]
'I think no one read the book'

b. *Was$_j$ denkst du [$_{CP}$ [$_{VP}$ t$_j$ gelesen]$_i$ [$_{C}$ hat$_k$ [keiner t$_i$ t$_k$]]]?
What$_j$ think you [$_{CP}$ [$_{VP}$ t$_j$ read]$_i$ [$_{C'}$ has$_k$ [no one t$_i$ t$_k$]]]?
'What do you think no one read'

c. Was$_j$ denkst du [$_{CP}$ t$_j$ [$_{C'}$ hat$_k$ [keiner [$_{VP}$ t$_i$ gelesen] t$_k$]]]?
What$_j$ think you [$_{CP}$ t$_j$ [$_{C}$ has$_k$ [no one [$_{VP}$ t$_i$ read] t$_k$]]]?
'What do you think no one read'

Descriptively speaking, when the VP is fronted to the embedded [spec,CP] subextraction cannot apply from there (103b), whereas subextraction can take place from the *in-situ* VP (103c).

In Japanese, the most embedded CP is scrambled to the higher embedded [spec, CP] (104b) and scrambling of NP 'that book-ACC' to the edge of the matrix CP (104c) is acceptable.

(104) Japanese

a. [$_{IP}$ John-ga [$_{CP}$ [$_{IP}$ Bill-ga [$_{CP}$ Mary-ga sono hon-o katta
[$_{IP}$ John-NOM [$_{CP}$ [$_{IP}$ Bill-NOM [$_{CP}$ Mary-NOM that book-ACC bought
to] itta] to] omotteiru]
that] said] that] think]
'John thinks that Bill said that Mary bought that book'

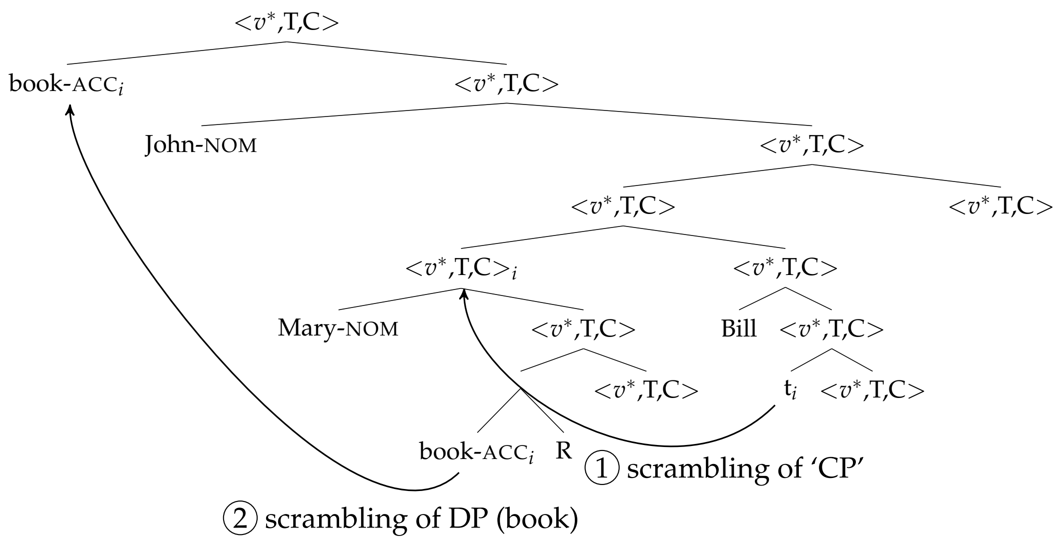
b. [$_{IP}$ John-ga [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-ga sono hon-o katta to]$_i$ [$_{IP}$
[$_{IP}$ John-NOM [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-NOM that book-ACC bought that]$_i$ [$_{IP}$
Bill-ga t$_i$ itta] to] omotteiru]
Bill-NOM t$_i$ said] that] think]

c. [$_{IP}$ sono hon-o$_j$ [John-ga [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-ga t$_j$ katta to]$_i$
[$_{IP}$ sono hon-ACC [John-NOM [$_{CP}$ [$_{IP}$ [$_{CP}$ Mary-NOM t$_j$ bought that]$_i$
[$_{IP}$ Bill-ga t$_i$ itta] to] omotteiru]
[$_{IP}$ Bill-NOM t$_i$ said] that] think]

(Saito & Fukui, 1998, 465-466, (65))

In our terms, CP is the amalgam <v*,T,C>. Since C is in the amalgam, the phase status is deactivated.

(105)



Since the edge condition (101b) only applies when the element is on the phasal edge, a scrambled CP in another embedded CP that is an amalgam is not regarded as a phasal edge. This predicts that extraction from the scrambled CP is fine and this prediction is borne out in (104c).

This set of examples suggests that English and German have strong phasehood, hence the edge condition applies, whereas Japanese does not.

In the previous subsection and this subsection, the theoretical consequences of the theory of amalgamation have been explored. In the previous subsection, the proposal provided a solution to the issue of the labeling of the subject-predicate structure in German and Japanese. In these languages, the movement strategy in labeling theory does not apply. It was proposed that the amalgam is preferred in order to label the whole structure. This does not add any extra mechanisms or assumptions to capture the labeling. Another consequence has to do with phase theory. In Chomsky (2015), phasehood is deactivated when the phase head becomes invisible (via head movement in the *v**P domain). The amalgamation process also deactivates phasehood. This correctly predicts the absence/presence of the edge condition effect in English, German, and Japanese. In English and German, the edge condition is observed regarding CP.[34] In our analysis, this is because the status of the phase head C is activated, since in English and German, C is introduced as a free-standing head, whereas in Japanese, the amalgam $<v^*$,T,C$>$ demotes the phase head C, hence the phase status is eliminated, which results in the absence of the edge condition effect.

3.8.4 Empirical Consequences

In §3.8.3.1, we discussed the following generalization, which holds for English, German, and Japanese.

(106) A generalization (=91)

a. If a functional category is free-standing (i.e., introduced in the narrow syntax by set-Merge), the specifier is available and the labeling algorithm forces valuation (spec-head relation), or the element in the specifier has to move out of it.

b. If multiple functional categories are amalgamated before merging in narrow syntax, the specifier is not available for valuation (i.e., for spec-head agreement).

In addition to this, there are more consequences that fall out from the amalgamation approach.

(107) a. If T is free-standing, VP-fronting/VP-ellipsis is licensed.

b. If T is free-standing, [spec,TP] is available; an expletive is inserted.

[34] Regarding the v^* phase edge, there is some difficulty in testing our hypothesis. The relevant discussion would involve the subject island: Subextraction from [spec, v^*P] in English (cf. Chomsky, 2008) and German (Fanselow & Ćavar 2002; cf. Ott 2011) is impossible, whereas subextraction from [spec, v^*P] is possible in Japanese (i.e., there is no subject island effect). Our proposal predicts that German consists of weak phases since it has [spec, $<v^*$,T$>$], but not [spec, v^*P]. However, since the connection between the phasal edge of v^* and the subject island is not clear (see §2.8.1), we cannot draw a conclusion at this point. I leave this issue for further research.

c. If v^* and T are amalgamated, [spec,TP] is not available for spec-head agreement.

d. If C is free-standing, *wh*-movement is obligatory.

e. If C is amalgamated with $v*$ and T, the *wh*-phrase will be *in situ*.

In what follows, we will confirm additional consequences for English, German, and Japanese. The summary is in the table below.[35]

	English	German	Japanese
Subject-verb agreement	yes	yes	no
Expletive	yes	no	no
VP-fronting	yes	$<v*$,T$>$P-fronting	no
VP-ellipsis	yes	no	no
wh-movement	yes	yes	no

Table 3.3: Properties captured by the amalgamation approach

3.8.4.1 Subject-Verb Agreement

Assuming that unvalued features on C-T have to be valued via Agree/minimal search, there are multiple outputs that the amalgamation approach can predict. The schema for each structure is below.

(108) a. $\{_{CP}$ C $\{_{TP}$ subj T $\{_{v*P}$ v^* ... $\}\}\}$ *English-type languages*

b. $\{_{CP}$ C $\{_{<v*,T>}$ subj ... $<v^*$,T$>$ $\}\}$ *German-type languages*

c. $\{_{<v*,T,C>}$ subj $<v^*$,T,C$>$ $\}$ *Japanese-type languages*

Under the feature inheritance approach (Chomsky, 2007, 2008; Richards, 2007), C, a phase head, has unvalued features that will be transmitted to T in the derivation.

[35] A cluster effect is indirectly derived from the amalgamation approach. Once the amalgamation takes place, the specifiers of those functional heads are not available. However, the specifiers of the amalgamated heads are available, even if it does not agree with the functional heads within the amalgam, since the latter become invisible by pair-Merge, by definition (see Chomsky 2015 and Epstein et al. 2016). Thus, the generalization (107) derives the cluster effect. But in other languages, there might be other ways of deriving another set of phenomena in a systematic way based on this amalgamation approach.

In (108a), after the unvalued features are inherited by T (a free-standing head), the subject and T agree and value the unvalued features on T.[36] In case of (108b), T is not available since it is amalgamated. The unvalued features on a free-standing C stay on C and minimal search finds C and the subject to value the unvalued features on C via the probe-goal ϕ-agreement configuration. Thus, in this case, subject raising is not forced. In Japanese, on the other hand, the relevant functional heads are unavailable since T and C are amalgamated.

The following table shows that the predictions above are borne out.

		English	German	Japanese
SG	1	I run.	Ich renne.	Watashi-ga hashi-ru.
	2	You run.	Du rennst.	Anata-ga hashi-ru.
	3	He/She/It runs.	Er/Sie/Es rennt.	Kare-ga/kanojyo-ga hashi-ru.
PL	1	We run.	Wir rennen.	Watashi tachi-ga hashi-ru.
	2	You run.	Ihr rennt.	Anata tachi-ga hashi-ru.
	3	They run.	Sie rennen.	Karera-ga/kanojyora-ga hasi-ru.

Table 3.4: Subject-verb agreement in English, German and Japanese

When C-T and the subject are in a ϕ-relation, the subject-verb agreement phenomenon takes place, whereas when C-T is amalgamated with v^*, like in Japanese-type languages, the inflection of the verb does not change regardless of the person- and number-features on the pronouns.

3.8.4.2 Expletive

Another consequence of the amalgamation approach is the presence/absence of the [spec,TP] position. Traditionally, it has been argued that the subject is necessary for grammaticality (the extended projection principle (EPP), cf. Chomsky 1982), and it occupies [spec,TP]. In English, the expletive functions as a place holder for the subject and in (109a), it is obligatory.

[36] For the root context in German, I assume that the amalgam <v^*,T> moves to C (cf. Obata, 2010; Legate, 2011; Goto, 2011a).

(109) a. English

because there is a man in the garden

b. German

weil (*es) getanzt wird
because (*it) danced was
'because (*it) danced was'

In German and Japanese, on the other hand, there is no obligatory expletive insertion to the [spec,TP] position, since T is amalgamated in German and Japanese and there is no TP-projection in our terms.

3.8.4.3 VP-fronting

VP-fronting in English is allowed; *v* is a free-standing head in our terms.

(110) English

The children said that they would cut the grass . . .

and [$_{CP}$ [$_{VP}$ cut the grass]$_i$ they did t$_i$]] (Travis & Massam, 2021, 37,(5))

In Japanese, since we argue that *v* is amalgamated with T and C, there is no way of moving the VP projection, since such a projection is not generable. The example of VP-fronting is below.

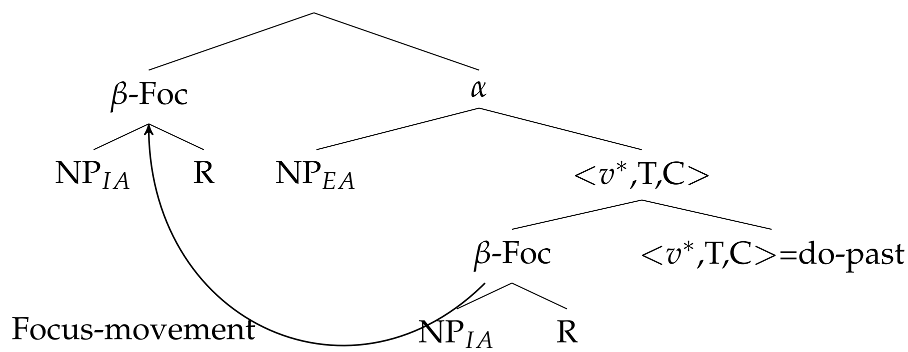
(111) *[$_{VP}$ Ringo-o tabe] Taroo-ga t$_{VP}$ (si)-ta.
[$_{VP}$ apple-ACC eat] Tarro-NOM t$_{VP}$ (do-)PAST
'Eat apple, Taroo did'

Descriptively speaking, once the verb and the tense morpheme are unattached, VP-fronting is not licensed. However, once a morpheme such as -sae/mo/wa/dake '-even/also/TOP/only' is attached to the verb and *su*-insertion takes place, the fronting example becomes acceptable (cf. Funakoshi, 2020).

(112) [$_{VP}$ Ringo-o tabe-sae/mo/wa/dake] Taroo-ga t$_{VP}$ si-ta.
[$_{VP}$ apple-ACC eat-even/also/TOP/only] Tarro-NOM t$_{VP}$ do-PAST
'Taroo ate apples'

We reinterpret this as follows: when a focus particle is attached to a Root phrase, the focus particle behaves as a head and fronting become possible.

(113)



We also assume that v^* has to be realized as the dummy verb *su-* 'do' when this focus movement happens (*-si* in 112). What is called VP-fronting is possible with this qualification. In our terms, it is Foc-P fronting.

In German, since v^* is amalgamated with T, we predict that VP-fronting is not available. However, German apparently does have VP-fronting as shown below.

(114) [$_{\alpha}$ [$_{Subj}$ Ein junger Hund] einen Briefträger gebissen] hat hier schon oft.
[$_{\alpha}$ [$_{Subj}$ a-NOM young dog] a-ACC mailman bitten] has here already often
'It has happened often here already that a young dog has bitten a mailman'

Since the subject is not forced to move to [spec,TP] in German (Wurmbrand, 2006), the subject remains *in situ*. We interpret this as α-fronting being $<v^*$,T$>$P-fronting.

3.8.4.4 VP-ellipsis

Sag (1976) suggests that the condition on licensing VP-ellipsis depends on the presence of the auxiliary located immediately before VP.

(115) a. Dennis rarely plays the piano, but Susan often *(does) __.
(Lobeck, 1990, 352, (12b))

b. John is sleeping, and Bill is __, too. (van Craenebroeck, 2017, (1))

We interpret this condition as follows: VP-ellipsis is licensed by a free-standing T. It also has been argued in the literature that a spec-head relation is a necessary condition. From the amalgamation approach, we predict that Japanese and German do not have VP-ellipsis, since the T is amalgamated. If there is no free-standing T, [spec,TP] is not available.

Furthermore, Lobeck (1990) and Saito & Murasugi (1990) argue that the VP is elided only when a T head has a specifier which it agrees with.

(116) a. *Mary doesn't smoke because [PRO to __] is dangerous.
(Lobeck, 1990, 353,(13b))

b. It is possible that Mary smokes, but it's certain that John does __.
(Lobeck, 1990, 353,(14a))

In (116a), the elided VP has an infinitival T head, which does not have a specifier which it agrees with, whereas there is a specifier of TP in (116b). Thus, there are two conditions for VP-ellipsis: (i) a free-standing T and (ii) the presence of the specifier that agrees with the subject.

In German, VP-ellipsis is not allowed, as shown in (117), as we predict.[37]

(117) German

*Leyla WOLLte die Hausaufgaben nicht machen, doch Franz meinte, dass
Leyla wanted the homework not make but Franz meant that
sie __ HAT.
she __ has

'Leyla didn't want to do the homework but Franz said that she has (done it).' (Repp & Struckmeier, 2020, 187)

[37]Capitals indicate stress on the modal and the auxiliary verb respectively.

In Japanese, recent literature suggests that the following example involves argument ellipsis, not VP-ellipsis.

(118) Japanese

Hanako-wa gakkoo-ni it-ta kedo, Taroo-wa __ ik-anak-atta.
Hanako-TOP school-to go-PAST but Taroo-TOP __ go-NEG-PAST

(intended) 'Hanako went to the school, but Taroo didn't go to the school.'

We take this as a consequence of the absence of a free-standing *v* and T. Since there is no free-standing T, the absence of T also indicates there is no specifier of TP. Thus, the conditions on VP-ellipsis cannot be satisfied.[38]

3.8.4.5 *Wh*-movement

The last set of examples is related to the CP domain. Our prediction is that when C is a free-standing head, the specifier is available, whereas when C is amalgamated with other functional categories, there is no specifier, which results in no obligatory A'-type movement.

In English and German, *wh*-movement is obligatory.

[38]In the previous studies on Japanese ellipsis, there is literature arguing for VP-ellipsis (where the verb is supposed to be raised to T before VP-ellipsis) (Otani & Whitman, 1991). See also Oku (1998); Takahashi (2008); Sakamoto (2017) among many others. Oku (1998) argues for argument ellipsis based on the following set of examples.

i. Bill-wa kuruma-o teineini arat-ta
Bill-TOP car-ACC carefully wash-PAST
'Bill washed the car carefully'

ii. John-wa __ arawa-nakat-ta
John-TOP __ wash-NEG-PAST
(lit)'John didn't not wash __' (Oku, 1998, 172)

The interpretation of (ii) is that John didn't wash the car at all and it does not mean that John did actually wash the car, but he didn't wash it carefully. This is different from the English VP-ellipsis example such as 'Bill washed the care carefully, but John didn't __'. In this example, the prominent interpretation of 'John didn't' is that John didn't wash the car carefully, which includes the interpretation of the adverbial. This is easily explained by the VP ellipsis analysis, since the VP including the adverbial is elided. If Japanese example is also VP-ellipsis, then it does not explain why the same reading cannot obtain in the English example.

(119) a. English

I don't know what$_i$ John bought t_i.

b. German

Ich	fragte	mich	wen$_i$	Hans	t_i	sah.
I	asked	REFL	who-ACC	Hans		saw

'I wondered who Hans saw.' (Sabel, 2000, 413,(12-b))

In contrast, Japanese does not have a requirement of *wh*-movement.

(120) Japanese

Boku-ga	John-ga	nani-o	katta	ka	siranai	(koto).
I-NOM	John-NOM	what-ACC	bought	Q	know-NEG-PRES	(fact)

'(the fact that) I don't know what John bought.' (Fukui, 1988, 256, (12))

In (121), the *wh*-phrase can stay in the object position in the embedded clause. We predict this result from the amalgamation approach, because there is no CP-projection.

Notice that scrambling of the *wh*-phrase to the front is also possible.

(121) Japanese

nani-o$_i$	Boku-ga	John-ga	t$_i$	katta	ka	siranai	(koto).
what-ACC	I-NOM	John-NOM	t	bought	Q	know-NEG-PRES	(fact)

'(the fact that) I don't know what John bought.'

There is no specifier dedicated to C since C is included in the amalgam <*v**,T,C>. The specifier of <*v**,T,C> is available, but it does not establish a spec-head relation with the *wh*-phrase since C is hidden in the amalgam. It is known that this type of scrambling will reconstruct to the base position (*aka* radical reconstruction (Saito 1989, 1992, 2003)).

In this subsection, I showed how to derive a cluster effect among English, German, and Japanese without assuming macro-parameters. I showed relevant data to test our prediction in (107), which suggests that underspecification of rule application exists with respect to merging heads (See also Table 3.3).

In the next section, I will show that Basque shows another kind of relevant I-language variation.

3.8.5 Intra-variation and Underspecification

In Obata & Epstein (2016) and Epstein et al. (2018b), it is argued that there are two types of variations: inter-variation and intra-variation. Inter-variation, the linguistic variation among languages, is what we observed with English, German, and Japanese in the previous subsections. On the other hand, intra-variation is when a given language system allows variation within its system. This aspect cannot be captured by the notion of parameters, since once parameters are fixed, it's not possible to switch them. In our underspecification approach, it is possible to capture intra variation since there is no parameter in the linguistic system. Rather, multiple optimal derivations are allowed in the linguistic system of a single I-language as long as the derivations converge. In what follows, I will show that this is the case in Basque.

In Bach (1971), there is an important generalization:

(122) a. Movement of question words will always be to the head (left) of the sentence.

b. Question Movement will occur only in SVO or VSO languages, never in (deep) SOV languages.

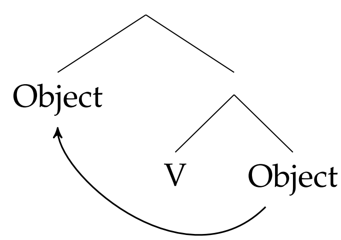
c. If a language marks the themes of a sentence, question words will never occur as themes.

(Bach, 1971, 164)

The relevant point in the discussion here is that (deep) or consistently head-final language tend to show the *wh-in-situ* property, namely, *wh*-movement is not obligatory in these languages.

Ormazabal et al. (1994) try to capture this by adopting Kayne's (1994) framework. In this framework, the universal word order becomes SVO based on asymmetrical c-command. In order to derive SOV languages, this theory assumes that the object moves up high enough to asymmetrically c-command V.

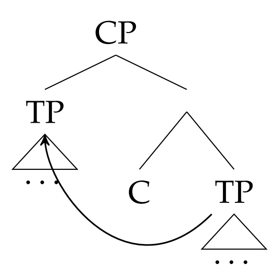
(123) a.



b. word order: Object < V ("<" means precedence)

For C to be in head-final position, TP moves up to the specifier of CP.

(124) a.



b. word order: TP < C

After TP raises to [spec,CP], TP asymmetrically c-commands C and the head-final property is captured. However, Ormazabal et al. (1994) show that Basque has a counterexample to this derivation.

(125) a. ... Mirenek Joni liburua irakurri dio-la ... [S-IO-O-V+I+C]

... Mary-ERG John-DAT book-the-ABS read AUX-C ...

'... that Mary read the book to John'

b. **Nori**$_i$ irakurri dio Mirenek t_i liburua? [Wh$_{IO}$-V-S-O]

who-DAT read AUX Mary-ERG book-the-ABS

'Who did Mary read the book to?'

c. *Mirenek **nori** liburua irakurri dio? *[S-WH$_{IO}$-O-V]

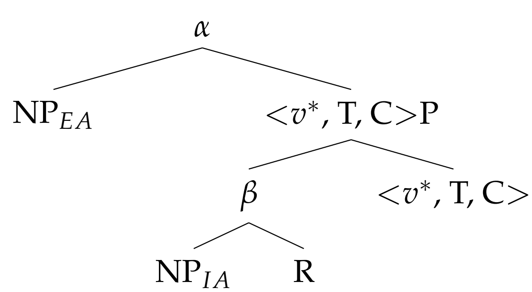
Mary-ERG who-DAT book-the-ABS read AUX?

(Ormazabal et al., 1994, 3,(9)-(10))

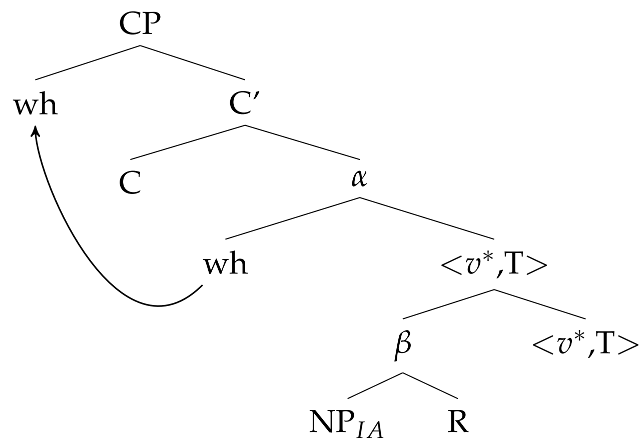
The declarative sentence (125a) indicates that Basque is a SOV language. However, (125b) and (125c) show that *wh*-movement is obligatory. Ormazabal et al. (1994) suggests that the *wh*-phrase moves to the specifier of C, but TP does not, which derives an interrogative sentence, whereas when TP moves to [spec,CP], it becomes a declarative sentence.

In the amalgamation theory, I propose here that within Basque, the derivation for the declarative sentence is a Japanese-type structure, while the derivation for the interrogative sentence is a German-type structure (here I do not adopt Kayne's approach).

(126) a. a derivation for a declarative sentence



b. a derivation for a interrogative sentence[39]



If this is on the right track, intra-variation can be captured by the theory of amalgamation.

In this section, we explore intra-variation within a particular language, Basque. Based on Bach's (1971) observation, SOV languages do not show overt *wh*-fronting. However, Basque interrogative sentences show an exception to this generalization, though Basque is a SOV language. Ormazabal et al.'s (1994) approach suggests that two options are available in this language: when TP is moved to [spec,CP], it becomes a declarative sentence; when a *wh*-phrase moves yo [spec,CP], it becomes an interrogative sentence. We interpret this as an intra-variation pattern where a declarative sentence includes the amalgam $<v*$,T, C$>$, whereas an interrogative sentence has a free-standing C, so the specifier is available for *wh*-fronting.

3.9 Conclusion

In this chapter, I have illustrated how operations in narrow syntax interact based on Obata et al. (2015), Obata & Epstein (2016), and Epstein, Obata & Seely (2018b). We started from an overview of the current state of theories on the locus of linguistic variation. Based on the minimalist view of UG, it is implausi-

[39] Although I omit some derivational steps, I assume that head movement of $<v^*$,T$>$ to C takes place later in the derivation.

ble to assume the notion of parameters as primitives of UG. It is assumed that the locus of linguistic variation comes from the lexicon or externalization (i.e., the Borer-Chomsky conjecture and the Berwick-Chomsky conjecture). In this chapter, we adopted the third factor principle-driven approach, where interaction between syntactic operations and underspecification derives multiple optimal derivations in the narrow syntax.

I also discussed another type of underspecification, namely, underspecification in the application of Merge to heads (i.e., optional amalgamation). This option allows us to generate different structures, some of which do not have specifiers. There are theoretical and empirical consequences of this approach. The theoretical consequences involve labeling theory and phase theory.

I proposed that the amalgamated elements are privileged and the labeling algorithm assigns those elements to be a label of the structure. This solves Chomsky's (2013) labeling problem with the subject-predicate structure where an unlabeled {XP,YP} structure arises. In particular, there is no solution for German and Japanese, at least, in Chomsky's (2013) framework, since the subject does not have to move to [spec,TP] in these languages. Once we adopt the proposed analysis, the amalgamated elements are properly selected as a label, eliminating the need for obligatory movement to [Spec, TP], which does not project separately.

I also argued that when amalgamation takes place for the phase heads, phasehood will disappear, based on Chomsky (2015) and Epstein et al. (2016). I showed that when C is a free-standing head, the edge condition applies (i.e., subextraction from the phasal edge is impossible), whereas there is no such condition when C is introduced as an amalgam with other functional categories.

Various empirical consequences are also attested under the amalgamation theory. I showed such consequences for different kinds of phenomena such as the presence/absence of subject-verb agreement, expletives, VP-fronting, VP-ellipsis, and *wh*-movement. Without referring to parameters, we deduce these either by applying or not applying the operation that creates amalgams. This derives different structures for each language. Namely, whether the specifier of a head (e.g., yielding spec-head agreement) exists depends on whether amalgamation applies to functional categories.

This underspecification approach also predicts that I-language-internal variation occurs, which is called intra-variation in Obata & Epstein (2016) and Epstein, Obata & Seely (2018b). I showed that Basque is a case where the declarative sentence is formed by amalgamation of $<v^*$,T,C$>$, like Japanese, whereas the interrogative sentence is formed by a free-standing C with obligatory *wh*-movement to the specifier of CP. This intra-variation cannot be captured by (macro-)parameter approaches.

This way, the underspecification approach captures a part of linguistic variation without resorting to (macro-)parameters. Other approaches to parameters by Biberauer & Roberts (2015) Roberts (2019) try to deduce (macro-)parametric effects only by assuming third factor principles (Chomsky, 2005). This emergent view of macro-parameters is not another kind of parameter, but rather it derives macro-parameters from the lexical items. The premise for this approach is to recapture 'parametric variation' that was captured by the macro-parameters. One of the implications of a macro-parameter analysis is that in the adult I-language system, the parameters should be fixed. As we discussed, however, there is intra-variation within a single I-language (e.g., English employs T-subject agreement, but it can employ T-object agreement for tough constructions, as Obata et al. 2015; Epstein et al. 2018b discussed), therefore the rule ordering can be different, which cannot be captured by the parameter approaches. Thus, it is desirable to adopt the underspecification approach, not only for the rule ordering, but also regarding which lexical items are amalgamated together.

Finally, this approach might have implications for language change and language acquisition. The underspecification approach suggests that children learn the rule ordering as they are exposed to input. There are also cases where the rule ordering is not acquired, which might cause language change, although exploring these implications would require further research which is beyond the scope of this dissertation.

CHAPTER IV

A Late-Insertion-Based Exoskeletal Approach to the Hybrid Nature of Functional Features in Creole Languages

4.1 Introduction

In the previous chapter, I discussed a different way of deriving parametric variation by appealing to underspecification of rule ordering in the narrow syntax, building upon an original proposal by Obata et al. (2015). Furthermore, I proposed that the combination of functional categories is also underspecified, which derives a cluster effect. In a similar vein, this chapter also approaches a different kind of variation (in multilingual contexts), by appealing to underspecification of functional feature combination, which allows Creole languages to emerge.

This chapter[1] investigates how to derive the possible linguistic structures for natural languages, with respect to Creole languages. I propose a hypothesis concerning the formation of functional categories (FCs) in Creole languages, based on the framework of what are called exoskeletal models (Borer, 2003, 2017; Grimstad et al., 2018; Riksem, 2018) and the Distributed Morphology (DM, Halle & Marantz 1993; Marantz 1997) framework. According to DM, syntactic categories are deter-

[1]Part of this chapter is published as Sugimoto & Baptista (2022).

mined by the combination of the categorizer (i.e., a nominalizer for *n*, a verbalizer for *v* etc.) and Root (an uncategorized category that includes only concepts, without categories such as nouns or verbs), which are combined by the single combinatory operation Merge. Furthermore, the categorized phrase is merged into the syntactic structures determined by the functional projections (e.g., CP, $v{*}$P, DP). In this way, FCs operate as a spine/backbone determining the domain of syntactic operations.

The research question in this chapter addresses how functional categories emerge, especially in Creole languages. In the context of the Minimalist Program for linguistic theory, lexical items have been stipulated to have phonological features, semantic features, and syntactic features that are stored in the lexicon (Chomsky, 1995, 2000). Lexical categories and functional categories are believed to be distinct; that is, lexical items typically have referential meaning or conceptual content, while functional categories do not have such meaning, but trigger 'grammatical' phenomena such as agreement.

In language mixing, words can be mixed within the same categories such as nominal phrases (e.g., a determiner is from one language and a noun phrase is from another), FCs are determined by one of the two languages and the stems (categorizer plus root) are determined by the other language. Thus, even if the language is 'mixed', it is taken not to mix the properties of FCs. I will argue that, in Creole languages, FCs can be recombined, giving them a hybrid nature. Thus, FCs in some Creole languages are not fixed or borrowed from one language (e.g., lexifiers/superstrates, substrates, etc.), but their features can be recombined (cf. Aboh, 2009, 2015).

In the principles and parameters (or Government and Binding (GB)) framework, possible language variation is restricted to parameters, which are binary values that are not specified at the beginning (e.g., word order, the pro-drop parameter), as well as variation in the properties of lexical items (cf. the Borer-Chomsky conjecture, Baker 2008). However, with the advent of minimalist syntax, syntacticians have been seeking to explain why UG has evolved in the way it has and not another way ("beyond explanatory adequacy," cf. Chomsky 2004. The same is true for parameters; if parameters exist, how have they evolved (cf. Richards, 2008)? One possibility is that there are no parameters at all (cf. Chomsky, 2017b; Rothman & Chomsky, 2018), as I discussed in the previous chapter.

Based on this discussion, I argue that Creole languages are, just as in monolingual scenarios and language mixing, I(nternal)-language systems where some parts are not specified at the beginning (in UG). In the previous chapter, I discussed how the way FCs are introduced can derive a cluster effect among English, German, and Japanese under the underspecification approach to the narrow syntax (§3.8). I took functional categories such as $v*$, T, and C as primitives in UG. In this chapter, I will argue with Roberts (2019) and his collaborators (Biberauer et al., 2014; Biberauer & Roberts, 2015) that there is also underspecification in formal features. My approach here is more specific to functional features under a model in which the combination of features in functional categories is underspecified. This process defines the kind of structural spine/backbone used in a given language. Furthermore, I argue that functional projections are not universal (Wiltschko, 2013, 2014; Rizzi & Cinque, 2016), but the features might be. In Grimstad et al. (2018) and Riksem et al. (2019) argue that the combination of choosing one functional category from one language system and choosing one stem from the other is restricted to language mixing. However, in the case of Creole languages, I propose that the formation of functional categories can be innovative; functional categories themselves could change through the process of competition and selection in the model proposed by Mufwene (2001, 2002). In this sense, unlike language mixing, the formation of functional categories does not solely depend on the superstrate or substrates.

The goal of this chapter is to demonstrate how the late-insertion-based exoskeletal model (Grimstad et al., 2018; Riksem et al., 2019; Åfarli & Subbarao, 2019) with feature recombination accounts for the properties of functional features in Creoles. In order to show this, this chapter is organized as follows: §4.2 introduces the literature on the notion of hybrid grammar and discusses how Creoles emerge in the first place. §4.3 presents the feature recombination model found in Mufwene (2001) and Aboh (2015), which seems to capture the nature of Creole hybrid grammars. I also discuss feature recombination in CVC as well as Aboh's (2009; 2015) analysis of Saramaccan. §4.4 introduces the issue of language mixing under the lexicalist minimalist syntax. §4.5 provides the framework that I will adopt in this chapter and §4.6 presents my proposal on feature recombination. Based on the model, I propose that the anterior marker *-ba* in Cabo Verdean Creole functions as a novel functional category which has a novel feature matrix that imposes a new insertion restriction on the exponent that is inserted (§4.7). I discuss other types of feature recombination in Creoles. §4.8 explores the implications of this study and §4.9 concludes this chapter.

4.2 Creole Genesis

Most theories of Creole formation assume that substrates (early Creole speakers' first languages) and superstrates (oftentimes colonial European languages) interact with each other within the socio/historical contexts of Creole emergence (see Baptista et al., 2018). As background to the discussion that follows, in this subsection, I briefly summarize Bickerton's (1981; 1984) language bioprogram hypothesis, Lefebvre's (1998) theory of relexification, Mufwene's (1996) founder principle, Kihm (1990) and Baptista's (2020) concept of convergence, Baker's (1994) interethnic medium of communication hypothesis, and the second language acquisition approach in Kouwenberg (2006).

Bickerton (1981, 1984) proposes the language bioprogram hypothesis, arguing that humans have a biological blueprint for language and that Creole languages most directly reflect this blueprint. He argues that prototypical features in Creoles (e.g., word order, tense-mood-aspect system, articles, etc.) are derived from the language bioprogram. McWhorter (2002) also argues that Creoles have prototypical features: (i) little or no use of tone, (ii) little or no inflectional morphology, (iii) few combinations or noncompositional combinations of derivational markers and roots.

Lefebvre (1998) proposes the "relexification hypothesis," arguing that the Creole languages are formed by adopting their syntax and semantics from their substrates and their phonological representation from their superstrate. On this issue, Muysken (1981) states that

> "Given the concept of lexical entry, relexification can be defined as the process of vocabulary substitution in which the only information adopted from the target language in the lexical entry is the phonological representation (Muysken, 1981, 61)."

Thus, this hypothesis predicts that the substrates have a major influence on Creole genesis. However, Mufwene (1996) proposes the opposite hypothesis, known as the founder principle. This principle assumes that "[t]he vernaculars spoken by the settlers of the new colony would establish themselves as the targeted norm (Velupillai, 2015, 181)," meaning that the main influence on Creole formation is from the superstrate/lexifier.

Kihm (1990) and Baptista (2020) argue that when source languages have common linguistic features independently of language contact, those common features are selected to participate in the formation of a given Creole language.

> "Given that a fortuitous formal similarity of really or apparently comparable elements from possibly very different languages is an attested and, after all, inevitable fact, one may expect spontaneous learners of

> a second language to grab at such elements and conflate them in their minds, by virtue of this principle [. . .] that you more easily learn what you think you already know (Kihm, 1990, 113)."

Based on a cross-linguistic comparison of 19 contact languages across 20 grammatical domains, Baptista (2020) offers the Pattern and Matter Mapping model to explain how convergence between substrates and superstrates operate in language contact (see also Baptista 2006).

Baker (1994) claims that innovations observable in Creoles emerge from interethnic communication between the first Creole speakers.

> "Pidgins and Creoles are successful solutions to problems of human intercommunication rather than the unhappy consequences of botched language learning or failed language maintenance (Baker, 1994, 65-66)."

Another dominant hypothesis about Creole genesis is L2 acquisition. For Wekker (1996), creolization is:

> "a gradual process of imperfect second-language acquisition by successive cohorts of adult slaves, extending over generations.(Wekker, 1996, 146)."

It has been argued in the literature (cf. Siegel, 2006) that both creolization and second language acquisition are connected, though these creolization has a different aspect; the target language (the language that the acquirers presumably want to learn) may not be the target of second language acquisition at some point. However, Kouwenberg (2006) points out that when second language acquisition is incomplete, the proficiency level is "frozen" (at a 'low' or 'intermediate' level), whereas in creolization, there might be language transfer from L1, which may create new grammatical functions in the target language, a phenomenon that does not last in typical second language acquisition.

These ideas are summarized in Table 4.1. The next section addresses the competition and selection model.

Principles/Approaches	Main influences	References
The language bioprogram	biologically universal properties	Bickerton (1981)
Relexification	Substrate	Lefebvre (1998)
The founder principle	Superstrate	Mufwene (1996)
Conflation/convergence	Substrate/superstrate	Kihm (1990), Baptista (2020)
Creativity	Innovation	Baker (1994)
Second Language Acquisition	SLA process	Kouwenberg (2006); Siegel (2006)
Competition and selection	Substrate/superstrate	Mufwene (2001); Aboh (2009, 2015)

Table 4.1: Approaches to Creole Genesis

All these approaches address the role of source languages (lexifier and substrates) in Creole formation. However, there has been less of a focus on proposing a formal grammatical approach to Creole genesis. In the next section, I introduce the competition and selection model as well as the notion of feature recombination.

4.3 Competition and Selection Model

Mufwene's (1996) notion of restructuring suggests that the process of creolization involves the selection of features from the feature pool that exists in the multilingual situation. Taking into account the linguistic features available in the contact situation, Aboh (2009, 2015) argues that a recombination of these features contributes to Creole grammars, based on Mufwene's (2002) assumption that languages are biological species. That is, "Creole languages are linguistic hybrids in the biological sense" (Aboh, 2009, 317). I provide more details below about these assumptions.

The notion of 'hybrid grammar' illustrated by Aboh (2009, 2015) captures some general aspects of Creole genesis (i.e. how Creoles emerge) by suggesting that formal features (syntactic, semantic, and phonological features) are 'recombined' in Creoles and that some of the Creole features can be traced back to the Creole superstrate and substrate(s). This idea is based on Mufwene's (2001; 2008) competition and selection model that stipulates that features from substrates and superstrates compete with each other in a multilingual setting and that some are selected from

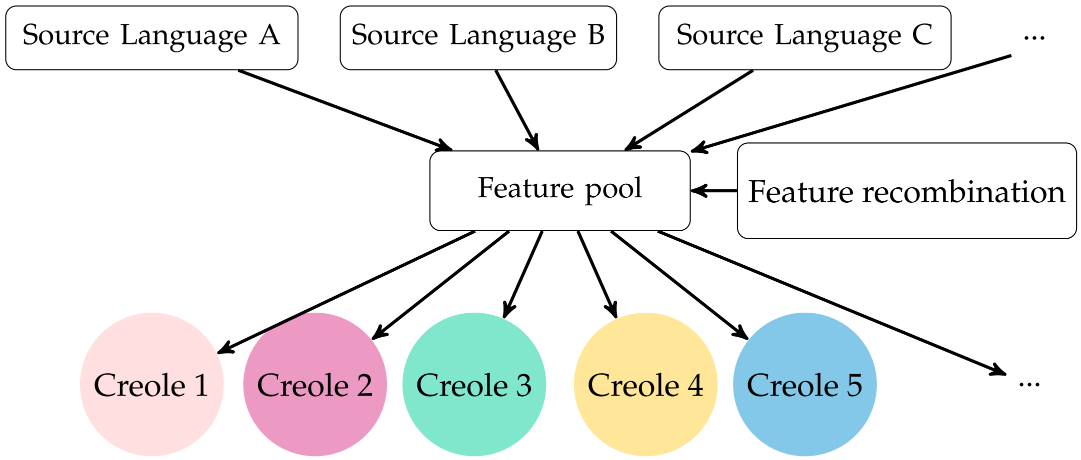


Figure 4.1: A competition and selection model

that feature pool whereas others die out. Although ecology plays a crucial role in Mufwene's model, Aboh (2009, 2015) argues that "competition and selection of linguistic features is free" (Aboh, 2009, 332), assuming that this process takes place within I-Creole (the internalized language system that is represented in the speaker's mind/brain). In sum, selecting formal features from the feature pool is free because external factors do not affect I-Creole.[2]

4.3.1 Feature Recombination: The Case of Saramaccan

Saramaccan is a language where one of the substrates is Gungbe and one of the superstrates is English. One can observe feature recombination in the Saramaccan light verb. For instance, so-called inherent complement verbs (ICV) in Gungbe "requires an object in their citation form" (Aboh, 2009, 328), and verbs like *ɖù* 'eat' change meaning depending on the object that follows them. To be more precise, Aboh (2009, 2015) argues that the verb *ɖù* 'eat' is a light verb, and the V is empty, resulting in the incorporation of N (i.e. N-to-V incorporation). As shown in (1), the

[2]Although we agree with Aboh that ecological factors do not interact with I-language, we suspect that feature recombination is not totally free. The process of acquiring formal features necessarily involves the environment/pool where the acquirers detect the formal features. Especially in the context of Creole genesis, the features of source languages are subject to feature recombination. In this sense, it is not entirely clear that feature recombination is totally free.

meaning of the verb ḍù 'eat' in Gungbe is altered by the following object: ḍù followed by a pronoun means 'to have a headache', as shown in (1a) but ḍù followed by 'money' means 'to spend' (1).

(1) Gungbe

a. Tà ḍù mì
Head eat 1SG
'I have a headache'

b. Kòfí ḍù kw´ɛ cè
Kofi spent my money
'Kofi spent my money' (Aboh, 2009, 329)

In contrast to Gungbe, the English verb eat has the specific meaning of 'ingesting N' and shows V to *v* movement (verb raising), which means that the verb 'eat' moves to the light verb position *v*.

Bearing this in mind, the verb *njan* 'eat' in Saramaccan has inherited hybrid features from both Gungbe and English. The example in (2) shows that the object can be realized in Saramaccan, which is compatible with the object requirement of Gungbe.

(2) Saramaccan

Amato njan di bakuba
Amato eta DET banana
'Amato ate banana.' (Aboh, 2009, 332, (13))

Another parallel with Gungbe is that Saramaccan also shows V-N incorporation. The Saramaccan examples in (3) align well with the Gungbe examples in (1).

(3) a. Njan moni
eat money
'to spend money'

b. Njan pena
eat pain
'to suffer' (Aboh, 2009, 333,(15))

On the other hand, the Saramaccan verb *njan* 'eat' can appear in sentences where the object is absent (as seen in (4)), which is compatible with English but not allowed in Gungbe, where the object is required.

(4) a. I njan kaa no?
2SG eat already Q
'Have you already eaten?'

b. Ai mi njan (kaa)
yes I eat already
'Yes, I have eaten (already).'

c. Ai mi njan soni
yes, 1SG eat something
'*Yes, I ate (something [non-specific]).'

'Yes, I ate something [specific]' (Aboh, 2009, 332, (13)-(14))

This state of affairs leads Aboh (2009, 2015) to conclude that "Saramaccan *njan* maps the semantic properties of English and Gbe 'eat' onto the syntax of English" (Aboh, 2009, 334), which recombines the properties of the verb 'eat' in Gungbe and English. Such examples show that feature recombination can involve phonological, semantic, and syntactic features of any given lexical item, in Aboh's (2009; 2015) proposal. The logical possibilities of feature recombination are summarized in table 4.2.

Lexical features of 'eat' in Saramaccan	
phonological features	*njan*
syntactic features	English
semantic features	English + Gungbe

Table 4.2: The 'hybrid' grammar in Saramaccan

These combination patterns should be possible and in principle play an important role in creolization. In sum, Aboh's feature recombination targets lexical items and shows how some of the properties of Saramaccan *njan* 'eat' can be argued to be derived from English (its superstrate) whereas others are derived from Gungbe (its substrate).

Based on the feature recombination mechanism, Aboh (2020) proposes a grammatical model that captures code-mixing.[3] He argues that the cognitive process that underlies feature-recombination is crucial for language acquisition, and the grammatical model that he proposes inserts vocabulary via the executive function at the Spell-out where the syntactic unit is sent off to the interfaces, i.e., the semantic component and phonological component (see Figure 4.2 below). The executive function is described as "a cover term for various cognitive processes involving attention control for the deliberate control of goal oriented actions" (Aboh, 2020).

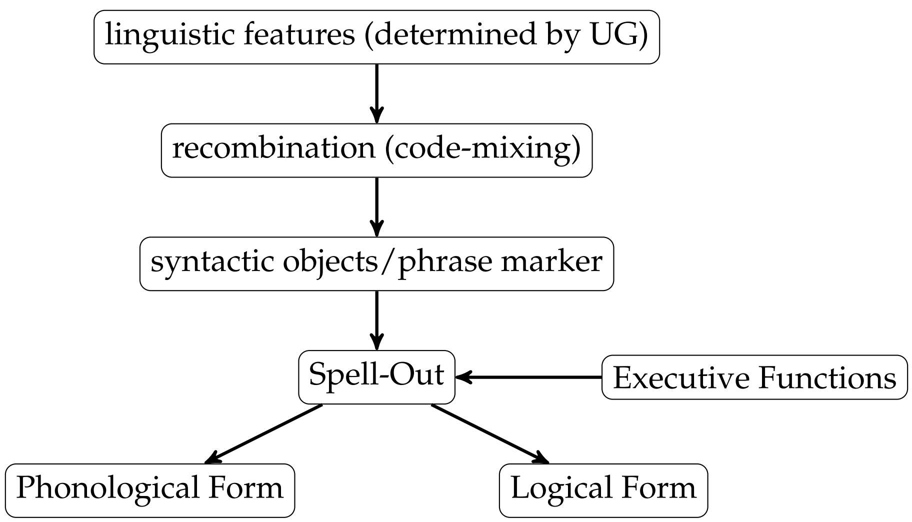


Figure 4.2: Aboh's (2020) model

In Figure 4.2, feature recombination takes place before introducing lexical items to the narrow syntax where syntactic objects are combined. Once the computation is done, it is sent off to the interfaces (i.e., phonological form and logical form). Spell-out is the point where the executive functions take place to select vocabulary in Aboh's (2020) model.

[3]Aboh (2020) defines code-mixing as follows: "the term code-mixing/switching refers to instances of language mixing in which speakers/signers combine properties of two or more languages in their utterances."

Several questions arise concerning the executive function. It seems the executive function plas an important role in Aboh's model (Abutalebi & Green 2008; Green & Abutalebi 2008). However, if the recombined features already have phonological, semantic, and syntactic features before being introduced into the narrow syntax, why does the executive function even exist? Is it possible to override the information before sending it to the interfaces? Why can't we just do code-mixing when vocabulary is inserted at the Spell-out point? Aboh (2020) assumes that his model is compatible with Distributed Morphology (DM) in that vocabulary insertion takes place at Spell-out (see §4.5 regarding the framework of DM). However, if the subset principle (Halle, 1997) is applied, vocabulary insertion takes place based on features in the case of functional features (see §4.5 in detail). Thus, an extra mechanism is unnecessary if DM is assumed. In this chapter, we adopt the idea that formal features are recombined, but we do not assume Aboh's (2020) model itself. Instead, I will assume the late-insertion-based exoskeletal approach and show that our approach captures multilingual data as well as monolingual data.

In this section, I discussed the competition and selection model. The examples from Saramaccan show that the linguistic features are not selected from just one language, but are recombined and used in a Creole language as a hybrid grammar. Although this competition and selection model is insightful, the competition and selection model does not have the full specific apparatus necessary to capture the hybrid grammar. That is, Aboh (2009, 2015) describes feature recombination by examining some attested data, but what is the mechanism behind the data?

In the next section, I contrast Aboh's feature recombination of lexical items like 'eat' with a different type of feature recombination that involves functional categories. More precisely, I show not only that feature recombination can apply to functional categories but also that while functional categories can draw some of their properties from their source languages, they can also display striking innovations. To demonstrate this point, we examine the Cabo Verdean anterior marker *-ba*.

4.3.2 Feature Recombination in Cabo Verdean Creole

The Cabo Verdean anterior marker *-ba* can append to both lexical verbs and auxiliaries, hence it participates in the Tense-Mood-Aspect-Mood (TMA) system of Cabo Verdean Creole (henceforth, CVC). On this issue, we do not label the phrases projected by TMA markers AuxP, AspP, MoodP, but we use instead FP for expository reasons,[4] as I elaborate in §4.7.

In order to support our feature recombination analysis of *-ba* in CVC which rests on contact between Portuguese (CVC superstrate) and Manjako (CVC substrate), we briefly provide below a sociohistorical introduction to CVC.

4.3.2.1 A Brief Sociohistorical Introduction

Cabo Verde (a former Portuguese colony) is an archipelago that lies in the Atlantic Ocean about 400 miles from Senegal (West Africa). The Portuguese arrived in Cabo Verde in 1445 and they were shortly thereafter followed by African enslaved populations (Kihm, 1994, 2). The enslaved populations are believed to have originated from the regions of Cacheu and Bissau and were composed of Jalofo, Peul, Bambara, Bolola, Manjako, Banhun, Mandinka, Balante, Bijago, and Feloupe people, among others (Brásio, 1962). As a result, CVC emerged from a mixture between nonstandard varieties of Portuguese and African languages, including Manjako, Mandinka, Wolof, and Temne, among others. In 1975, Cabo Verde became independent from Portugal, but CVC remained in close contact with Portuguese while moving away long ago from the African substrates that contributed

[4]See Baptista's (2002:160-161) TMA templates, which show the possible and impossible patterns of the combination of TMA markers. I limit myself to the anterior marker *-ba* in the analysis in this chapter since the treatment of the entire TMA system is beyond the scope of this chapter.

to its genesis.

4.3.2.2 The Functions of *-ba* in CVC

As a result of such contacts, the anterior marker *-ba* has been argued to have originated from Portuguese *-va*, a past tense marker that modifies Portuguese verbs whose infinitive form is *-ar* and that belong to the first conjugation of verbs (ex: *andar* 'to walk' > past tense *andava*, or *falar* 'to talk' > past tense *falava*), and from the Manjako form *ba* which marks the completion of an event (Kihm, 1994, 103). In order to demonstrate how feature recombination operates with respect to the anterior marker *-ba* in CVC, I provide below relevant data featuring *-ba* in CVC, Portuguese and Manjako. We start by examining the functions of *-ba* in CVC.

In CVC, the anterior marker *-ba* can be affixed to a nonstative verb to convey, pluperfect as in (5a), or to a stative verb to convery simple past, as in (5b), and can convey imperfective when combined with the markers *sta* and *ta*, as in (5c). The simple past of a nonstative verb involves a bare verb stem (no suffixation), as shown in (5d). When combined with the CVC TMA markers *sta* and *ta*, yielding *staba ta*, the combination can express past imperfective, as shown in (5e).

(5) Verbal Domain (TAM domain): *-ba* in CVC

a. Paulo kumeba katxupa.
Paulo eat+ba katxupa
'Paulo had eaten katxupa.'

b. Paulo staba duenti.
Paulo was sick
'Paulo was sick.'

c. Paulo staba ta kume katxupa to ki bu txiga
Paulo PROG+ANT MOOD eat katxupa when you arrive
'Paulo was eating katxupa when you arrived' (Baptista, 2020, 171,(13))

d. N kanta
I sing
'I sang'

e. N staba ta kanta
I sta-ba ta sing
'I was singing'

As mentioned earlier, the marker *-ba* has been argued to be derived from both the Portuguese anterior marker *-va* and from Manjako *(pe-)ba* which means 'finish' and marks the completion of an event (Kihm, 1994, 102-103).

These functions are summarized in (6).

(6) The functions of *-ba* (cf. Baptista 2002: 84)

a. Past perfect with nonstative verbs

b. Simple past with stative verbs

c. Imperfective when combined with TMA marker *sta* (and *ta*)

Let us turn first to the function of *-va* in European Portuguese (as the marker behaves differently in Brazilian Portuguese).

4.3.2.3 The Functions of *-va* in Portuguese

In Portuguese, *-va* appends to verbs belonging to the first conjugation ending in *–ar*, such as *cantar* 'to sing'.[5] For instance, *eu cantava* can be interpreted as 'I sang', 'I would sing or I used to sing', as shown in (7a). The simple involves on a different inflection on the verb, as shown in (7b). In addition, *-va* can append to the auxiliary *estar* to convey imperfectivity, as in (7c).

[5]Data come from Nélia Alexandre (p.c.) and Ana Luís (p.c.). I thank them for their thorough input on the various interpretations of *-va* in European Portuguese.

(7) *-va* in Portuguese

a. Eu cantava
   I sing-va
   'I sang, I would sing, I used to sing'

b. Eu cantei
   I sang
   'I sang'

c. Eu estava a cantar.

   'I was singing.'

4.3.2.4 The Functions of *ba* in Manjako

In Manjako, which is one of the substrates of CVC, *ba* is a verb meaning 'finish' that, when combined with other verbs, marks the completion of an event (8).

(8) Manjako

a-reala ba

'He finished eating'

(Kihm, 1994, 103)

One should also note that *ba* is a form that has been attested in the speech of enslaved individuals reported in 16^{th}-century Portuguese plays, suggesting that the languages in contact could indeed have contributed to its emergence in CVC.

(9) como mi saba primeyro
    as I was the-first
    'as I was the first' (Gil Vicente, 16th century, Língua de Preto)

(Teyssier, 1959, 235)

4.3.2.5 The Summary of the Functions of *-va* in Portuguese, *ba* in Manjako, and *-ba* in CVC

Having examined the formal features of *-ba* in CVC and its source languages, a feature recombination analysis would account for the fact that *-ba* in CVC combines the marking of anteriority and completion in addition to the postverbal position found in the source languages.

	-va in Portuguese	*ba* in Manjako	*-ba* in CVC
simple past	✓		✓
past habitual	✓		
imperfective with an auxiliary	✓		✓
completion		✓	✓
pluperfect			✓

Table 4.3: Summary of relevant items

Recall that the feature recombination analysis offered in Aboh (2009,2015) accounts for the superstratal and substratal sources of inherited properties of the verb *nyan* 'to eat' in Saramaccan, for instance, but would not be able to account for the innovation of the CVC stative-nonstative verb distinction (see (5a) and (5b) above), a distinction that is absent from source languages. To be more precise, the traditional feature recombination analysis does not account for the pluperfect reading of *-ba* when modifying the non-stative verb like *kume*, 'to eat' in (5a), a reading that is genuinely innovative in that it does not obtain in its source languages. In order to explain the innovative feature on *-ba* in CVC, I propose a different way of capturing feature recombination using a late-insertion-based exoskeletal model in §4.6. In the next section, I introduce issues raised by language mixing regarding the problems of lexicalist minimalist syntax.

4.4 Language Mixing and the Minimalist Lexicalist Syntax

Grimstad et al. (2018) point out that the minimalist lexicalist syntax (Chomsky, 1995, 2000, among others) poses a problem when analyzing grammatical agreement in language mixing (MacSwan, 2009). In the probe-goal Agree system, unvalued features must delete under Agree, otherwise the derivation crashes (Chomsky, 2000, 2001). The Agree system is below.[6]

(10) Agree

a. Matching is feature identity

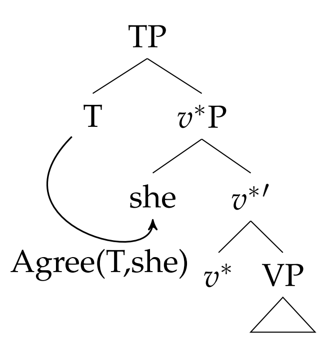
b. D[omain](P) is the sister of P (where P stands for probe)

c. Locality reduces to closest "c-command" (Chomsky, 2000, 122, (40))

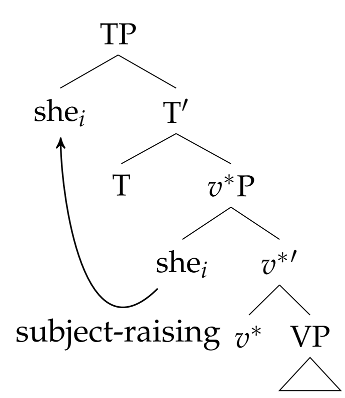
Agree is the mechanism that establishes the grammatical relation between the probe and the goal. Normally the probe has unvalued features which need to be valued under Match by another syntactic object, namely the goal.

The schematic structure in (21a) shows that T has unvalued ϕ-features and a valued case feature. T finds its goal in its minimal domain (in this case, *she*), then T gets its ϕ-features valued and *she* gets its nominative case valued.

(11) a.



b.



In the Spanish-English language mixing, a determiner should be Spanish and Noun should be English.

[6]I will not discuss the precise definition of Agree. See Chomsky (2000) and Chomsky (2001) for details. See also chapter 3 for relevant discussion.

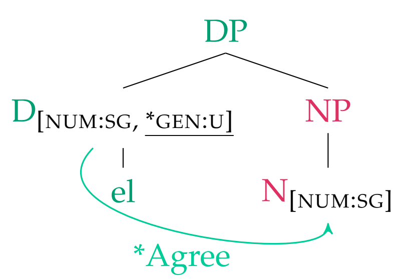
(12) Spanish-English mixing

a. el employer
the employer
'the employer.'

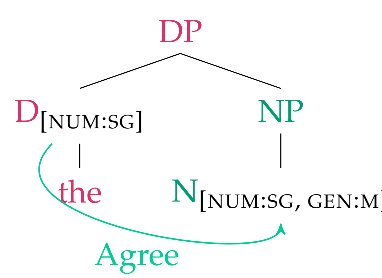
b. *the casa
the house
'the house'

Assuming that the determiner of Spanish has number-/gender-features, the D head *el* finds (in its c-command domain) the noun *employer*, which has only a number-feature, since English nouns do not have gender-features. In this analysis, the gender-feature remains unvalued, which should cause a crash, though this whole nominal phrase is grammatical (12a). On the other hand, Moro (2014) argues that the ungrammaticality of (12b) is because of the lack of a gender-feature on the English determiner.[7] However, if we assume that the determiner in English does not have an unvalued gender-feature, the sentence should be grammatical.

(13) a.



b.



Hence Moro's (2014) analysis should predict that (12a) is ungrammatical and that (12b) is grammatical, which is not borne out empirically.

Another example of language mixing involves American and Norwegian. According to Alexiadou & Lohndal (2018); Grimstad et al. (2018), American Norwegian is "a minority language existing in the midst of a language community heavily dominated by English (Grimstad et al., 2018)." First of all, in Norwegian, shown in (14b) and (14c), number agreement is obligatory.

[7]However, Liceras et al. (2008) claim that (12b) is not ungrammatical. They actually show that some bilinguals of Spanish and English (L1 English) actually prefer (12b). Liceras et al. (2008) argue that the Grammatical Features Spell-out Hypothesis offers an account for these examples. See their work for more details.

(14) Norwegian: grammatical gender language

a. dette hus-et
this.SG.N house-SG.DEF.N
'this house'

b. *denne hus-et
this.SG.M/F house-SG.DEF.N
'this house'

c. *dette hus-a
this.SG.N house-PL.DEF.N
'this houses'

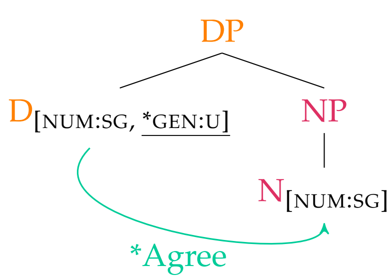
(15) illustrates language mixing in American Norwegian, where the nominal phrase also includes a determiner and a noun.

(15) American-Norwegian

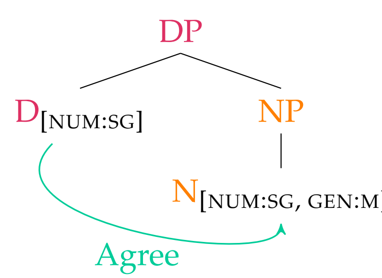
a. en blanket
a.M blanket
'a blanket'

b. the by
the city.SG
'the city'

(16) a.



b.



In (16a), the determiner is from Norwegian and it has an unvalued gender-feature, and the noun is from English. Assuming that the probe is D, it cannot find a valued gender-feature, since the English noun does not have a gender-feature. The condition on Agree is feature matching. However, there is no such matching in this example. Thus, the Agree system cannot value the unvalud feature. In (16b), a determiner is from English and the noun is from Norwegian. In terms of the number-feature, D can find a valued feature on N. Thus, there should not be a problem with this derivation. The problem in these examples is that (16a) is predicted to be ungrammatical, contrary to fact, since it is attested in American Norwegian.

In this section, I discussed word-internal language mixing where the functional category comes from one language and the complement of the functional head comes from another language. Since grammatical features on these languages do not necessarily match, the unvalued-feature problem arises in some cases under the lexicalist minimalist syntax model (cf. Chomsky, 2001), though the expressions are acceptable. Thus, descriptive adequacy is not satisfied. To address this issue, the next section provides an alternative grammatical model, namely a late-insertion-based exoskeletal model.

4.5 A late-insertion-based Exoskeletal Model

In order to overcome the problems discussed in §4.4, Grimstad et al. (2018) introduce an exoskeletal approach, which is a new approach to capture language mixing. An exoskeletal model, which is also called a neo-constructivist model, is proposed by Borer (2003, 2005a,b, 2013, 2017). The constructivist's claim is that the form-meaning pair is the linguistic unit, that is, the construction itself plays an important role in grammar (cf. Goldberg, 1995, 2003, 2006; Croft, 2001). The difference between a constructivist and a neo-constructivist/ late-insertion-based exoskeletal model is that the exoskeletal model has a generative engine (i.e. the structure-building operation, combing the basic building blocks of grammar) while the constructionist regards constructions as language-specific products that are stored in the lexicon.

It is important to note that although the model we assume here is a combination of Borer's (2003; 2005a; 2005b; 2013; 2017) exoskeletal model and of tenets from Distributed Morphology (Halle & Marantz 1993, among others), we do not necessarily adopt all assumptions underlying these two approaches. In what follows, we introduce only the most relevant assumptions to our analysis.

I assume that "all aspects of the computation emerge from properties of structure, rather than properties of (substantive) listemes" (Borer, 2005a, 21). The basic idea is that structure is not formed by lexical items (e.g., verbs). The assumption is instead that functional features are determined by syntactic structure and roots do not have grammatical features. To be more precise, roots might only refer to a concept and they are uncategorized. The category of the root is determined by the syntactic environment (e.g., combining a root with *v*, which stands for a verbalizer, will express a verbal element; see Alexiadou et al. 2014 and Alexiadou & Lohndal 2017 for a thorough discussion of roots). At Spell-out (when the derivation is completed phase by phase, e.g., Chomsky 2000, 2001, 2004), vocabulary items (phonological exponents) are inserted. As for functional features, the Subset Principle applies (Halle, 1997). The subset principle is defined as follows:

> "The phonological exponent of a Vocabulary Item is inserted into a position if the item matches all or a subset of the features specified in that position. Insertion does not take place if the Vocabulary item contains features not present in the morpheme. Where several Vocabulary Items meet the conditions of insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen."
>
> (Halle, 1997, 428)

By this principle, the exponents of the functional features are restricted. That is, the functional exponents are required to match all or a subset of functional features and if there is more than one candidate for insertion, the one matching the most will win.

This approach is combined with Distributed Morphology (DM). One assumption is that lexical items are not full-fledged/pre-packaged elements. Rather, the features are separated and introduced through the derivation. Although it has been an object of debate, let's assume that the primitive uncategorized element *root* is only a concept. Since it does not have a category, it has to be realized with a categorizer (such as verbalizer(*v*), nominalizer(*n*), etc.) to form a linguistic unit.

One important property DM is Late Insertion (List 2 in Figure 4.3): lexical items are inserted later, so their phonological form is realized *post-syntactically*. At the spell-out of the syntactic structure the vocabulary is inserted, not as a given property of the lexical item. The schema of the model is below (cf. Embick & Noyer, 2007).

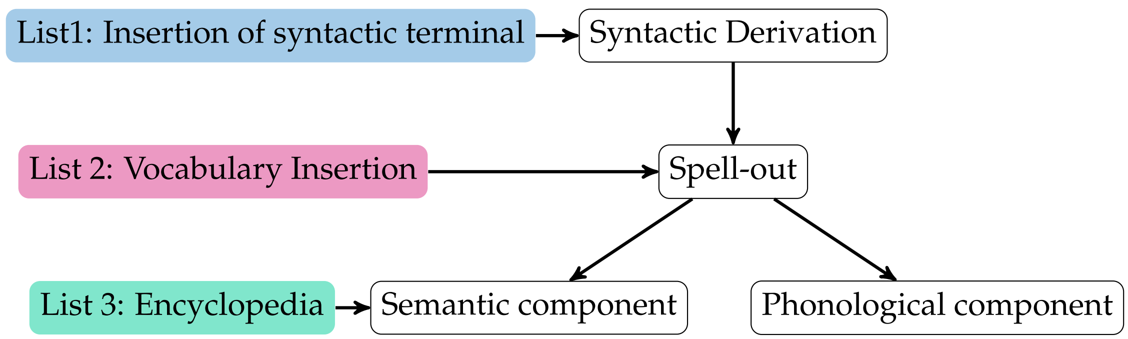


Figure 4.3: The grammatical model in Distributed Morphology (Embick & Noyer, 2007)

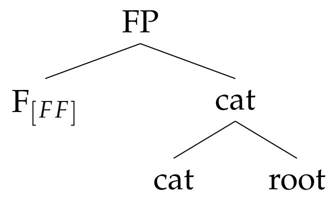
The syntactic terminals in List 1 in Figure 4.3 consist of roots (uncategorized element) and functional features/bundles. The roots can be realized as different forms (e.g., *destroy /destruction* in English) by combining with the categorizers, which categorize the roots during the derivation.

Functional features normally have their own grammatical features, such as [+PRES], [-PL], etc, based on the list in a particular language, whereas roots do not have such features. Therefore, roots are not subject to the Subset Principle. We also adopt the root-categorizer assumption.

(17) Categorization Assumption (Embick & Noyer 2007: 296, cf. Marantz 1997; Arad 2005 among others)
Roots cannot appear without being categorized; Roots are categorized by combining with category-defining functional heads.

The root combines with the categorizer, cat. The category of the stem depends on the type of categorizer (e.g., if cat is *v*, the stem becomes a verb). Another type of syntactic structure has to do with functional features.

(18)



In (18), FP stands for functional projection and the features of F vary, depending on the syntactic context. FP can be realized, as CP, TP, VoiceP/*v*P, DP, etc/ (see also Borer 2003; Ramchand 2008; Lohndal 2014 for variants of structures of functional projections). Compared to the functional/syntactic terminal, the stem (the categorizer + root) is less restricted. since the Subset Principle does not have to apply to these positions.

Overall, this approach is compatible with the underspecification approach that I developed in previous chapters. First, roots are uncategorized until they combine with categorizers. Second, late-insertion assumes that the vocabulary insertion takes place at spell-out. Until then the phonological features are underspecified, but they are subject to the Subset Principle if the insertion takes place for functional features. One of the consequences is that this model can capture language mixing, which is illustrated in the next subsection. Furthermore, as I will propose, the underspecification also applies to the functional features themselves, which derives their novel features and aspects of linguistic variation.

In the following subsections, I apply the late-insertion-based exoskeletal model to language mixing.

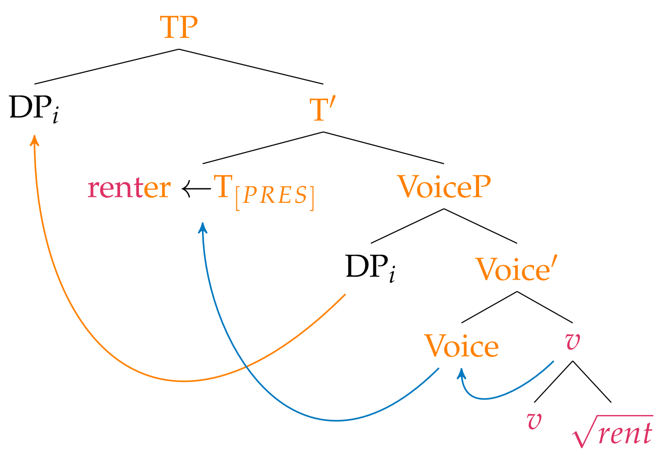
4.5.1 Language Mixing in American Norwegian

In the late insertion-based exoskeletal models proposed by Grimstad et al. (2018); Riksem et al. (2019), the syntactic feature bundles form the syntactic structure (Borer, 2005a; Lohndal, 2014) and the morphological exponents of the functional features are inserted later, subject to the Subset Principle (Halle, 1997). For instance, (19a) shows that the functional exponent *-er* is from Norwegian and the stem (the verbalizer + root) *rent* is from English.

(19) American-Norwegian

a. rent-er
rent-PRES
'rent(s)'

b.



(Riksem et al., 2019, 201,(8))

The tree in (19b) shows that since Norwegian does not have subject-verb agreement, the functional feature on T is only [PRES], standing for Present Tense. The verb moves up to T via the voice head and the functional exponent is *-er*.[8] Since the functional features on T are from Norwegian, the functional exponent has to come from Norwegian, not from English (see Table 4.4).

[8]Note that American Norwegian has a V2 rule, so the head moves up to C, which is not illustrated here.

TP in American Norwegian in (19a)		
	FP	stem
Functional Exponent	Norwegian	
Functional Feature	Norwegian	
Phonological Exponent		English

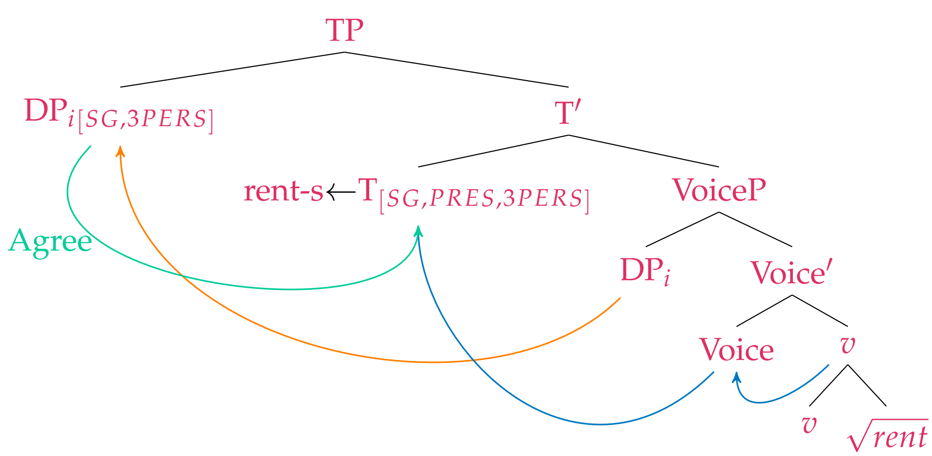
Table 4.4: A summary of the *v*P-TP structure of American Norwegian

If the functional features are from English, T has valued tense, unvalued number, and unvalued person features.

(20) English

a. she rents

b.



After valuation (e.g., via Agree; Chomsky 2000, 2001), the phonological exponent of the functional features [PRES], [SG], and [3PERS] is *-s* in English (i.e., *rents*). If the verb is already inflected in the lexicon (e.g., Chomsky 1995), this language mixing pattern cannot be captured at all. Furthermore, Riksem et al. (2019) capture this linguistic phenomenon without any additional assumptions beyond the late insertion-based exoskeletal model (using the null theory approach to language mixing, "is an approach that claims that the same theory that accounts for monolingual data should account for language mixing as well" (Riksem et al., 2019, 194)).

A similar language mixing pattern is observable in the nominal domain: Norwegian has double definiteness (the definiteness appears both as a determiner equivalent to English *the*, and as a suffix expressing number and gender agreement on both adjectives and nouns), as shown in (21a), and as such, F is responsible for these features and is located between D and the stem ([*n* Root]).[9]

(21) Norwegian: double definiteness

a. den gaml-e maskin-a
the.DF.SG.F old-DF.SG.F machine-DF.SG.F
'the old machine'

b. $[_{DP}$ D $[_{\alpha P}$ α $[_{FP}$ F $[_{NP}$ N ...]]]]

Here, I assume with Riksem (2018) that FP is realized between DP and *n*P and when adjectives and quantifier phrases are present, αP and CardP are also realized.[10]

In American Norwegian, the phonological exponent of the functional features (double-definiteness and masculine gender) is *-en* in (22) and (23a), whereas the phonological exponent is realized as feminine gender *-a* in (24a), which is from Norwegian, not from English. Note that the stem (*n*P) moves to [spec,FP].

(22) Nominals in American Norwegian

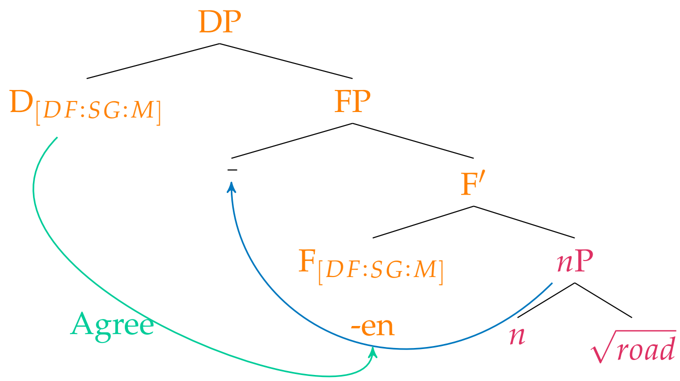
denne heritage tour-en
this-M heritage tour-SC.DEF.M
'This heritage tour' (Grimstad et al., 2018, 200,(13b))

(23) Nominals in American Norwegian

a. road-en
road-DEF.SG.M
'the road'

[9]See Julien (2003, 2005) for more details on the internal structure of the DP in Norwegian.
[10]See Riksem et al. (2019) for details of the nominal structure of Norwegian.

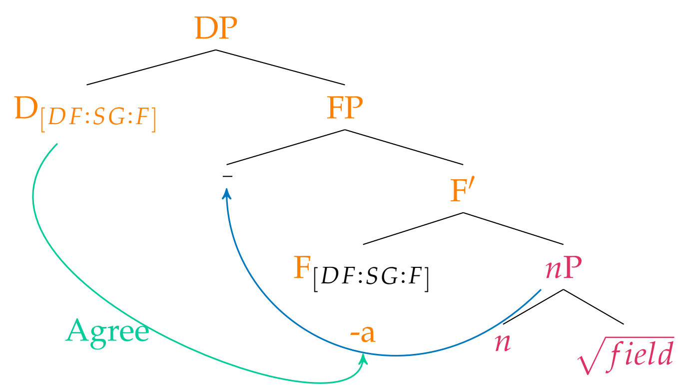
b.



(Riksem, 2018, 505,(19))

(24) a. den field-a
that.DF.SG.F field-DEF.SG.F
'that field'

b.



(Riksem, 2018, 508,(22a))

Again, in (23a) and (24a), the functional features for DP structures are from Norwegian (Table 4.5).

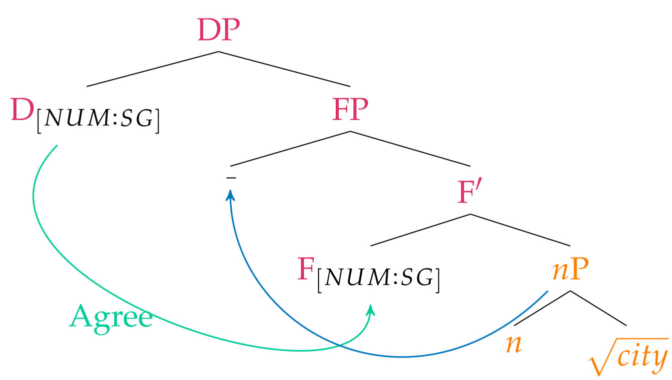
DP in American Norwegian in (23a) and (24a)		
	FP	stem
Functional Exponent	Norwegian	
Functional Feature	Norwegian	
Phonological Exponent		English

Table 4.5: A summary of the DP structures of American-Norwegian 1

The opposite pattern is also illustrated in Grimstad et al. (2018) in which functional features come from English and the stem originates from Norwegian.

(25) a. the by
the city
'the city'

b.



(Grimstad et al., 2018, 206,(17))

In Grimstad et al. (2018), it is assumed that F agrees with D, but in this particular case, there is no overt realization of the morpheme for this functional head. The summary is in Table 4.6 below.

DP in American Norwegian in (25a)		
	FP	stem
Functional Exponent	English	
Functional Feature	English	
Phonological Exponent		Norwegian

Table 4.6: A summary of the DP structure of American-Norwegian 2

This way, Grimstad et al.'s (2018) study demonstrates how the late-insertion-based exoskeletal model can capture effectively the language mixing patterns of American Norwegian nominal and verbal domains.

Compared to the lexicalist minimalist syntax, the late-insertion-based exoskeletal model is argued to better capture the possible grammatical patterns of American Norwegian.

4.5.2 Language Mixing in Dakkhini

In sum, Grimstad et al. (2018) and Riksem et al.'s (2019) late-insertion-based exoskeletal model provides an account for language mixing that proposes that functional features can come from one language (such functional features are language-specific and are subject to the Subset Principle), whereas the stem can come from another language. However, in CVC (see §4.3.2), the morphological realization of functional features is not as clear as in the American Norwegian case.

On this issue, Åfarli & Subbarao (2019) suggest in the context of language change that there are two possible changes in the functional domain. The first one is a reconstitution of the functional exponents, so that the existing exponent may be inserted by new criteria of the Subset Principle. To illustrate this point, they use examples from Dakkhini, a language that is the outcome of long-term contact between Hindi/Urdu and Telugu, and the authors provide a three-way comparison of complementizers in Dakkhini, Hindi/Urdu and Telugu. They show that in embedded questions, Dakkhini uses the same complementizer *ki* as Hindi/Urdu but with the crucial difference that the Dakkhini complementizer is clause-final (see (27)), just like the Telugu complementizer *-o:* (see (28)), whereas it is clause initial in Hindi/Urdu (see (26)). In sum, Dakkhini aligns with Telugu in having a clause-final complementizer.

(26) Hindi/Urdu: *ki* as Initial complementizer (IC)

Mujhe kyā patā [$_S$ **ki** rām kab āyega]?
I+DAT what known [$_S$ IC Ram when will-come]

'How do I know when Ram will come' (Åfarli & Subbarao, 2019, 32,(3))

(27) Dakkhini: *ki* as Final Complementizer (FC)

[$_S$ rām kab ātāē **ki**] mere ku kyā mālum?
[$_S$ ram when comes FC] I+DAT what known

'How do I know when Ram will come' (Åfarli & Subbarao, 2019, 32,(4))

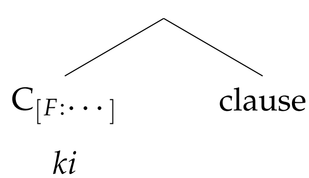
(28) Telugu: *o*: as FC

[$_S$ rāmuDu yeppuDu ostāD **-o:**] nā.ku yēmi telusu?
[$_S$ Ram when comes FC] +DAT what known

'How do I know when Ram will come?' (Åfarli & Subbarao, 2019, 32,(5))

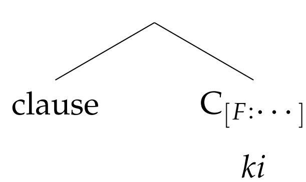
The CP structures featuring the complementizer of each language are shown below. These structures show that the Dakkhini CP structure headed by the complementizer *ki* (30) uses the same structure as Telugu in (31), whereas the functional exponent is from Hindi/Urdu (29).

(29) Hindi/Urdu



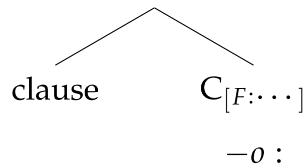
(Åfarli & Subbarao, 2019, 43,(13))

(30) Dakkhini



(Åfarli & Subbarao, 2019, 43,(14))

(31) Telugu



(Åfarli & Subbarao, 2019, 43,(15))

Åfarli & Subbarao (2019) suggest that Dakkhini *ki* is reconstituted to become a head-final complementizer and to match the functional feature of the complementizer *-o:* in Telugu. This is different from American Norwegian, where the functional exponent is related to the functional feature of a particular language (English or Norwegian in Grimstad et al. 2018; Riksem et al. 2019). Long-term contact with Hindi/Urdu and Telugu has also affected Dakkhini's *that*-clauses which also display innovations.

The second pattern that Åfarli & Subbarao (2019) consider is where a new functional exponent is inserted with new insertion criteria of the Subset Principle. As shown in the examples below, the complementizer in Dakkhini is in a final position in the embedded clause (26), and is argued by Åfarli & Subbarao (2019) to be inherited from Telugu. However, the functional exponent for the Dakkhini complementizer *bol ke* in (33) is neither from the functional exponent of the complementizer of Hindi/Urdu (*ki* in (32) nor Telugu (*ani* in (34)). Thus, the functional exponent is novel.[11]

(32) Hindi/Urdu: *ki* as IC

Mujhe nahi: patā [$_S$ **ki** si:tā gã:v cali: gayi: hai].
I+DAT NEG known [$_S$ IC Sita village has gone is]

'I did not know that Sita has gone to the village'

(Åfarli & Subbarao, 2019, 32,(6))

[11] Åfarli & Subbarao (2019:44) point out that *bol ke* is a calque from Telugu.

(33) Dakkhini: *bol ke* as FC

[$_S$ si:tā	g=a	gã:v ku	cale gayi:	**bol ke**]	mere ku	mãlum	nai:
[$_S$ sita	village	DAT	went away	FC]	I+DAT	known	not

'I didn't know that Sita had gone to the village'

(Åfarli & Subbarao, 2019, 32,(7))

(34) Telugu: *ani* as FC

[$_S$ sīta	u:ri	ki	wellindi	**-ani**]	nāku	teliyadu.
[$_S$ Sita	village	DAT	went	FC]	I+DAT	not known

'I did not know that Sita had gone to the village'

(Åfarli & Subbarao, 2019, 33,(8))

In this subsection, I showed that there exist different patterns of language mixing regarding the relation between functional feature and functional exponent. I reviewed Åfarli & Subbarao's (2019) illustration of two patterns that exist in Dakkhini due to a long-term contact effect. The first pattern, which is shown in (27), occurs when an existing exponent (e.g., *ki*) is inserted by new insertion criteria (i.e., the Subset Principle) due to reconstitution (Table 4.7).

Dakkhini complementizer *ki*	
Functional Exponent	from Hindi/Urdu *ki* (reconstitution)
Functional Features	from Telugu (final complementizer)

Table 4.7: The complementizer *ki* in Dakkhini

The other pattern, which is shown in (33), consists of the realization of *bol ke* in the embedded clause, where it is inserted as a new exponent by new insertion criteria (Table 4.8).

In Åfarli & Subbarao (2019) focus on functional exponents and their insertion criteria. In the following sections, I propose that the functional features themselves can be novel. Thus, what I will show is that the locus of the new criteria for the

bol ke in Dakkhini	
Functional Exponent	*bol ke* (a novel exponent)
Functional Features	from Telugu (final complementizer)

Table 4.8: The complementizer *bol ke* in Dakkhini

-ba in CVC	
Functional Exponent	*-ba* (a novel exponent)
Functional Features	*pluperfect* (a novel feature)

Table 4.9: The anterior marker *-ba* in CVC

functional exponents is the recombination of the functional features, so new insertion criteria are created (see Table 4.9). The study of Creoles like CVC adds to our understanding of the range of possible patterns of language mixing. Indeed, as we discussed in §4.3.2, the anterior marker *-ba* in CVC has multiple functions, some of which overlap with source languages (*-ba* expresses anteriority as in Portuguese and completion, as in Manjako), whereas others do not overlap with source languages and are totally novel (the pluperfect reading with non-stative verbs).

In order to account for this type of language mixing, the next section introduces our proposal that functional features themselves can be recombined, leading to the rise of novel functional features and novel functional exponents.

4.6 A Proposal

In the previous sections, I summarized the application of the late-insertion-based exoskeletal model to language mixing based on Grimstad et al. (2018), Riksem et al. (2019), and Åfarli & Subbarao (2019). In American Norwegian, the functional features are from one language (either from English and Norwegian) and the functional exponent is subject to the Subset Principle. Dakkhini shows that the feature matching between functional features and the functional exponent are changed for the complementizers *ki* and *bol ke*. Åfarli & Subbarao (2019) suggest that the functional exponent could be an innovation. Note that in Dakkhini, the functional exponent for the complementizer is novel, but the structure is still from Telugu and the function of the complementizer remains the same in Dakkhini and Telugu (i.e., indicating the *that*-clause).

In contrast, in CVC, the novel pluperfect reading of the anterior marker *-ba* does not originate from source languages (Portuguese and Manjako). This is an entirely novel pattern. Although there are several overlaps between *-ba* in CVC, *-va* in Portuguese (CVC and Portuguese share the expression of anteriority), and ba in Manjako (CVC and Manjako share the expression of completion), *-va* in Portuguese and ba in Manjako do not have exactly the same function (pluperfect) as *-va* in CVC.

In order to capture this innovation and the novel structure, I propose the following syntactic structure for Creoles , based on key assumptions presented in (35).

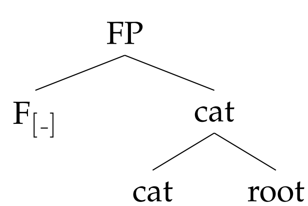
(35) A proposal:

In Creole languages, functional categories can be but need not be directly inherited from source languages. When such features are not directly inherited from source languages, they are decomposed into features and are recombined as "hybrid functional categories."

(36) Schema of the syntactic structure (see also (18))

a. [$_{FP}$ $F_{[-]}$ [cat root]]] (where cat is a categorizer)

b.



Based on the late-insertion-based exoskeletal model and (35), I assume here that the functional feature F is formed through competition and selection (between the languages in contact) and forms the syntactic structure. The structure illustrated in (36) is identical to the structure in (18), while the functional feature in (36) is recombined. This is distinct from the language mixing pattern observable in American Norwegian, in that functional features do not originate from just one language, and from the patterns that are shown in the Dakkhini data from Åfarli & Subbarao (2019). The logical possibilities for Creoles are that functional features are (i) from one or several substrates, (ii) from the superstrate, or (iii) can be novel/recombined features. Possibility (iii) suggests that the recombination of functional features could allow for a novel functional exponent. Functional exponents, nonetheless, might come from source languages since they are available during Creole genesis.

Functional Features	
phonological features	late insertion
functional features	(i) substrate, (ii) superstrate, (iii) recombination
semantic features	late insertion

Table 4.10: Functional features under the proposed model

Moreover, the late-insertion-based exoskeletal model is, in principle, applicable not only to language mixing or monolingual grammar, but also to all sorts of other scenarios of language contact, involving for instance bilingualism and heritage languages.

In the next section, I will not only illustrate the pattern (iii) in Table 4.10 but also demonstrate how feature recombination is operationalized in CVC.

4.7 Analysis

4.7.1 The Anterior Marker *-ba* in CVC

Going back to CVC examples regarding the anterior marker *-ba*, recall that some of its functions come from Portuguese (anteriority) and others from Manjako (completion), whereas others represent a novel feature, namely pluperfect, which does not exist in *-va* in Portuguese or *ba* in Manjako (37 = 5a).

(37) (=(5a))

Paulo kumeba katxupa.
Paulo eat+ba katxupa

'Paulo had eaten katxupa.'

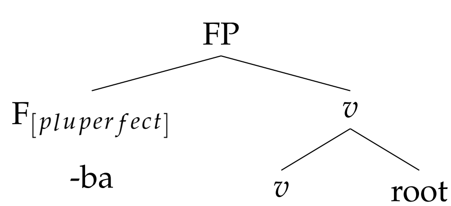
See the table below for a summary.

	-va in Portuguese	*ba* in Manjako	*-ba* in CVC
simple past	✓		✓
past habitual	✓		
imperfective with an auxiliary	✓		✓
completion		✓	✓
pluperfect			✓

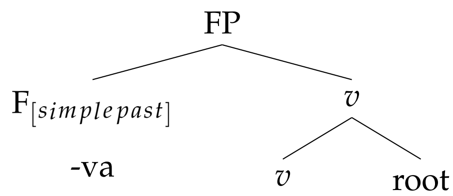
Summary of relevant items (=Table 4.3)

The relevant structures for each language are represented in (38) for CVC, (39) for Portuguese and (40) for Manjako.

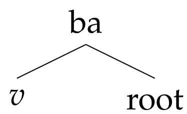
(38) The structure for *-ba* in CVC



(39) The structure for *-va* in Portuguese (in a case of simple past)



(40) The structure for *ba* in Manjako



The anterior marker *-va* in Portuguese and *ba* in Manjako do not have a pluperfect function. The exponent of pluperfect in Portuguese is not *-va*. For instance, the Portuguese equivalent of CVC *N kumeba katxupa* 'I had eaten katxupa' is *Eu tinha comido katxupa* 'I had eaten katxupa'. As for Manjako, *ba* is realized as the verb 'finish'. Thus, CVC must have a unique functional head/projection, which I call F (since the TMA system is a complex system, I assume the general functional feature here for the purpose of exposition), which includes pluperfect aspect. I also assume that the stem [*v* root] moves to F to get the right order ("V-ba," Baptista 2002). The entire structure includes the spell-out of the functional exponent, along with the [*v* root] structure, pronounced as 'V-ba'.

As a result of feature recombination of the Tense marker, the functional feature itself ([pluperfect]) might change in a unique way, and the exponent is easy to insert here since the insertion restriction (i.e., the Subset Principle) becomes new due to the recombined functional feature of *-ba* (pluperfect). This pattern is different from Dakkhini in that it is not clear which language among the source languages provides the functional projections in CVC. In the example of the complementizer *bol ke* in (33), the functional features are still from Telugu, as Åfarli & Subbarao (2019) argue, which is a distinct pattern from what *-ba* in CVC shows. Of course, this does not undermine the idea that CVC might be the result of long-term language contact between its source languages, leading to the emergence of *-ba* in CVC being able to express a novel feature, pluperfect, when modifying non-stative verbs.

Interestingly, the morphemes *-ba* in CVC, *-va* in Portuguese, and *ba* in Manjako look similar in their forms. Although each form has a different function, there are still some overlaps, as we discussed in §4.3.2 (see Table 4.3). This suggests that 'similar form and function' in the languages in contact could lead to congruence, which in turn may make it easier to implement a novel functional exponent (see Baptista 2020 on the role of congruence in the genesis of Creole languages).[12]

In the following two subsections, I will add more evidence of feature recombination of functional features under the late-insertion-based exoskeletal model.

4.7.2 Nominal Structures of Saramaccan

In this subsection, I discuss the nominal structure of Saramaccan. I will argue here that the nominal structure in Saramaccan is a novel structure that results from feature recombination.

Saramaccan's lexifiers are English and Portuguese, and one of the substrates is Fongbe. Lefebvre (2013) and Lefebvre (2015) argue that the Fongbe nominal structure and that of Saramaccan display similar properties except for word order, as shown in below.

(41) a. Fongbe

àsɔ́n [nyɛ̀ tɔ̀n] élɔ́ ɔ́ lɛ́
crab [me GEN] DEM DEF PL
'these/those crabs of mine' (Lefebvre, 2015, 19,(2))

[12] Another way of analyzing *-ba* in CVC is that the functional features are from Portuguese, and its functional features for pluperfect are reconstituted, so that *-ba* can be inserted based on the new insertion criteria.

b. Saramaccan

déé/dí físi [u mí] akí
PL/DEF fish [CASE me] DEM
'these/this fish(es) of mine' (Lefebvre, 2015, 19,(3))

c. English

*the my these crabs (Lefebvre, 2015, 19,(4))

In Saramaccan, as in Fongbe, a possessor phrase, a demonstrative, and the definite determiner can all co-occur in a nominal structure, whereas English does not have such a structure. However, the word order in Saramaccan is different from that of Fongbe. The word order of the Saramaccan nominal structure is similar to that of English in that the determiner and possessive can precede the noun.

(42) word order in nominal structures

a. Fongbe: POSSP NOUN POSSP DEM DEF PL

b. Saramaccan: PL DEF POSSP NOUN POSSP DEM

c. English: {DEF, POSS, DEM} NOUN(.PL) POSSP (Lefebvre, 2015, 61,(81))

The fact that the definite determiner, possessive, and demonstrative can co-occur in Saramaccan and Fongbe nominal structures and the fact that the determiner and/or possessive can precede a noun in Saramaccan and English suggest that the nominal structure in Saramaccan might come from Fongbe, while the word order might come from English.

However, a closer look at Fongbe and Saramaccan shows that the number category is realized as a free morpheme, whereas the number category and the definite determiner are in complementary distribution in Saramaccan as shown below (cf. Lefebvre, 2013).

(43) a. Fongbe

àsɔ́n ɔ́ lɛ́
crab DEF PL

'the crabs'

b. Saramaccan

déé / dí físi
PL / DEF fish

'the fish(es)' (Lefebvre, 2013, 45,(31)-(32))

More specifically, the morpheme *Déé* is a plural definite determiner (McWhorter & Good, 2012).

(44) a. Dí dí mɛ́síte tá léi, h´ɛ déé míi tá woóko gó dóu
when DEF teacher IMF reading, then DEF.PL child IMF work go arrive
'While the teacher was reading, the children continued working'
(McWhorter & Good, 2012, 77,(11))

b. A léi mi déé fóótóo.
3S show 1S DEF.PL photo
'He showed me the photographs.' (McWhorter & Good, 2012, 77,(14))

dí is realized for the singular determiner. Notice that there is no plural indefinite determiner in Saramaccan, according to McWhorter & Good (2012).

These data suggest that the functional heads in Saramaccan are not strictly from Fongbe. For the purposes of this chapter, I limit myself to analyzing the properties of the definiteness and number features in Fongbe, Saramaccan, and English.

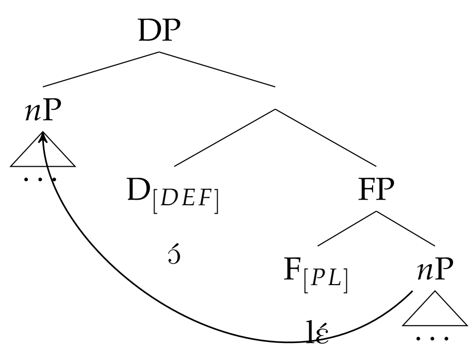
To summarize the observations so far, Fongbe, English, and Saramaccan have determiners, but the realization is different across these languages when plural features are expressed. In Fongbe, the determiner and the number-feature are realized separately as free morphemes, whereas in Saramaccan, the number-feature is realized on the determiner. In English, the number-feature is realized as a suffix on the noun (i.e., a bound morpheme). As for word order, English and Saramaccan have the same pattern (DEF>noun), while Fongbe does not. Table 4.11 is a summary of the relevant features.

	Fongbe	English	Saramaccan
Determiner	✓	✓	✓
Plural determiner			✓
Separate projections (determiner and plural)	✓		
Free morpheme (PL)	✓		✓
DEF >noun (word order)		✓	✓

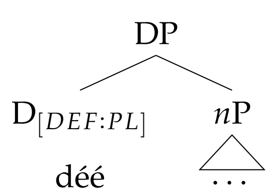
Table 4.11: A summary of the properties of nominal structure in Fongbe, English, and Saramaccan

Based on these data, I assume that the DP structure for each language is represented below.

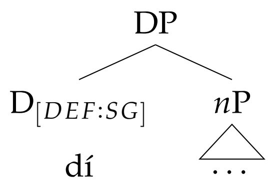
(45) a. Fongbe (cf. Aboh, 2019)



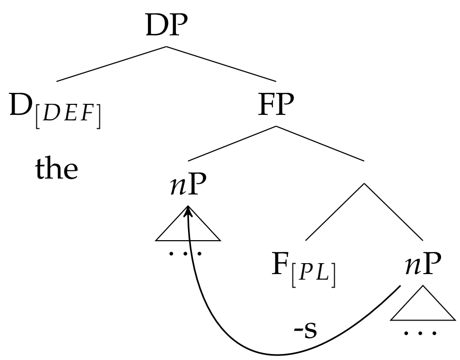
b. i. Saramaccan: definite + plural



ii. Saramaccan: definite + singular



c. English



Since the number category and the definite determiner are in a complementary distribution in Saramaccan, I assume that a [pl]-feature is on the determiner head in (45b-i), while a [sg]-feature is on the determiner head in (45b-ii). In Fongbe, however, the determiner head and the number head are assumed to be separate since the number feature and definite head can co-occur (cf. Aboh, 2019). In English, I assume that the determiner head and the number head are distinct projections because the plural marker can be realized as a bound morpheme, whereas the number feature is realized on the determiner in Saramaccan.

As represented in these structures above, I argue that the properties of the nominal structures in Saramaccan emerge from a novel functional projection. In Fongbe, the definite feature and the number feature are realized as separate functional projections and each functional feature has its own functional exponent; for the determiner, ɔ́ is realized and for the plural marker, *lɛ̌* is realized, as shown in (45a). In English, unlike Saramaccan and Fongbe, the determiner feature is realized as *the* and the plural marker is realized as a bound morpheme *-s* (45c). I propose here that the determiner feature and the plural feature are recombined and form a novel functional head in Saramaccan. The determiner is realized as *dí*, while the plural determiner is realized as *déé*, as shown in (45b-i). Accordingly, the functional exponent of these functional features becomes novel.

4.7.3 The Complementizer *ki* in CVC

In §3.7, I discussed different ways of deriving the possible complementizer agreement in HC, CVC, and BP. In this subsection, I explore the complementizer *ki* in CVC, its substrate Wolof, and its superstrate European Portuguese.

First, I discuss the complementizer in Wolof, which is one of the substrates of CVC. In Wolof, there are no complementizer markers that are exclusively used in declarative sentences. In embedded, non-interrogative sentences, *ne* is realized as a force complementizer.

(46) Defe-na-a ne macc-na-ñu màngo b-i.
think-FIN-1SC that suck-FIN-3SG mango CL-DEF.PROX
'I think that they sucked the mango' (Torrence, 2013, 77,(43))

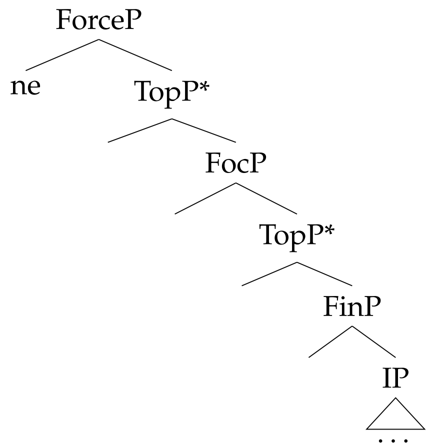
According to Torrence (2013), *ne* is homophonous with the verb *say* or *tell*.

(47) Ma ne Ayda (*ne) macc-na-a màngo b-i.
1SG say ayda that suck-FIN-1SG mango CL-DEF.PROX
'I told Ayda that I sucked the mango.' (Torrence, 2013, 78,(44))

Torrence (2013) assumes that *ne* is a Force head, which is the highest functional projection in Rizzi's (1997) left periphery analysis where the CP projection is divided into some distinct projections.[13]

[13] In the following, I won't adopt Rizzi's (1997) left periphery analysis, but see Rizzi & Cinque (2016) for an overview of cartographic structures in syntax.

(48)



When it comes to interrogative sentences in Wolof, there are two types of *wh*-constructions. The first one forms is an overt *wh*-fronting with an optional Q-particle in a cleft form (50). The other form involves null *wh*-phrase fronting with complementizer agreement (51).

Wh-expressions consist of a noun class consonant (a class marker, CL will be used for this marker in glosses) and the *wh*-element *-an* (Torrence, 2013).

(49) *wh*-forms in Wolof

a. k-an 'who'

b. f-an 'where'

c. l-an 'what'

d. . . . (Torrence, 2013, 90,(85))

Wolof has *wh*-movement with an optional Q-particle.

(50) An interrogative sentence in Wolof

(An-a/i) l-an l-a Isaa lekk?
Q_{wh}-DET CL-*an* XPL-COP isaa eat

'What is it that Isaa ate?' (Torrence, 2013, 91,(87a))

In this example, *l-an* is an overt *wh*-expression, which follows the optional Q-particle.

Torrence (2013) also proposes the existence of a null *wh*-expression that agrees with the complementizer *k-u* in (51).

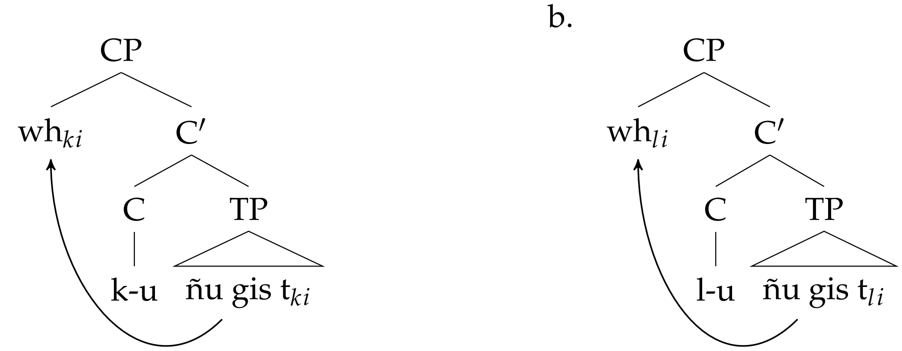
(51) K-u ñu gis?
CL-*u* 3PL see
'who did they see' (Torrence, 2013, 164,(2a))

The *u*-form of the complementizer depends on what it agrees with. In (51), it is realized as *k-u*; in this case, the question is asking about a single person. When the question is asking about a thing, the complementizer becomes *l-u*.[14]

(52) L-u ñu gis?
CL-*u* 3PL they see
'What did they see' (Torrence, 2013, 164,(2b))

Torrence (2013) then proposes the structures for (51) and (52) below. The fronted *wh*-expression is a null element, but the complementizer *k-u* agrees with the null element (*wh* in the tree structure below) in [spec,CP].

(53) a. b.



(Torrence, 2013, 165,(4))

In Torrence (2005), the evidence for this structure comes from the following; (i) the *-u* forms do not have the same distribution as the lexical DPs, (ii) the *-u* forms have the same distribution as the C head, (iii) C agrees with a DP, and (iv)

[14]See Torrence (2013: 167, (9)) for different complementizer agreement patterns with different *wh*-phrases.

the *-u* forms and the *-an* forms are sensitive to both general and language-specific constraints on movement.[15]

The distribution of the *-u* form (complementizer) depends on what kind of *wh*-phrase it is and where it is from.

(54) a. subject

k-u togg ceeb bi ak jën wi
CL-u cook rice the and fish the

'who cooked the rice and the fish?'

b. direct object

y-u jieéén ji togg
CL-u woman the cook

'what(pl) did the woman cook?'

c. locative adjunct

f-u jieéén ji togg-e ceeb bi ak jën wi
CL-u woman the cook-LOC rice the and fish the

'where did the woman cook the fish and the rice?'

d. applied object

ñ-u negeen ubbéél bunt bi
CL-u 2PL open-ben door the

'who(pl) did y'all open the door for?'

e. instrumental object

l-u Isaa ubbéé bunt yi
CL-u isaa open-instr door the.PL

'what did Isaa open the doors with' (Torrence, 2005, 80,(4))

In this way, Wolof expresses two types of interrogative sentences, and one with overt *wh*-fronting or one with a null *wh*-expression in [spec,CP] where complementizer agreement takes place.

[15]See more details in Torrence (2005), chapter 2.

In the null *wh*-movement pattern, complementizer agreement is obligatory for the highest CP clause, while agreement in the lower CPs is optional. Recall that *k-u* agrees with the null *wh*-element (which is represented as wh_{k_i} in the examples below), whereas *l-a* does not.

(55) Optional complementizer agreement in Wolof

a. [wh$_{ki}$ k-u Kumba wax [ne k-u Isaa defe [ne k-u Maryam dóór t$_{ki}$]]]?
[WH CL-*u* kumba say [that CL-*u* isaa think [that CL-*u* Maryam dóór t$_{ki}$]]]
'Who did Kumba say that Isaa thought that Maryam hit?'

b. [wh$_{ki}$ k-u Kumba wax [ne l-a Isaa defe [ne l-a Maryam dóór t$_{ki}$]]]?
[WH CL-*u* kumba say [that XPL-COP isaa think [that XPL-COP Maryam dóór t$_{ki}$]]]
'Who did Kumba say that Isaa thought that Maryam hit?'

c. [wh$_{ki}$ k-u Kumba wax [ne l-a Isaa defe [ne k-u Maryam dóór t$_{ki}$]]]?
[WH CL-*u* kumba say [that XPL-COP isaa think [that CL-*u* Maryam dóór t$_{ki}$]]]
'Who did Kumba say that Isaa thought that Maryam hit?'

d. [wh$_{ki}$ k-u Kumba wax [ne k-u Isaa defe [ne l-a Maryam dóór t$_{ki}$]]]?
[WH CL-*u* kumba say [that CL-*u* isaa think [that XPL-COP Maryam dóór t$_{ki}$]]]
'Who did Kumba say that Isaa thought that Maryam hit?'

(Torrence, 2013, 258,(66))

Notice that in these examples, the complementizer *ne* co-occurs with the complementizers *k-u* and *l-a*.[16]

So far, I have discussed the CP structures fro declarative sentences and interrogative sentences in Wolof. The complementizer for a declarative sentence can be

[16] When C does not agree with a *wh*-phrase, it is realized as *l-a*, which is an expletive-copula form. That is, when C-agreement does not occur, the cleft formation is applied (Torrence, 2013, 258).

ne; for interrogative sentences, there are two types of *wh*-constructions (Torrence, 2013). The first one is formed by a cleft form where overt *wh*-fronting takes place with an optional Q particle. The other way of expressing interrogative sentences is that a null *wh*-phrase moves to the front of the sentence and the complementizer agrees with it. In this case, the realization of the complementizer depends on what kind of *wh*-phrase it agrees with. For the purposes of this section, I discuss the complementizer *ki* in CVC and the complementizer *quem* in European Portuguese (a superstrate of CVC) to compare them with the complementizer in Wolof (a substrate of CVC) below. I show that the complementizer *ki* in CVC has a novel feature, distinct from the complementizer in Wolof and the complementizer *quem* in European Portuguese.

Let us discuss the distribution of the complementizer *ma* and *ki* in CVC. The complementizer *ma* 'that' introduces an embedded declarative clause when illocutionary verbs are involved.

(56) a declarative complementizer in CVC

a. Maria fla [$_{CP}$ ma ses fidju ta bai skola]
Maria say(PFV) [$_{CP}$ that POSS.3PL son IPVF go school]
'Maria said that her sons go to school.'

b. Nu atxa [$_{CP}$ ma mininu ka djuga bola n'es kau]
1PL think(PFV) [$_{CP}$ that boy NEG play(PFV) ball in-DEM place]
'We think that the boys didn't play ball in this place'

c. Djon odja [$_{CP}$ ma Maria kunpra sukrinha]
Djon see(PFV) [$_{CP}$ that Mari buy(PFV) sweet]
'John saw that Mary bought sweets'

d. Ta parse-m [$_{CP}$ ma bu sta mariadu]
IPFV parecer-1SG [$_{CP}$ that 2SG be bored]
'It seems to me that you are bored.' (Alexandre, 2012, 64)

Ma obligatorily appears after illocutionary verbs, while the other CVC complementizer, *ki* cannot.

(57) João fra-m ma/*ki/*∅ Maria kupra libru.
John told+me C Maria bought book
'John told me Mary bought the book' (Baptista & Obata, 2015, 171,(32))

However, when the *wh*-phrase is fronted, the complementizer is realized as *ki*, not as *ma* (see also §3.7.2).

(58) Kenhi ki fra-m kuze ki/*ma/*∅ Maria kunpra?
who C told+me what C Maria bought
'Who told me what Mary bought?' (Baptista & Obata, 2015, 171,(33))

Baptista & Obata (2015) argue that "[t]he complementizer *ma* changes to *ki* iff a *wh*-phrase is interpreted at its Spec position; in other words, if a *wh*-phrase is interpreted in the embedded Spec-CP, then *ki* must appear" (Baptista & Obata, 2015, 172).

Furthermore, when a *wh*-phrase appears in an embedded clause, the embedded complementizer has to be realized as *ki*.

(59) Kenhi *(ki) odja *(ki) João kai di bisikleta?
who C saw C João fall from bicycle
'Who saw that João fell from the bicycle?'

(60) Kenhi *(ki) bu ubi *(ki) João konbida onti?
who C you hear C João invite yesterday
'Who(m) did you hear that João invited yesterday?'

(Obata et al., 2015, 6,(8)-(9))

Finally, we discuss complementizer agreement in European Portuguese. In the case of a *wh*-object sentence (see (61a)), a *wh*-phrase is fronted with a cleft form and the complementizer is realized as *que*. I assume here with Kato (2013) that the *wh*-formation involves a cleft formation, as in (61a).[17]

[17] Kato (2013) assumes that the cleft formation is involved in *wh*-constructions in BP. See also (64) below.

(61) a. O que é que ele disse?
DEF that is that he said
'What did he say?'

b. Quem viu João
who saw John
'Who saw John?'

In (61b), the *wh*-phrase is fronted without a complementizer being realized. Notice that the asymmetry with the *wh*-subject sentence in European Portuguese is not observed in Brazilian Portuguese in §3.7.3.

Table 4.12 below shows a summary of the complementizer systems of Wolof, CVC, and European Portuguese (I will limit myself to focusing on the complementizer agreement system, not the entire left periphery system).

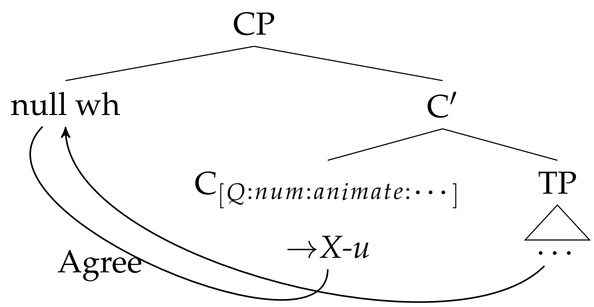
	Wolof	CVC	European Portuguese
Wh-fronting with a cleft form	yes	no	yes
overt Wh-movement and complementizer agreement	no	yes	only for wh-object
An agreed complementizer form	*k-u*/*l-u*, etc.	*ki*	*que*
Agreement optionality	yes (for embedded clauses)	no	no

Table 4.12: The summary of the complementizer agreement system in Wolof, CVC, and European Portuguese

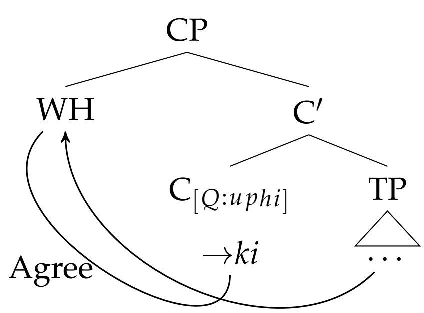
In Wolof, there are two types of *wh*-constructions, one with *wh*-fronting in a cleft form and the other with the null *wh*-phrase fronted to [spec,CP] where the *-u* form agrees with it. The form of *-u* depends on the type of the null *wh*-phrase. In CVC, however, there is no such system; whenever a *wh*-phrase is fronted to [spec,CP], it has to agree with the complementizer, which is always realized as *ki*. The other complementizer (*ma*) cannot co-occur with *ki*, which is distinct from Wolof where the force complementizer *ne* can co-occur with the complementizer *k-u* or *l-a* when the null *wh*-phrase is fronted in (55). European Portuguese shows *wh*-fronting with a cleft form and the complementizer is realized as *que*.

The proposed structures for the complementizer systems across Wolof, CVC and European Portuguese are as follows.

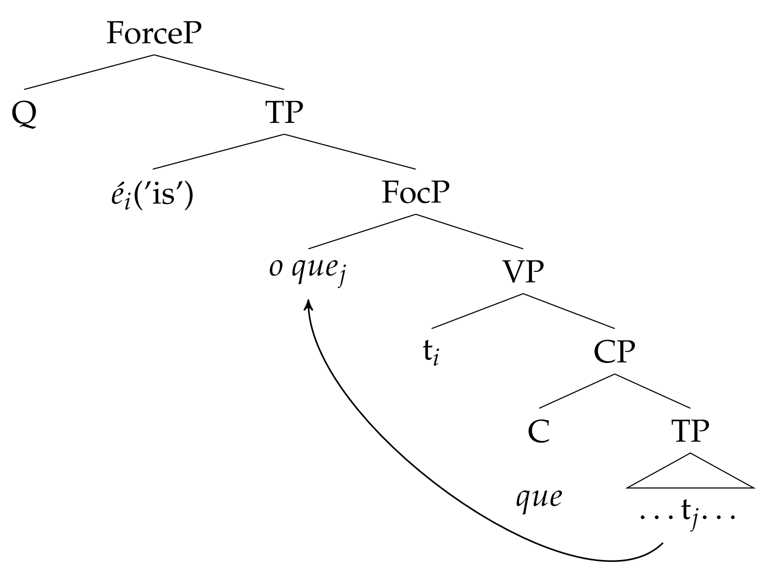
(62) Wolof



(63) CVC



(64) European Portuguese



As shown (62), in Wolof, C has more specific features in terms of the number-feature and the animate-feature, and the morphological realization of the complementizer depends on the type of null *wh*-phrase (and where the *wh*-phrase is base-generated) where X in X-*u* represents the variable.[18] In CVC (63), there is less specification of the features on C since the morphological realization of the complementizer is always *ki* whenever an overt *wh*-phrase is in its [spec,CP], though the ϕ-features still have to agree with the *wh*-phrase (see also §3.7.2). In European Portuguese (64), I assume with Kato (2013) that *wh*-fronting with *que* is a cleft formation where the *wh*-phrase moves to [spec, FocP] (a part of the left periphery, Rizzi 1997). In this case, it is not clear whether there is an element that agrees with *que*.[19]

The summary in Table 4.12 and the syntactic structures proposed above clearly suggest that the CVC complementizer system does not come from Wolof or European Portuguese. In Aboh's feature recombination approach, the expectation may be that, for example, the syntactic features are from Wolof or European Portuguese, but as already discussed, this is clearly not the case. Here I argue that feature recombination takes place on the C head, and CVC develops its own unique complementizer agreement system. Complementizer agreement also present in Wolof, but in CVC an overt *wh*-phrase has to be in [spec,CP] to agree with the complementizer *ki*.

To summarize this section, I explored different syntactic domains (the TMA system, nominal structures, and the CP domain) in Creoles and illustrated that functional features can be recombined and can have novel structures.

The next section discusses the implications of the proposed model.

[18]For expository reasons, the number-feature and animate-feature are in the representation (62).

[19]Since my purpose in this section is not to come up with an analysis of the *wh*-construction in European Portuguese, I leave this issue for my future research.

4.8 Theoretical Implications

In the previous section, I argued that functional features are recombined and the novel functional features and their exponents are realized in Creoles under the framework of the late-insertion-based exoskeletal model. In this section, I will discuss the theoretical implications of the proposed model.

As discussed in the previous chapter, some aspects of the linguistic variation can be best captured by different rule ordering in syntax, not by parameters. The theoretical implication is significant if we can explain this variation without relying on parameters, as it contributes to ultimate goal of the strong minimalist thesis (SMT), the idea that language is an optimal solution to the interfaces (cf. Chomsky, 2000). SMT itself implies that the derivation in the narrow syntax should be uniform (cf. Boeckx 2011; Miyagawa 2010 for the stronger version of SMT), though there is good reason to assume that rule ordering is not fixed in the narrow syntax. In other words, why should we assume that rule ordering in the narrow syntax is the same across languages? Rule ordering needs to be underspecified, otherwise the narrow syntax could be more complicated than the simplest-Merge model, or it could involve look-ahead (i.e., the syntax already knows what it has to do before the derivation).

Within this underspecified syntax framework, it is also natural to assume underspecification of the features on functional categories. For example, Roberts (2019) and his collaborators suggest that formal features on heads are underspecified and as acquirers learn languages, features will be added to the heads. The approach in this chapter is very similar to Roberts'. The functional features on the functional categories are not fixed; instead, features are selected and recombined, and this derives the novel features available in Creoles, which is another way of deriving linguistic variation.

Notice that the structure of Creole languages does not have to show the recombination of functional categories. Language mixing could have "hybrid functional categories" in principle, and the genesis of Creole languages that typically occurs in multilingual settings could result in language mixing structures. I focus on the process of Creole formation rather than on the Creole languages themselves. This model provides the hypothetical grammatical patterns and restricts the range of possible patterns. Also note that labels such as 'language mixing' or 'pidgins/Creoles' do not matter in this model, since it assumes the null theory approach that stipulates no differences between monolingual or multilingual/hybrid grammars (Grimstad et al., 2018; Riksem et al., 2019) (see §4.5).

To sum up, one of the key implications of this model is that there is a possibility that linguistic variation can be captured by feature recombination. In the minimalist literature (Chomsky, 1995), functional features (e.g., inflectional features) are the locus of linguistic variation (*aka* the Borer-Chomsky Conjecture (Baker, 2008)). This suggests that feature recombination can result in linguistic variation.

Crucially, the model makes predictions about the range of possible and impossible grammars among natural languages. In this sense, the proposed structures capture the fact that the Creole languages are natural languages, in addition to being clear representatives of language mixing.

4.9 Conclusion

In this chapter, I adopted a late-insertion-based exoskeletal model that captures multilingual data to show the hybrid nature of Creole languages. The important part of the model is the restriction on the functional exponents, namely the Subset Principle. In monolingual grammars, the specification of the functional features is given in the lexicon (more precisely, a list that includes functional features, categorizers, and roots) and the functional exponent is inserted at Spell-out.

Under this model, I propose that functional features are assembled from multiple sources, not selected from just one source. Language mixing patterns in American Norwegian suggest that the functional projection could be either English or Norwegian, but once the structure is formed based on the functional features, the functional exponent has to match the functional features (Grimstad et al., 2018; Riksem et al., 2019). However, in Dakkhini, Åfarli & Subbarao (2019) argue that the exponent for the functional features can be novel/reconstituted due to long-term language contact.

Considering Åfarli & Subbarao's (2019) data and CVC data, I have argued that feature recombination is the mechanism that forms a novel functional feature, allowing the spell out of a novel functional exponent. This not only captures the Dakkhini data reported in Åfarli & Subbarao (2019) and Aboh's feature recombination analysis, but also captures the novel function of the anterior marker *-ba* in CVC. This anterior marker has a novel feature, pluperfect, which does not exist as such in the forms of *-va* in Portuguese and *ba* in Manjako. I argued that this novel feature is due to feature recombination, in that functional features are re-combined in a novel way, which is different from Aboh's feature recombination analysis, where formal features such as phonological, syntactic, and semantic features are traced back to specific source languages. Based on our proposal, the structure of the functional feature pluperfect (affecting non-stative verbs) in CVC is a novel functional projection (FP). The nominal structure of Saramaccan and the complementizer system in CVC were also examined and I have shown that their structures are not simply transmitted from the substrates or the superstrates; by virtue of recombination of the functional features, novel features can arise. Again, this cannot be fully captured by Aboh's (2009; 2015) version of feature recombination.

My proposal suggests that functional features on functional heads are not grouped together and fixed in the lexicon, but rather they are recombined. As a result of the recombination of functional features, we suggest that the insertion restriction (i.e., the Subset Principle, Halle 1997) of the functional features becomes a new criterion, so the novel functional exponent becomes easy to insert at Spell-out, although the functional exponent tends to come from different source languages, due to language contact. Based on this model, the executive functions in Aboh (2020) do not have any role to play (see §4.3.1) in the syntax, hence we eliminate it from the grammatical model (notice again that our proposed model follows the null theory approach.).

In a broader sense, I proposed a more desirable model based on Obata et al. (2015) and a late-insertion-based exoskeletal model (cf. Grimstad et al. 2018; Riksem et al. 2019), with two underspecifications: (i) rule ordering of operations and (ii) the lexicon. In this model, the Merge operation is the fundamental operation that is powerful enough to characterize the fundamental properties of human language (including language mixing and Creole languages) and linguistic variation.

CHAPTER V

General Conclusions

In this thesis, I have explored multiple ways of deriving possible patterns of functional feature combination, (im)possible derivations in the narrow syntax, and a rule-ordering approach to linguistic variation within a restrictive theory of grammar (i.e., MERGE framework). This restrictive theory is based on MERGE, the

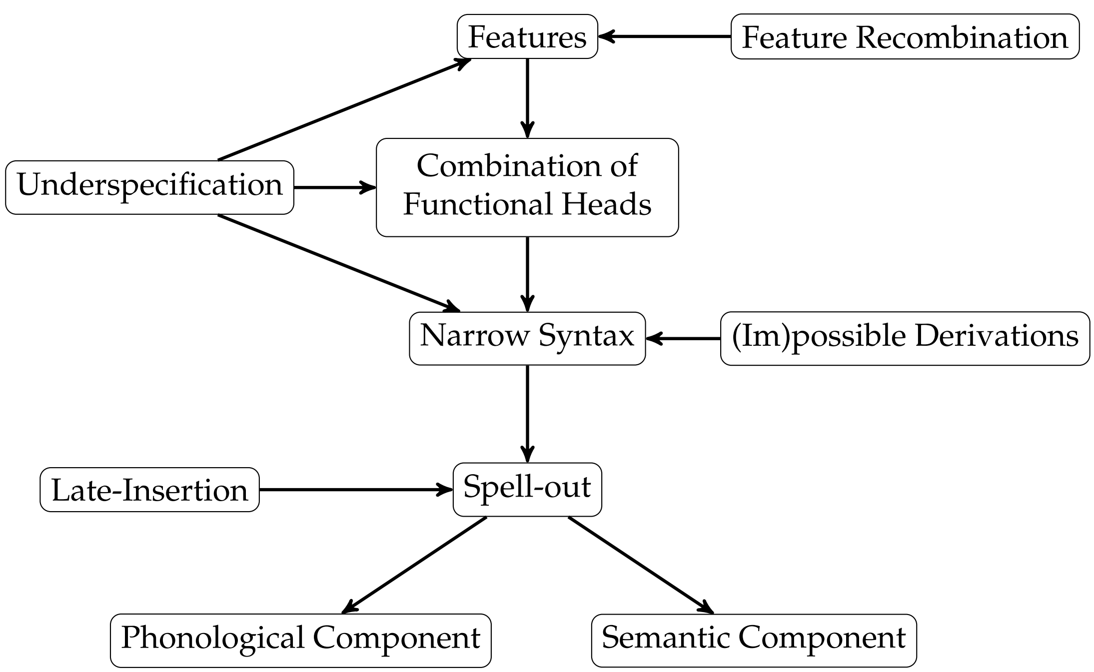


Figure 5.1: The proposed grammatical model

structure-building operation in syntax. It is impossible to discuss linguistic variation without a restricted grammatical theory. To explain the linguistic variation, I adopted the MERGE framework to restrict what is (im)possible.

A central idea in this thesis is underspecification under the restrictive theory. The faculty of language provides and restricts what we can or cannot do in the narrow syntax, but the more specific implementation of feature combinations or derivational paths (e.g. rule ordering) is not fixed within a given language. This is desirable within the current minimalist program in that the locus of the linguistic variation is not (macro-)parameters; rather, the interaction between the possible combinations of features and the possible rule orderings results in different optimal derivations, which can derive linguistic variation. This approach does not have to introduce extra notions such as parameters, but rather each optimal derivation is subject to third factor principles.

In chapter 2, I first pointed out that the proposal in Saito (2016, 2018) has technical problems in terms of the labeling theory in Japanese. Instead of assuming his analysis, I proposed an alternative approach to labeling. One consequence is that what Chomsky (2013) considers a problem regarding subject-predicate labeling is not a problem even in English. Furthermore, I also argued that movement constraints on scrambling in Japanese are not solely governed by labeling.

Instead of seeking an explanation of constraints on movement in terms of labeling, I discussed how Chomsky's (2021a; 2021b) MERGE works and how it gives an account of different kinds of movement. More specifically, I discussed Epstein, Kitahara & Seely's (2018a) analysis, and I analyzed remnant movement in Japanese and German. I showed that improper movement is ruled out because the A and A′-systems do not interact with each other, but they interact with theta theory. However, MERGE can derive other configurations that have been argued to be ungrammatical. The subject island, an instance of the freezing effect, should arguably not be the result of a syntactic principle, given that MERGE can generate the derivations, suggesting that extra-syntactic approaches could account for the (un)acceptability of the sentences that involve the freezing effect configuration. In a similar vein, I argued that adjunct islands are also generable by the general structure-building operation (i.e., *Form Set*).

In chapter 3, based on Obata et al. (2015), I adopted the notion of underspecification of rule ordering in the narrow syntax, and discussed how it can derive some linguistic variation cases involving complementizers. I pointed out that the derivations in Obata et al. (2015) for CVC leave open questions regarding cyclicity, movement and case-valuation of the subject. Based on Epstein et al. (2018b); Chomsky (2017b), I showed the possible complementizer agreement patterns in HC, CVC, and BP.

I also proposed the underspecification of rule application in terms of Merge. I focused on the functional categories and explored the possible combinations of the functional categories among English, German, and Japanese. I specifically proposed that v^* and T become an amalgam in German and v^*, T, and C become an amalgam in Japanese. On the other hand, English functional categories are introduced to the narrow syntax separately. The consequence is that if the functional head is introduced as a free-standing head, the specifier is available, for example, for spec-head agreement, whereas if the functional head is introduced as an amalgam, the specifier is not available. As a result, English, German, and Japanese have systematic patterns regarding the correlation between the absence of amalgamation and the availability of the specifier, which derives a new cluster effect.

In chapter 4, I proposed that formal features can be recombined and form a new syntactic spine, which is one locus of Creole genesis. I have adopted the late-insertion-based exoskeletal model of Grimstad et al. (2018); Riksem et al. (2019) to capture the emergence of novel functional features. In his insightful work, Aboh (2009, 2015) proposes that feature recombination takes place in the I-language, though the details of that proposal are not satisfactory when it comes to the anterior marker *-ba* in CVC, nominal structures in Saramaccan, or the complementizer *ki* in CVC. This is because Aboh's (2009; 2015) analysis treats only the combination of the superstrate and the substrates in terms of phonological, syntactic, and semantic features, in that each formal feature is transmitted by the source languages. However, the anterior marker *-ba* in CVC, the plural determiner in Saramaccan, and the complementizer *ki* have novel features, which is hard to explain in Aboh's (2009; 2015) feature recombination analysis. To solve this problem, I proposed that the functional features are underspecified in UG and they will be acquired how primitive features are assembled to form functional features through language acquisition or language contact. Based on the late-insertion-based exoskeletal model (Grimstad et al., 2018; Riksem et al., 2019), the functional exponent is subject to the Subset Principle (Halle, 1997) that restricts the possible exponents that can be inserted. Thus, the functional exponent has to match its feature to the functional features. I proposed that the recombined functional features have novel instructions for the exponent, thus, a novel functional exponent could fit into the relevant position. Notice that the novel exponents tend to derive from one of the source languages, but my proposal allows completely novel exponents to be inserted as well.

The proposal that bundles of functional features in the lexicon are not fixed in the UG, but rather underspecified, implies that potential recombination of the functional features can yield aspects of linguistic variation, which is compatible with the Borer-Chomsky conjecture (the locus of the linguistic variation is the lexicon). As is discussed in this thesis, a third factor principle approach is adopted, which is compatible with the Strong Minimalist Thesis (SMT). How to derive the linguistic variation in minimalism is still under debate, but I hope that what I presented in this thesis is on the right track, at least under the spirit of SMT (of course, each specific proposal and analysis can be further improved in future research).

BIBLIOGRAPHY

Abels, Klaus & Ad Neelman. 2012. Linear asymmetries and the LCA. *Syntax* 15(1). 25–74.

Aboh, Enoch Oladé. 2009. Competition and selection. In *Complex processes in new languages*, 113–170. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Aboh, Enoch Oladé. 2015. *The emergence of hybrid grammars: Language contact and change*. Cambridge: Cambridge University Press.

Aboh, Enoch Oladé. 2019. Our creolized tongues. In Edit Doron, Malka Rappaport Hovav, Yael Reshef & Moshe Taube (eds.), *Language contact, continuity and change in the genesis of modern Hebrew*, 287–320. Amsterdam/Philadelphia: John Benjamins.

Aboh, Enoch Oladé. 2020. Lessons from neuro-(a)-typical brains: Universal multilingualism, code-mixing, recombination, and executive functions. *Frontiers in Psychology* 11. 488. doi:10.3389/fpsyg.2020.00488.

Abutalebi, Dr Jubin & David W. Green. 2008. Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes* 23(4). 557–582. doi:10.1080/01690960801920602. `https://doi.org/10.1080/01690960801920602`.

Agbayani, Brian, Chris Golston & Toru Ishii. 2015. Syntactic and prosodic scrambling in japanese. *Natural Language and Linguistic Theory* 33. 47–77.

Alexandre, Nélia. 2012. *The defective copy theory of movement: Evidence from wh-constructions in Cape Verdean Creole*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Alexiadou, Artemis, & Elena Anagnostopoulou. 2001. The subject-in-situ generalization and the role of case in driving computations. *Linguistic Inquiry* 32(2). 193–231.

Alexiadou, Artemis, Hagit Borer & Florian Schaäfer. 2014. *The syntax of roots and the roots of syntax*. Oxford: Oxford University Press.

Alexiadou, Artemis & Terje Lohndal. 2017. The structural configurations of root categorization. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 203–231. Berlin, Boston: De Gruyter.

Alexiadou, Artemis & Terje Lohndal. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* 9. doi:10.3389/fpsyg.2018.01719. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01719.

Arad, Maya. 2005. *Roots and patterns: Hebrew morpho-syntax*. Dordrecht: Springer.

Bach, Emmon. 1971. Questions. *Linguistic Inquiry* 2(2). 153–166.

Baker, Mark. 2008. The macroparameters in a microparamatric world. In Theresa Biberauer (ed.), *The limits of syntactic variation*, 351–373. Amsterdam: John Benjamins.

Baker, Philip. 1994. Creativity in creole genesis. In Dany Adone & Ingo Plag (eds.), *Creolization and language change*, 65–84. Berlin, New York: Max Niemeyer Verlag. doi:doi:10.1515/9783111339801.65. https://doi.org/10.1515/9783111339801.65.

Baptista, Marlyse. 2002. *The syntax of Cape Verdean Creole*. Amsterdam: John Benjamins.

Baptista, Marlyse. 2006. When substrates meet superstrate: The case of Cape Verdean Creole. In Jürgen Lang, John Holm, Jean-Louis Rougé & Maria João Soares (eds.), *Cabo Verde - origens dasua sociedade e do seu criouls*, 91–116. Tübingen: Narr.

Baptista, Marlyse. 2020. Competition, selection, and the role of congruence in creole genesis and development. *Language* 96(1). 160–199.

Baptista, Marlyse, Jinho Baik, Ken Kollam & Alton B.H. Worthington. 2018. Modeling dynamic processes in creole genesis. Ms.

Baptista, Marlyse & Miki Obata. 2015. Complementizer-alternation in Creole Languages: New evidence for spec-head agreement. *PAPIA, São Paulo* 25(2). 155–176.

Bayer, Josef & Kornfilt. 1994. Agasint scrambling as an instance of move-alpha. In Norbert Corver & Henk van Riemsdijk (eds.), *Studies on scrambling*, 17–60. Berlin: Mouton de Gruyter.

Berwick, Robert & Noam Chomsky. 2011. The biolinguistic program: the current stat of its evolution and development. In *The biolinguistic enterprise: New perspectives*, .

Berwick, Robert C & Noam Chomsky. 2016. *Why only us: Language and evolution*. Cambridge, MA: MIT Press.

Biberauer, Theresa. 2017. Factor 2 and 3: A principled approach. *Cambridge occasional papers in linguistics* 10. 38–65.

Biberauer, Theresa. 2019. Children always go beyond the input: The maximise minimal means perspective. *Theoretical Linguistics* 45(3-4). 211–224.

Biberauer, Theresa & Ian Roberts. 2015. Rethinking formal hierarchies: A proposed unification. *COPiL* 7. 1–31.

Biberauer, Theresa, Ian Roberts & Michelle Sheehan. 2014. No choice parameters and the limits of syntactic variation. In R. Santana-LaBarge (ed.), *Proceedings of proceedings of the 31st west coast conference on formal linguistics (WCCFL31)*, 46–55. Cascadilla Press.

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma.

Bickerton, Derek. 1984. The language bioprogram hypothesis. *Behavioral and brain sciences* 7. 173–221.

Blümel, Andreas. 2017. Exocentric root declaratives: Evidence from V2. In L Bauke & A Bümel (eds.), *Labels and roots*, De Gruyter.

Blümel, Andreas, Nobu Goto & Yushi Sugimoto. 2021. When the grammar doesn't mind which merge it chooses. To appear in Proceedings of WCCFL39.

Blümel, Andreas, Nobu Goto & Yushi Sugimoto. 2022. Privileges of the rich. Paper presented at PLC45.

Bode, Stefanie. 2020. *Casting a minimalist eye on adjuncts*. New York: Routledge.

Boeckx, Cedric. 2000. Quirky agreement. *Studia Lingusitica* 54(3). 354–380.

Boeckx, Cedric. 2003. *Islands and chains: Resumption as stranding*. Amsterdam: John Benjamins.

Boeckx, Cedric. 2008. *Bare syntax*. Oxford/New York: Oxford University Press.

Boeckx, Cedric. 2011. Approaching parameters from below. In Cedric Boeckx & A.-M. Di Sciullo (eds.), *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, 205–221. Oxford: Oxford University Press.

Boeckx, Cedric. 2012. *Syntactic islands*. Cambridge: Cambridge University Press.

Boeckx, Cedric & Evelina Leivada. 2013. Entangled parametric hierarchies: Problems for an overspecified universal grammar. *PLOS ONE* 8(9). 1–11. doi:10.1371/journal.pone.0072357. `https://doi.org/10.1371/journal.pone.0072357`.

Boeckx, Cedric & Juan Uriagereka. 2007. Minimalism. In Gillian Ramchand & Charles Reiss (ed.), *The oxford handbook of linguistic interfaces*, 541–573. Oxford: Oxford University Press.

Bondevik, Ingrid, Dave Kush & Terje Lohndal. 2020. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics* 1–32. doi:10.1017/S0332586520000207.

Bondevik, Ingrid, Dave Kush & Terje Lohndal. 2021. Variation in adjunct islands: The case of norwegian. *Nordic Journal of Linguistics* 44(3). 223–254. doi:10.1017/S0332586520000207.

Borer, Hagit. 1984. *Parametric syntax: Case studies in semitic and romance languages*. Berlin, Boston: Mouton De Gruyter.

Borer, Hagit. 2003. Exo-skeletal vs. endo-skeletal explanations: syntactic projections and the lexcion. In *The nature of explanation in linguistic theory*, 31–67. Stanford, CA: CSLI Publications.

Borer, Hagit. 2005a. *Structuring sense: In name only*. Oxford: Oxford University Press.

Borer, Hagit. 2005b. *Structuring sense: The normal course of events*. Oxford: Oxford University Press.

Borer, Hagit. 2013. *Structuring sense: Taking form*. Oxford: Oxford University Press.

Borer, Hagit. 2017. The generative word. In *The cambridge companion to Chomsky*, 110–133. Camrbidge: Cambridge University Press.

Borgonovo, Claudia & Ad Neeleman. 2000. Transparent adjuncts. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 45(3-4). 199–224.

Bošković, Željko. 2002. A-movement and the EPP. *Syntax* 5(3). 167–218.

Bošković, Željko. 2007. On the locality and motivation of move and agree: An even more minimal theory. *Linguistic Inquiry* 38(4). 589–644.

Bošković, Željko. 2016. On the timing of labeling: Deducing comp-trace effects, the subject condition, the adjunct condition, and tucking in from labeling. *The Linguistic Review* 33(1). 17–66.

Bošković, Željko. 2020. On the coordinate structure constraint, across-the-board-movement, phases, and labeling. In Jeroen van Craenenbroeck an Cora Pots & Tanja Temmerman (eds.), *Recent developments in phase theory*, 133–182. Berlin, Boston: De Gruyter Mouton.

Broekhuis, H. 2006. Extraction from subjects: Some remarks on chomsky's 'on phases'. In H. Broekhuis, N. Corver & R. Huybregts (eds.), *Organizing grammar*, 59–68. Berlin//Boston: de Gruyter Mouton.

Brown, Jessica M.M. 2017. *Heads and adjuncts: An experimental study of subextraction from participials and coordination in english, german and norwegian*: University of Cambridge dissertation.

Bruening, Benjamin. 2002. Raising to object and proper movement. Ms, University of Delaware.

Bruening, Benjamin. 2010. Language-particular syntactic rules and constraints: English locative inversion and do-support. *Language* 86. 43–84.

Brásio, P.A.D. 1962. *Descobrimento, povoamento, evangelização do arquipélago de cabo verde*. Lisboa: Centro de Estudos Históricos Ultramarinos.

Carstens, Vicki. 2003. Rethinking complementizer agreemnet: Agree with a case-checked goal. *Linguistic Inquiry* 34. 393–412.

Carstens, Vicki. 2005. Agree and EPP in Bantu. *Natural Language & Linguistic Theory* 23(2). 219–279.

Carstens, Vicki & Michale Diercks. 2010. Parameterizing case and activity: Hyper-raising in Bantu. In *Proceedings of NELS40*, GLSA Publiscations.

Cattell, Ray. 1976. Constraints on movement rules. *Language* 52(1). 18–50.

Chaves, Rui. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 49(2). 285–327.

Chaves, Rui & Jeruen Dery. 2014. Which subject island will the acceptability of improve with repeated exposure? In Robert E. Santana-LaBarge (ed.), *The 31st west coast conference on formal linguistics (WCCFL31)*, 96–106. Cascadilla Proceedings Projects.

Chaves, Rui & Michael T. Putnam. 2020. *Unbounded dependnecy constructions: Theoretical and experimental perspectives*. Oxford: Oxford University Press.

Chomsky, Noam. 1955/1975. The logical structure of linguistic theory. Ms, Harvard University, 1955. Published in part in 1975, New York: Plenum.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1973. Conditions on transformations. In S.Anderson nd P. Kiparsky (ed.), *A festschrift for morris halle*, 232–286. New York: Holt, Rinehart, and Winston.

Chomsky, Noam. 1977. On wh-movement. In P.Culicover, T. Wasow & A.Akmajian (eds.), *Formal syntax*, 71–132. New York: Academic Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht, Holland/Cinnaminson, USA: Foris Publication.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In D. Michaels R. Martin & J. Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of howard lasnik*, chap. 3, 89–156. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and beyond: The cartography of syntactic structures*, 1–52. Oxford: Oxford University Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1). 1–22.

Chomsky, Noam. 2007. Approaching UG from below. In U Saureland & H.-M Gärtner (eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from semantics*, 1–29. Berlin and New York: Mouton de Gruyter.

Chomsky, Noam. 2008. On phases. In Robert Fredin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory*, 133–166. The MIT Press.

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

Chomsky, Noam. 2015. Problems of projection extensions. In Simona Matteini Elisa Di Domenico, Cornelia Hamann (ed.), *Structures, strategies and beyond: Studies in honour of adriana belletti*, 3–16. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Chomsky, Noam. 2017a. The galilean challenge. *Inference* 3(1). `https://inference-review.com/article/the-galilean-challenge`.

Chomsky, Noam. 2017b. The language capacity: architecture and evolution. *Psychonomic Bulletin & Review* 24(1). 200–203.

Chomsky, Noam. 2017c. Lecture at University of Arizona.

Chomsky, Noam. 2017d. Lecture at University of Reading.

Chomsky, Noam. 2019a. Lecture at MIT. April10,April12, 2019.

Chomsky, Noam. 2019b. Lecture at UCLA. April29, April30, May1, May2, 2019, transcript available at lingbuzz.

Chomsky, Noam. 2019c. Some puzzling foundational issues: The reading program. *Catalan Journal of Linguistics* Special Issue. 263–285.

Chomsky, Noam. 2020. puzzles about phases. In Ludovico Franco & Paolo Lorusso (eds.), *Linguistic variation: Structure and interpretation*, 163–168. Berlin, Boston: De Gruyter Mouton. doi:https://doi.org/10.1515/9781501505201.

Chomsky, Noam. 2021a. Genuine explanations. talk at the 39th of West Coast Conference on Formal Linguistics (WCCFL39), University of Arizona, April 9.

Chomsky, Noam. 2021b. Minimalism: Where are we now, and where can we hope to go. *Gengo Kenkyu* 160. 1–41.

Chomsky, Noam, Ángel J. Gallego & Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics: Special Issue* 229–261.

Chomsky, Noam & George Miller. 1963. Introduction to the formal analysis of natural languages. In R.D.Luce, R.R.Bush & G. Eugene (eds.), *Handbook of mathematical psychology 2*, 269–321. New York: Wiley & Sons.

Citko, Barbara. 2005. On the nature of merge: External merge, internal merge, and parallel merge. *Linguistic Inquiry* 36(4). 475–496.

Citko, Barbara. 2011. *Symmetry in syntax*. Cambridge: Cambridge University Press.

Citko, Barbara & Martina Gračanin-Yuksek. 2021. *Merge:binarity in (multidominant) syntax*. Cambridge, MA: MIT Press.

Collins, Chris. 2002. Eliminating labels. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivational and explanation in the minimalist program*, 42–64. Malden, MA/Oxford: Blackwell.

Collins, Chris. 2017. Merge(X,Y)={X,Y}. In Leah Bauke & Andreas Blümel (eds.), *Labels and Roots*, 47–68. Berlin and Boston: De Gruyter.

Collins, Chris & T. Daniel Seely. 2020. Labeling without labels. Ms, New York University and Eastern Michigan University, available at lingubuzz.

Corver, Norbert. 2017. Freezing effects. In Martin Everaert & Henk C. van Riemsjik (eds.), *The wiley blackwell companion to syntax. second edition*, John Wiley & Cons, Inc. doi:10.1002/9781118358733.wbsyncom055.

van Craenebroeck, Jeroen. 2017. VP-ellipsis. In Martin Everaert & Henk C. van Riemsjik (eds.), *The wiley blackwell companion to syntax. second edition*, John Wiley & Cons, Inc. doi:https://doi.org/10.1002/9781118358733.wbsyncom049.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Culicover, Peter W. & Susanne Winkler. 2022. Parasitic gaps aren't parasitic, or, the case of the univited guest. *The Linguistic Review* 39(1). 1–35.

Dadan, Marcin R. 2019. *Head labeling preference and language change*: University of Connecticut, Storrs dissertation.

den Dikken, Marcel. 2012. The phase impenetrability condition and successive cyclicity: A reconsideration. Ms, CUNY.

Duguine, Maia, Aritz Irurtzun & Cedric Boeckx. 2018. Linguistic diversity and granularity: Two case-studies against parametric approaches. *Linguistic Analysis* 43(3-4). 445–474.

Embick, David. 2015. *The morpheme: A theoretical introduction*. Berlin and Boston: De Gruyter Mouton.

Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax-morphology interface. In Gillian Ramchand & Charles Reiss (eds.), *The oxford handbook of linguistic interfaces*, 289–324. Oxford: Oxford University Press.

Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2012. Structure building that can't be. In M. Uribe-etxebarria & V. Valmala (eds.), *Ways of structure building*, 253–270. Oxford: Oxford University Press.

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

Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2015. From *aspects* 'daughterless mothers' (aka delta nodes) to POP's 'motherless'-set (aka non-projection): A selective history of the evolution of simplest merge. In Ángel Gallego & Dennis Ott (eds.), *50 years later: Reflections on chomsky's aspects*, 99–112. MIT working papers in linguistics 77.

Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2016. Phase cancellation by external pair-merge of heads. *The Linguistic Review* 33(1). 87–102.

Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2018a. Can "determinacy + PIC" explain descriptions of remnant movement asymmetries? Paper presented at the 157th meeting of the linguistic society of Japanese, November 17, 2018.

Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2020. Unifying labeling under minimal search in "single-" and "multiple-specifier" configurations. In *Coyote papers: Working papers in linguistics, linguistic theory at the university of arizona*, University of Arizona Linguistics Circle.

Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2021. *A minimalist theory of simplest merge*. New York: Routledge.

Epstein, Samuel David, Miki Obata & T. Daniel Seely. 2018b. Is linguistic variation entirely linguistic? *Linguistic Analysis* 41(3-4). 481–516.

Epstein, Samuel David & T. Daniel Seely. 2002. Rule applications as cycles in a level-free syntax. In *Derivation and explanation in the minimalist program*, 65–89. Malden, MA: Blackwell.

Epstein, Samuel David & Jae-Young Shim. 2015. Two notes on possible approaches to the unification of theta relations. *Linguistic Analysis* 40(1-2). 3–20.

Ernst, Thomas. 2022. The adjunct condition and the nature of adjuncts. *The Linguistic Review* doi:https://doi.org/10.1515/tlr-2021-2082.

Fanselow, Gisbert & Damir Ćavar. 2002. Distributed deletion. In *Theoretical approaches to universals*, 65–107. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Fiengo, Robert. 1977. On trace theory. *Linguistic Inquiry* 8(1). 35–61.

Fleisher, Nicholas. 2013. On the absence of scope reconstruction in tough subject A-chains. *Linguistic Inquiry* 44(2). 321–332.

Fong, Sandiway, Robert Berwick & Jason Ginsburg. 2019. The combinatorics ofmerge and workspace right-sizing. Paper presented at Evolinguistics Workshop, May 25-26, 2019.

Fong, Suzana. 2018. Featural definition of syntactic positions: Evidence from hyper-raising. *Proceedings of Lingusitic Society of America* 3(1). 1–15.

Fortuny, Jordi. 2008. *The emergence of order in syntax*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Fox, Danny. 1999. Reconstruction, binding theory, and the interpretation of chains. *Linguistic Inquiry* 30(2). 157–196.

Fox, Danny & David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theotical Linguistics* 31. 1–45.

Freidin, Robert. 2021. The strong minimalist thesis. *Philosophies* 1. 97. https://doi.org/10.3390/philosophies6040097.

Fujita, Koji. 2007. Facing the logical problem of language evolution. *English Linguistics* 24(1). 78–108.

Fujita, Koji. 2009. A prospect for evolutionary adequacy: Merge and the evolution and development of human language. *Biolinguistics* 3(2-3). 128–153.

Fukui, Naoki. 1986. *A theory of category projection and its application*: MIT dissertation. Published with revisions as Fukui (1995b).

Fukui, Naoki. 1988. Deriving the differences between English and Japanese: A case study in parametric syntax. *English Linguistics* 5. 249–270.

Fukui, Naoki. 1991. Strong and weak barriers: remarks on the proper characterization of barriers. In Heizo Nakajima (ed.), *Current english linguistic in Japan*, 77–93. Berlin/New York: Mouton de Gruyter. Reprinted in Fukui (2006).

Fukui, Naoki. 1993a. A note on improper movement. *The Linguistic Review* 10. 111–126.

Fukui, Naoki. 1993b. Parameters and optionality. *Linguistic Inquiry* 24(3). 399–420.

Fukui, Naoki. 1995a. The principles-and-parameters approaches: a comparative syntax of english and japanese. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 327–371. Reprinted in Fukui (2006).

Fukui, Naoki. 1995b. *Theory of projection in syntax*. Stanford, California and Tokyo: CSLI Publications and Kurosio Publishers.

Fukui, Naoki. 2001. Phrase structure. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 374–406. Malden,MA/Oxford: Blackwell.

Fukui, Naoki. 2006. *Theoretical comparative syntax: Studies in marcoparameters*. London and New York: Routledge.

Fukui, Naoki & Hiromu Kasai. 2004. Spelling-out scrambling. In Pierre Pica (ed.), *Linguistic variation yearbook 4*, 109–141. Amsterdam/Philadelphia: John Benjamins.

Fukui, Naoki & Hiroki Narita. 2014. Merge, labeling, and projection. In Andres Carnie, Yosuke Sato & Daniel Siddiqi (eds.), *The routledge handbook of syntax*, 3–23. Routledge.

Fukui, Naoki & Hiromu Sakai. 2003. The visibility guideline for functional categories: verb raising in japanese and related issues. *Lingua* 113. 321–375.

Fukui, Naoki & Margaret Speas. 1986. Specifiers and projection. In N. Fukui, T. Rapoport & E. Sagey (eds.), *MIT working papers in linguistics: Papers in theoretical linguistics 8*, 128–172. Reprinted in Fukui (2006).

Fukui, Naoki & Yuji Takano. 1998. Symmetry in syntax: Merge and demerge. *Journal of East Asian Linguistics* 7. 27–86.

Funakoshi, Kenshi. 2009. Theta-marking via agree. *English Linguistics* 26(1). 1–32.

Funakoshi, Kenshi. 2020. Verb-raising and VP-fronting in Japanese. *The Linguistic Review* 37(1). 117–146.

Gallego, Ángel & Juan Uriagereka. 2007. Sub-extraction from subjects: A phase theory account. In José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez & María José Cabrera (eds.), *Romance linguistics 2006: Selected papers from* *the 36th linguistic symposium on romance languages (lsrl), new brunswick, march-april 2006*, 146–162. Amsterdam/Philadelphia: John Benjamins Publishing Company.

van Gelderen, Elly. 2018. Problems of projection: The role of language change in labeling paradoxes. *Studia Linguistica* 72(1). 113–127. doi:https://doi.org/10.1111/stul.12041. https://onlinelibrary.wiley.com/doi/abs/10.1111/stul.12041.

Georgi, Doreen. 2013. Opaque interaction of internal merge and agree. In Fabian Heck & Anke Assmann (eds.), *Rule interaction in grammar*, 413–461. Linguistische Arbeits Berichte 90, UNiversität Leipzig 2013.

Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7(5). 219 – 224.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goto, Nobu. 2011a. *Feature-ineritance: Its effects on agree, move, and delete*: Tohoku Gakuin University dissertation.

Goto, Nobu. 2011b. Interplay between agree and move. In *Tohoku:essays and studies in english language and literature 44*, 39–68. Tohoku Gakuin University.

Goto, Nobu. 2013a. Labeling and scrambling in Japanese. *Tohoku:Essay and Studies in English Language and Literature* 46. 39–73.

Goto, Nobu. 2013b. Toward unconstrained merge. In Il-Jae Lee & Uujinbai Dolgormaaa (eds.), *Proceedings of the 15th seoul international conference on generative grammar (sicogg15): Universals and parameters*, 91–110. Seoul, Korea: Hankuk Publishing Co.

Goto, Nobu. 2016. Labelability=extractability: Its theoretical implications for the free. In *Proc. nels46*, 335–248.

Goto, Nobu. 2017. Eliminating the strong/weak parameter on T. In Michael Ypshitaka Erlewine (ed.), *Proccedings of GLOW in Asia xi, volume 2*, 57–71. MITWPL.

Goto, Nobu & Toru Ishii. 2019. Determinacy theory of movement. Paper presented at NELS50.

Goto, Nobu & Toru Ishii. 2021. Multiple nominative and form sequence. Ms, available at Lingbuzz.

Grano, Thomas & Howard Lasnik. 2018. How to neutralize a finite clause boundary: Phase theory and the grammar of bound pronouns. *Linguistic Inquiry* 49(3). 465–499.

Green, David & Jubin Abutalebi. 2008. Understanding the link between bilingual aphasia and language control. *Journal of Neurolinguistics* 21. 558–576. doi:10.1016/j.jneuroling.2008.01.002.

Grewendorf, Günther. 2015. Problems of remnant movement. In Günther Grewendorf (ed.), *Remnant movement*, 3–34. Berlin, München, Boston: De Gruyter Mouton.

Grewendorf, Günther & Joachim Sabel. 1999. Scrambling in german and japanese: Adjunction versus multiple specifiers. *Natural Language & Linguistic Theory* 17(1). 1–65.

Grimstad, Maren Berge, Brita Ramsevik Riksem, Terje Lohndal & Tor A Åfrli. 2018. Lexicalst vs. exoskeletal approaches to langauge mixing. *The Linguistic Review* 35(2). 187–218.

Haegeman, Liliane, Ángel L. Jiménez-Fernández & Andrew Radford. 2015. Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1). 73–150.

Haegeman, Liliane & Marjo van Koppen. 2012. Complementizer agreement and the relation between C^0 and T^0. *Linguistic Inquiry* 43(3). 441–454.

Haider, Hubert. 1988. Matching projections. In *Constituent structure. papers from the 1987 GLOW conferences*, 101–121. Venice.

Haider, Hubert. 2006. Mittelfield phenomena. In Martine Everaert & Hen van Riemsdijk (eds.), *The blackwell companion to syntax, volume 1*, 204–274. Malden, MA/Oxford/Victoria: Blackwell Publishing Ltd.

Halle, Morris. 1997. Distributed morphology: Impoverishment and fission. In *Pf: Papers at the interfaces*, 425–449. MIT Working papers in Linguistics 30.

Halle, Morris & Alec Marantz. 1993. *Distributed morphology and the pieces of inflection* chap. 3, 111–176. Cambridge, MA: MIT Press.

Halpert, Claire. 2016. Raising parameters. In Kyeong min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson & Lisa Shorten (eds.), *Proceedings of WCCFL33*, 186–195. Somerville, MA: Cascadilla Proceedings Project.

Halpert, Claire. 2018. Raising, unphased. *Natural Language & Linguistic Theory* doi:https://doi.org/10.1007/s11049-018-9407-2.

Hartman, Jeremy. 2011. (non-)intervention in A-movement. In *Linguistic variation. vol.11, issue 2*, 121–148. Amsterdam/Philadelphia: John Benjamins.

Haspelmath, Martin. 2008. Parametric versus functional explanations of syntactic universals. In Theresa Biberauer (ed.), *The limits of syntactic variation*, 75–107. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hiraiwa, Ken. 2010. Spelling out the double-o constraint. *Natural Language and Linguistic Theory* 28. 723–770.

Holmberg, Anders. 2002. Expletives and agreement in scandinavian passives. *Journal of Comparative Germanic Linguistics* 4. 85–128.

Holmberg, Anders & Thorbjörg Hróarsdóttir. 2003. Agreement and movement in icelandic raising constructions. *Lingua* 113(10). 997 – 1019.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30(1). 69–96.

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

Huang, C.-T. James. 1982. *Logical relations in chinese and the theory of grammar*: MIT dissertation.

Huang, C.-T. James & Ian Roberts. 2017. Principles and parameters of universal grammar. In Ian Roberts (ed.), *The oxford handbook of universal grammar*, 307–354. Oxford: Oxford University Press.

Ishii, Toru. 1997. *An asymmetry in the composition of phrase structure and its consequences*: University of California, Irvine dissertation.

Julien, Marit. 2002. *Syntactic heads and word formation*. Oxford: Oxford University Press.

Julien, Marit. 2003. Double definiteness in scandinavian. *Nordlyd* 31(1). 230–244.

Julien, Marit. 2005. *Noun phrases from a scandinavian perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kasai, Nironobu. 2004. Two notes on ATB movement. *Language and Linguistics* 5(1). 167–188.

Kato, Mary Aizawa. 2013. Deriving "wh-in-situ" through movement in Brazilian Portuguese. In Victoria Camacho-Taboada, Ángel L. Jiménez-Fernández, Javier Martín-González & Mariano Reyes-Tejedor (eds.), *Information structure and agreement*, 175–192. Amsterdam: John Benjamins Publishing Company.

Kayne, Richar S. 2011. Why are there no directionality parameters? In *Proceedings of the 28th west coast conference on formal linguistics (WCCFL28)*, 1–23. Somerville, MA: Cascadilla Proceedings Project.

Kayne, Richard S. 1976. French relative *que*. In Fritz Hensey & Marta Luján (ed.), *Current studies in romance linguistics*, 255–299. Washington DC: Georgetown University Press.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Ke, Alan Hezao. 2019. *The syntax, semantics and processing of agreement and binding grammatical illusions*: University of Michigan, Ann Arobor dissertation.

Keine, Stefean & Rajesh Bhatt. 2016. Interpreting verb clusters. *Natural Language and Linguistic Theory* 34. 1445–1492.

Kihm, Alain. 1990. Conflation as a directive process in creolization. In *Beiträge zum 4. essener kolloquium*, 111–137. Bochum: Brockmeyer.

Kihm, Alain. 1994. *Kriyol syntax: The portuguese-based creole language of guinea-bisau*. Amsterdam: John Benjamins Publishing Company.

Kikuchi, Akira. 1987. Comparative deletion in japanese. Ms, Yamagata University, Yamagata, Japan.

Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In Emmon Bach & Robert Harms (eds.), *Universals in linguistic theory*, 170–210. New York: Holt.

Kiparsky, Paul. 1971. Historical linguistics. In *A survey of linguistic science*, 576–642. University of Maryland Linguistic Program, College Park.

Kiparsky, Paul. 1976. Abstractness, opacity and global rules. In *Three dimensions of linguistic theory*, 57–86. Tokyo Institute for Advanced Studies of Language.

Kitahara, Hisatsugu. 1997. *Elementary operations and optimal derivations*. MIT Press.

Kitahara, Hisatsugu & T. Daniel Seely. 2021. Structure building under MERGE. Poster presented at WCCFL39.

Koizumi, Masatoshi. 2000. String vacuous overt verb raising. *Journal of East Asian Linguistics* 9. 227–285.

Kouwenberg, Silvia. 2006. L1 transfer and the cut-off point for L2 acquisition processes in creole formation. In Claire Lefebvre, Lydia White & Chirstine Jourdan (eds.), *L2 acquisition and creole genesis: Dialogues*, 205–219. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kuno, Masakazu. 2001. Why does movement sometimes leave a trace and sometimes not? Ms, University of Tokyo.

Kuno, Susumu. 1972. Evidence for subject raising in Japanese. *Papers in Japanese Linguistics* 1. 24–51.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.

Kuno, Susumu. 1976. Subject raising. In *Syntax and semantics 5: Japanese generative grammar*, 17–49. NewYork: Academic Press.

Kural, M. 1993. V-to (I-to)-C in Turkish. In F. Beghelli M. Kural (ed.), *UCLA occasional papers in linguistics 11*, 17–54. Los Angeles: University of California, Los Angeles Department of Linguistics.

Kuroda, S.-Y. 1988. Whether we agree or not. *Linguisticae Investigationes* 12. 1–47.

Lakoff, George. 1986. Frame semantic control and the theory of gramaar. In *Papers from the parasession on pragmatics and grammatical theory*, 152–167. Chicago Linguistic Society.

Lasnik, Howard. 1999. Chains of arguments. In Samuel David Epstein & Norbert Hornstein (eds.), *Working in minimalism*, 185–215. Cambridge, MA: MIT Press.

Lasnik, Howard. 2011. What kind of computing devive is the human language faculty? In Anna Maria Di Sciullo & Cedric Boeckx (eds.), *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, 354–365. Oxford: Oxford University Press.

Lasnik, Howard & Myung-Kwan Park. 2003. The EPP and the subject condition under sluicing. *Linguistic Inquiry* 34(4). 49–660.

Lasnik, Howard & Mamoru Saito. 1992. *Move alpha: Conditions on its application and output*. Cambridge, MA: MIT Press.

Lefebvre, Claire. 1998. *Creole genesis and the acquisition of grammar: The case of haitian creole*. Cambridge: Cambridge University Press.

Lefebvre, Claire. 2013. A comparison of the nominal structures of Saramaccan,Fongbe and English with reference to Haitian Creole: Implications for a relabelling-based account of creole genesis. *Lingua* 129. 32–66.

Lefebvre, Claire. 2015. The nominal structures of Saramaccan, Fongbe and English with reference to Haitian Creole. In Claire Lefebvre (ed.), *Functional categories in three atlantic creoles: Saramaccan, Haitian, and Papiametu*, 17–68. Amsterdam: John Benjamins Publishing Company.

Legate, Julie Anne. 2011. Under-inheritance. Presented at NELS42, University of Toronto.

Liceras, Juana M., Raquel Fernández Fuertes, Susana Perales, Rocío Pérez-Tattam & Kenton Todd Spradlin. 2008. Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua* 118(6). 827 – 851.

Lobeck, Anne. 1990. Functional heads as proper governors. In *Proceedigns of north east linguistics society 20. vol. 20, article 6*, 348–362.

Lohndal, Terje. 2011. Freezing effects and objects. *Journal of Linguistics* 47(1). 163–199.

Lohndal, Terje. 2014. *Phrase structure and argument structure*. Oxford: Oxford University Press.

Lohndal, Terje. 2019. The exoskeletal model. Ms, NTNU.

Longenbaugh, Nicholas. 2017. Composite A/A′-movement. Ms, MIT.

MacSwan, Jeff. 2009. Generative approaches to codeswitching. In B.E.Bullock & A.J.Toribio (eds.), *Cambridge handbook of linguistic codeswitching*, 309–335. Cambridge: Cambridge University Press.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *University of pennsylvania working papers in linguistics*, vol. 4 2, 201–225.

Marantz, Alec. 2013. Verbal argument structure: Events and participants. *Lingua* 130. 152–168. doi:https://doi.org/10.1016/j.lingua.2012.10.012. https://www.sciencedirect.com/science/article/pii/S0024384112002306. SI: Syntax and cognition: core ideas and results in syntax.

May, Robert. 1979. Must COMP-to-COMP movement be stipulated? *Linguistic Inquiry* 10. 719–725.

McInnerney, Andrwe & Yushi Sugimoto. 2022. On dissociating adjunct island and subject island effects. *Proceedings of the LSA* 7(1). 5207. doi:https://doi.org/10.3765/plsa.v7i1.5207.

McWhorter, John & Jeff Good. 2012. *A grammar of Saramaccan creole*. Berlin and Boston: De Gruyter Mounton.

McWhorter, John H. 2002. The rest of the story: restoring pidginization to creole genesis theory. *Journal of Pidgin and Creole Languages* 17(1). 1–48.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and identity in ellipsis*. Oxford: Oxford University Press.

Michel, Dan & Grant Goodall. 2012. Finiteness and the nature of island constraints. In Nobu Goto, Koichi Otaki, Atsushi Sato & Kensuke Takita (eds.), *The proceedings of GLOW in Asia IX*, 187–197.

Miyagawa, Shigeru. 2010. *Why agree? why move?: Unifying agreement-based and discourse-configurational languages*. Cambridge, MA: MIT Press.

Miyagawa, Shigeru. 2017. *Agreement beyond phi*. Cambridge, MA: MIT Press.

Miyagawa, Shigeru, Dafeng Wu & Masatoshi Koizumi. 2019. Inducing and blocking labeling. *Glossa: a journal of general linguistics* 4(1). 141.

Miyamoto, Yoichi. 2012. On transparent adjuncts in Japanese. In M. Uribe-etxebarria & V. Valmala (eds.), *Ways of structure building*, 330–365. Oxford: Oxford University Press.

Moro, Monica Quintanilla. 2014. In *Grammatical theory and bilingual codeswitching*, 213–226. Cambridge,MA: MIT Press.

Mufwene, Salikoko S. 1996. The founder principle in creole genesis. *Diachronica* 13(1). 83–134.

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Mufwene, Salikoko S. 2002. Competition and selection in language evolution. *Selection* 3(1). 45–56.

Mufwene, Salikoko S. 2008. *Language evolution: Contact, competition and change*. London/New York: Continum International Publihsing Group.

Müller, Christiane. 2017. Extraction from adjunct islands in swedish. *Norsk Lingvistisk Tidsskrift* 35(1). 67–85.

Müller, Georeon. 2013. On the order of syntactic operations. Ms, Universität Leipzig, handout, Lecture 1: A Brief History of Rule Interaction in Syntax.

Müller, Gereon. 1996. A constraint on remnant movement. *Natural Language and Linguistic Theory* 14. 335–407.

Müller, Gereon. 1998. *Incomplete category fronting*. Dordrecht: Kluwer.

Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41(1). 35–82.

Muysken, P.C. 1981. Halfway between quechua and spanish: the case for relexification. In A. Highfield (ed.), *Historicity and variation in creole studies*, 52–78. Ann Arbor: Karoma Publisher.

Narita, Hiroki. 2010. The tension between explanatory and biological adequacy: Review of fukui (2006). *Lingua* 120(5). 1313–1323.

Narita, Hiroki. 2011. *Phasing in full interpretation*: Harvard University dissertation.

Narita, Hiroki. 2012. *{XP, YP}, the independence of the LSA and antisymmetry, and the LCA-free phrase-based account of the CED effect. Ms, Waseda Institute for Advanced study.

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

Narita, Hiroki & Koji Fujita. 2010. A naturalist reconstruction of minimalist and evolutionary biolinguistics 4(4). 356–376.

Narita, Hiroki & Naoki Fukui. 2022. *Symmetrizing syntax: Merge, minimality, and equilibria*. London and New York: Rougtledge.

Nemoto, Naoko. 1995. Scrambling in Japanese, AGROP, and economy of derivation. *Lingua* 97. 257–273.

Nevins, Andrew. 2005. Derivations without the activity condition. In Martha McGinnis & Norvin Richards (eds.), *MITWPL49: Perspectives on phases*, 283–306. Cambridge, MA: MIT working papers in linguistics.

Newmeyer, Frederick J. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.

Newmeyer, Frederick J. 2017. Where, if anywhere, are parameters? a critical historical overview of parametric theory. In Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond)*, Berlin: Language Science Press.

Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry* 32(2). 303–344.

Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.

Nunes, Jairo & Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3(1). 20–43.

Obata, Miki. 2010. *Root, successive-cyclic and feature-splitting internal merge: Implications for feature-inheritance and transfer*: University of Michigan dissertation.

Obata, Miki. 2017. Is transfer strong enough to affect labels? In L Bauke & A Bümel (eds.), *Labels and roots*, 117–126. De Gruyter.

Obata, Miki & Samuel Epstein. 2016. Eliminating parameters from the narrow syntax: rule ordering variation by third factor underspecification. In Koji Fujita & Cedric Boeckx (eds.), *Advances in biolinguistics: The human language faculty and its biological basis*, 128–138. London: Routledge.

Obata, Miki, Samuel Epstein & Marlyse Baptista. 2015. Can crosslinguistically variant grammars be formally identical? third factor underspecification and the possible elimination of parameters of ug. *Lingua* 156. 1 – 16.

Obata, Miki & Samuel David Epstein. 2011. Feature-splitting internal merge: Improper movement, intervention, and the A/A′-distinction. *Syntax* 14(2). 122–147.

Oishi, Masayuki. 2015. The hunt for a label. In Hiroki Egashira, Hisatsugu Kitahara, Kazuo Nakazawa, Tadao Nomura, Masayuki Oishi, Akira Saizen & Motoko Suzuki (eds.), *Untiring pursuit of better alternatives*, 322–334. Tokyo: Kaitakusha.

Oku, Satoshi. 1998. *A theory of selection and reconstruction in the minimalist perspective*: University of Connecticut, Storrs dissertation.

Omaki, Akira, Shin Fukuda, Chizuru Nakao & Maria Polinsky. 2020. Subextraction in Japanese and subject-object symmetry. *Natural Language & Linguistic Theory* 38. 627–669.

Ormazabal, Javier, Juan Uriagereka & Myriam Uribe-Etxebarria. 1994. Word order and WH-movement: towards a parametric account. Handout. GLOW 17, Vienna.

Oseki, Yohei. 2015. Eliminating pair-merge. In Ulrike Steindl, Thomas Borerand Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the 32th west coast conference on formal linguistics*, 303–312. Cascadilla Proceedings Project.

Otani, Kazuyo & John Whitman. 1991. V-raising and VP-ellipsis. *Linguistic Inquiry* 22. 345–358.

Ott, Dannis. 2011. A note on free relative clauses in the theory of phases. *Linguistic Inquiry* 42(1). 183–192.

Ott, Dennis. 2015. Symmetric merge and local instability: Evidence from split topics. *Syntax* 18(2). 157–200.

Petersen, Carolina & Arhonto Terzi. 2015. Hyperraising and locality: a view from brazilian portuguese and greek. In *Proceedings of 50th chicago linguistics society (cls50)*, .

Pires, Acrisio & Wil Nediger. 2018. (hyper-)raising in Brazilian Portuguese: Interaction betwee ncase and agreement. In *The complement clauses in portuguese: syntax and acquisition*, 187–212. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Pires, Acrisio & Heather Taylor. 2007. The syntax of wh-in-situ and common ground. In *Proceedings of the annual meeting of the chicago linguistic society 43, vol.2*, 201–215.

Polinsky, Maria, Carlos Gómez Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan & Anne Sturgeon. 2013. Subject islands are different. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and islands effects*, 286–309. Cambridge: Cambridge University Press.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20(3). 65–424.

Åfarli, Tor A & Karumuri V. Subbarao. 2019. Models of grammar and the outcomes of long-term language contact: Language mixing in dakkhini. In Albert Alvarez-Gonzalez, Zarina Estrada-Fernández & Claudine Chamoreau (eds.), *Diverse scenarios of syntactic complexity*, 27–49. Amsterdam: John Benjamins.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences* 46. 152–174.

Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon*. Cambridge: Cambridge University Press.

Repp, Sophie & Volker Struckmeier. 2020. *Syntax – Eine Einführung*. J.B. Metzler.

di Ricerca, Tesi. 2020. *To be or not to be an island*: Università Ca' Foscari dissertation.

Richards, Marc. 2007. On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry* 38(3). 563–572.

Richards, Marc. 2008. Two kinds of variation in a minimalist system. In F Heck, G Müller & J Trommer (eds.), *Varieties of competition. linguistische arbeits berichte, vol. 87*, 133–162.

Richards, Marc. 2019. Problems of 'problems of projection': breaking a conceptual tie. *Catalan Journal of Linguistics* Special Issue: Generative Syntax, Questions, Crossroad, and Challenges. 139–152.

Richards, Marc D. 2012. On feature inheritance, defective phases, and the movement-morphology connection. In *Phases: Developing the framework*, 195–232. Berlin/Boston: De Guyter.

Riksem, Brita Ramsevik. 2018. *Language mixing in american norwegian noun phrases: An exoskeletal analysis of synchronic and diachronic patterns*: Norwegian University of Science and Technology dissertation.

Riksem, Brita Ramsevik., Maren.Berg. Grimstad, Terje. Lohndal & Tor A Åfarli. 2019. Language mixing within verbs and nouns in American Norwegian. *The Journal of Comparative Germanic Linguistics* 22. 189–209.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Lilian Haegeman (ed.), *Elements of grammar*, 281–337. Dordecht: Springer.

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

Rizzi, Luigi. 2007. On some properties of criterial freezing. In *Ciscl working papers on language and cognition, 1: Stil studies in linguistics*, 145–158. University of Siena.

Rizzi, Luigi. 2015. Notes on labeling and subject positions. In Simona Matteini Elisa Di Domenico, Cornelia Hamann (ed.), *Structures, strategies and beyond: Studies in honour of adriana belletti*, 17–46. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Rizzi, Luigi & Guglielmo Cinque. 2016. Functional categories and syntactic theory. *Annual Review of Linguistics* 2(1). 139–163.

Roberts, Ian. 2019. *Parameter hierarchies & universal grammar*. Oxford: Oxford University Press.

Roberts, Ian, Theresa Biberauer, Michelle Sheehan & Anders Holmberg. 2014. Complexity in comparative syntax: the view from modern parametric theory. In Frederick J. Newmeyer & Laureal B. Preston (eds.), *Measuring gramamtical complexity*, 103–127. Oxford: Oxford University Press.

Roeper, Tomas. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2(3). 169–186.

Ross, J. R. 1967. *Constraints on variable in syntax*: MIT dissertation.

Rothman, Jason & Noam Chomsky. 2018. Toward eliminating arbitrary stipulation related to parameters: linguistic innateness and the variational model. *Linguistic approaches to biligualism* 8(6). 764–769.

Sabel, Joachim. 2000. Partial Wh-movement and the Typology of Wh-Questions. In Uli Lutz, Gereo Müller & Arnim von Stechow (eds.), *Wh-scope marking*, 409–446. Amsterdam: John Benjamins Publishing Company.

Safir, Ken. 2019. The A/Ā Distinction as an Epiphenomenon. *Linguistic Inquiry* 50(2). 285–336. doi:10.1162/ling_a_00305. https://doi.org/10.1162/ling_a_00305.

Sag, Ivan Andrew. 1976. *Deletion and logical form*: MIT dissertation.

Saito, Mamoru. 1985. *Some asymmetries in japanese and their theoretical implications*: MIT dissertation.

Saito, Mamoru. 1989. Scrambling as semantically vacuous A′-movement. In *Alternative concepts of phrase structure*, 182–200. Chicago: University of Chicago Press.

Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1(1). 69–118.

Saito, Mamoru. 2003. A derivational approach to the interpretation of scrambling chains. *Lingua* 113. 481–518.

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

Saito, Mamoru. 2018. Kase as a weak head. *McGill Working Papers in Linguistics* 25(1). 382–391.

Saito, Mamoru & Naoki Fukui. 1998. Order in phrase structure and movement. *Linguistic Inquiry* 29(3). 439–474.

Saito, Mamoru & Keiko Murasugi. 1990. N′-deletion in japanese: A preliminary study. In *Japanese and korean linguistics 1*, 51–81.

Sakai, Hiromu. 1994. Derivational economy in long distance scrambling. In *Formal approaches to japanese linguistics 1: Mit working papers in lingusitics 24*, 294–314.

Sakai, Hiromu. 1996. *Derivational uniformity: A study of syntactic derivation in parameter setting*: University of California, Irvine dissertation.

Sakamoto, Yuta. 2017. *Escape from silent syntax*: University of Connecticut, Storrs dissertation.

Salzmann, Martin. 2013. Rule ordering in verb cluster formation. In Anke Assmann & Fabian Heck (eds.), *Rule interaction in grammar, volume 90 of linguistische arbeitsberichte*, 65–121. University of Leipzigi.

Seely, T. Daniel. 2006. Merge, derivational c-command, and subcategorization in a label-free syntax. In Cedric Boeckx (ed.), *Minimalist essays*, 182–217. Amsterdam: John Benjamins.

Sheehan, Michelle, Theresa Biberauer, Ian Roberts & Anders Holmberg. 2017. *The final-over-final condition: A syntactic universal*. Cambridge, MA: MIT Press.

Sheehan, Michelle L. 2013. Some implications of copy theory of labeling. *Syntax* 16(4). 363–396.

Siegel, Jeff. 2006. Links between SLA and creole studies: Past and present. In Claire Lefebvre, Lydia White & Christine Jordan (eds.), *L2 acquisition and creole genesis*, 15–46. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sorida, Masanobu. 2014. Multiple-branching structures in syntax. In *Proceedings of the 16th seoul international conference on generative grammar SICOGG16*, 411–419.

Sorida, Masanobu. 2017. A step toward a deeper understadning of parameters. Talk at Keio Colloquium, Tokyo, Japan, July 9th, 2017.

Stepanov, Arthur. 2007. The end of CED?: Minimalism and extraction domains. *Syntax* 10(1). 80–126.

Sugimoto, Yushi. 2021. Labeling defective T. *Studia Linguistica* 75(1). 150–164.

Sugimoto, Yushi & Marlyse Baptista. 2022. A late-insertion-based exoskeletal approach to the hybrid nature of functional features in creole languages. *Languages* 7(2). doi:10.3390/languages7020092. `https://www.mdpi.com/2226-471X/7/2/92`.

Tada, Hioraki. 1993. *A/A-bar partition in derivation*: MIT dissertation.

Takahashi, Daiko. 1994. *Minimalisty of movement*: The University of Connecticut dissertation.

Takahashi, Daiko. 2008. Quantificational null object and argument ellipsis. *Linguistic Inquiry* 39. 173–193.

Takahashi, Shoichi & Martina Gračanin-Yuksek. 2008. Morphosyntax of movement dependencies in haitian creole. *Syntax* 11(2). 223–250.

Takano, Yuji. 1994. Unbounded traces and indeterminacy of derivation. In Masaru Nakamura (ed.), *Current topics in English and Japanese*, 229–253. Tokyo, Japan: Hituzi Syobo.

Takano, Yuji. 1995. Predicate fronting and internal subjects. *Linguistic Inquiry* 26(2). 327–340.

Takano, Yuji. 2000. Illicit remnant movement: An argument for feature-driven movement. *Linguistic Inquiry* 31(1). 141–156.

Takita, Kensuke. 2010. *Cyclic linearization and constraints on movement and ellipsis*: Nanzan Univeristy, Aichi, Japan dissertation.

Takita, Kensuke, Nobu Goto & Yoshiyuki Shibata. 2016. Labeling through spell-out. *The Linguistic Review* 33(1). 177–198.

Tanaka, Hidekazu. 2002. Raising to object out of CP. *Linguistic Inquiry* 33(4). 637–652.

Teyssier, Paul. 1959. *La langue de gil vicente*. Paris: Klincksieck.

Tiskin, Daniel. 2017. Movement out of adjunct clauses in russian: Evidence from semantics. *Comparative Slavic Syntax and Semantics* 9(1). 33–43.

Torrence, Harold. 2005. *On the distribution of complementizer in Wolof*: University of California, Los Angeles dissertation.

Torrence, Harold. 2013. *Clause structure of wolof: Insights into the left periphery*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Travis, Lisa & Diane Massam. 2021. VP-fronting within a movement typology. In Vera Lee-Schoenfeld & Dennis Ott (eds.), *Parameters of predicate fronting*, New York/Oxford: Oxford University Press. doi:10.1093/oso/9780197545553.001.0001.

Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117(8). 1355 – 1377.

Truswell, Robert. 2011. *Events, phrases, and questions*. Oxford: Oxford University Press.

Ura, Hiroyuki. 1994. Hyper-raising and the theory of *pro*. In C. Giordano & D. Ardron (eds.), *Proceedings of the 6th annual student conference in linguistics, MIT working papers in linguistics 23*, 297–316. MITWPL.

Uriagereka, Juan. 1999. Multiple spell-out. In Samuel David Epstein & Norbert Hornstein (eds.), *Working minmalism*, 252–282. Cambridge, MA: MIT Press.

Velupillai, Viveka. 2015. *Pidgins, creoles and mixed languages: An introduction*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wekker, Herman. 1996. Creolization and the acquisition of English as a second language. In Herman Wekker (ed.), *Creole languages and language acquisition*, 139–149. Berlin: Mouton de Gruyter.

Westergaard, Marit. 2014. Linguistic variation and micro-cues in first language acquisition. In Kleanthes K. Grohmann (ed.), *Three factors and beyond: Socio-syntax and language acquisition*, 26–45. Amsterdam: John Benjamins.

Wexler, K. & P. W. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Wiltschko, Martina. 2013. The anatomy of universal categories: Developing discovery procedures. In Stefan Keine & Shayne Sloggett (eds.), *Proceedings of the 42nd meeting of hte north east linguistic society*, 257–276.

Wiltschko, Martina. 2014. *The universal structure of categories:towards a formal typology*. Cambridge, MA: MIT Press.

Wurmbrand, Susi. 2006. Licensing case. *Journal of Germanic Linguistics* 18(3). 175–236.

Wurmbrand, Susi. 2007. How complex are complex predicates. *Syntax* 10. 243–288.

Yoshida, Masaaya. 2006. *Constraints and mechanisms in long-distance dependency formation*: University of Maryland, College Park dissertation.

Zeijlstra, Hedde. 2020. Labeling, selection, and feature checking. In *Agree to agree: Agreement in the minimalist programme*, 31–70. Berlin: Language Science Press. doi:10.5281/zenodo.3541745.

Zyman, Erik. 2017. P'uherpecha hyperraising to object: An argument for purely altruistic movement. *Proceedings of Linguistic Society of America* 2(53). 1–15.

Zyman, Erik. 2021. Antilocality at the phase edge. *Syntax* 24(4). 510–556.