

Coalescence: A unification of bundling operations in syntax

Brian Hsu (University of North Carolina at Chapel Hill)

Keywords: head movement, head bundling, M-Merger, Feature Scattering, Coalescence

Author's note: This is the accepted version of this article (to appear in *Linguistic Inquiry*), which reflects all changes made during peer review prior to pre-publication copy-editing.

Abstract

This article revisits the status of two proposed bundling operations that affect heads: Feature Scattering (Georgi and Pianesi 1997), which accounts for variation in the distribution of features across functional heads, and M-Merger (Matushansky 2006), which accounts for head adjunction in head movement. While these mechanisms have been situated in the presyntactic lexicon and a postsyntactic module, respectively, I argue that they can receive a unified analysis in terms of one syntactic operation called Coalescence, which bundles structurally adjacent heads in particular configurations. This eliminates redundancies in the architecture of the grammar while maintaining prior empirical coverage, and sheds new light on long-puzzling properties of head movement. The proposal is illustrated in the analysis of several patterns of head bundling in the inflectional and clausal domains.

1. Introduction

Syntactic operations applying to heads, defined as syntactically indivisible bundles of features (Matushansky 2006), have played a pivotal role in explaining numerous word order patterns. This paper examines several properties of *head movement*, the displacement of heads within an extended projection, and *head bundling*, the occurrence of multiple features of an extended projection on a single head. My central claim is that

key properties of both patterns are best explained in terms of a prominence-based licensing restriction on features. In brief, I will propose that each category feature, defined provisionally as a feature that can head a projection in some language, is specified as either *dominant* or *recessive*, with a distribution subject to language-internal and cross-linguistic variation. Derivations are subject to the DOMINANCE CONDITION, a requirement that all heads contain a dominant feature. Although all category features are first Merged on distinct heads, the need to satisfy the Dominance Condition triggers a syntactic head-bundling operation COALESCENCE, which in some instances is fed by head movement. Informally, Coalescence takes place to prune phrase structure trees by combining weak or unproductive branches of the tree with adjacent stronger branches.

In terms of its main theoretical contributions, I argue that the proposal provides a unified account for a broad range of movement and bundling patterns, while resolving key theoretical problems of head movement in Minimalist syntax. Furthermore, it provides new empirical coverage of two less-discussed syntactic patterns involving heads: *delayed gratification* (a.k.a. *delayed EPP*) patterns in which head movement must precede phrasal movement to the same projection (Den Dikken 2007; Kandybowicz 2009; Gallego 2006, 2010), and *unrestricted edge feature* patterns where multiple probes on one head compete to trigger phrasal movement (Fanselow and Lenertová 2010).

I first present an overview of M-Merger (Matushansky 2006), a proposed syntactic bundling operation that forms featurally complex heads as the result of head movement; and the Feature Scattering Hypothesis (Georgi and Pianesi 1997), which claims that languages vary in the number of heads on which a set functional features is distributed

even in the absence of movement. Although the phenomena that motivate the two theories have not previously received a unified treatment, I show that they share three fundamental similarities: a head-adjacency locality restriction on bundling, a two-way distinction between “deficient” features that must be bundled versus “prominent” ones that do not, and the unique ability of “prominent” features to support phrasal specifiers. Consequently, I claim that both bundling patterns are generated by a single syntactic operation that applies to terminal nodes during the derivation, Coalescence. Broadly, bundling in Feature Scattering results from Coalescence operations that apply in the absence of head movement, while Coalescence fed by movement accounts for head concatenation in head movement. Coalescence subsumes Feature Scattering and M-Merger effects, thereby eliminating redundancies in the architecture of the grammar while maintaining prior empirical coverage.

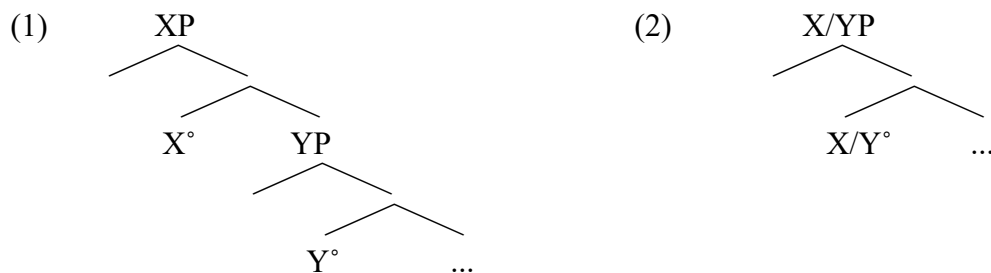
I then turn to a formalization of the Dominance Condition, dominant versus recessive features, and Coalescence. This is done in the context of a general theory of head movement and phrasal movement that extends and refines the system of Matushansky (2006). This system of features and operations permits new explanations for several patterns that have challenged prior approaches. First, it provides an explanatory trigger for head movement, and is able to reconcile the apparent local nature of head movement with anti-locality restrictions on movement. Finally, the additional claim that the EPP property is unique to dominant heads enables an account for delayed gratification and unrestricted edge feature patterns.

The paper is organized as follows. Section 2 reviews key properties of Feature Scattering and head movement, and shows that these bundling patterns share the same structural definition and locality restrictions. Section 3 outlines the formal properties of Coalescence, and the derivational ordering of bundling, head movement, and phrasal movement. Section 4 presents key case studies: strict and relaxed verb second patterns, perfect aspect marking in Catalan, and English contracted negation. Section 5 discusses the timing of Coalescence with respect to spell-out, and implications of the approach with theories of affix ordering. Section 6 concludes the paper.

2. Head bundling in grammar

2.1 Feature Scattering

Although languages appear to share a large inventory of hierarchically ordered features (Rizzi 1997; Cinque 1999), they vary in the number of positions in phrase structure that can be used to instantiate them. This tension can be resolved by a *bundling* parameter on the distribution of category features (to be defined at the end of this subsection), such as the Feature Scattering Hypothesis (Giorgi and Pianesi 1997). As a schematic example, some pairs of features [X] and [Y] can either occur on separate heads X° , Y° (1), or bundled on a single head X/Y° (2).



This contrasts with standard “cartographic” analyses in which each feature occurs on a single head in a universal hierarchy of projections (Cinque 1999; Kayne 2005a; Cinque and Rizzi 2009). On this view, (1) is the only possible configuration of features [X], [Y]. While both approaches are compatible with the observation that extended projections contain an intricate, possibly universal hierarchy of features (Cinque 1999), they make different predictions about possible movements and the availability of specifier positions (Bobaljik and Thráinsson 1998; Erlewine 2016; Douglas 2017; Hsu 2017).

To illustrate, consider Giorgi and Pianesi’s (1997; 2004) analysis of subjunctive embedded clauses in Italian. In brief, the presence of a complementizer affects the possible placement of subject DPs in the embedded clause, in a way that suggests that fewer specifier positions are available in clauses without a complementizer. First, consider the placement of subject DPs in subjunctive clauses without a complementizer. Here, speakers vary in where they allow subjects to occur. One set of speakers allows subjects to precede auxiliaries (3a), while a second set does not. For the second group, subjects in these clauses must occur in a post-verbal position (3b).

- (3) a. Gianni credeva [Maria avesse telefonato]
 Gianni believed Maria had.SBJV called
 ‘Gianni believed that Maria had called.’
- b. Gianni credeva [avesse telefonato Maria]
 Gianni believed had.SBJV called Maria
 ‘Gianni believed that Maria had called.’

These contrast with embedded subjunctive clauses with a complementizer. Here, all speakers permit preverbal subjects within the embedded clause (4).

- (4) Gianni credeva che [Maria avesse telefonato]
 Gianni believed that Maria had.SBJV called
 ‘Gianni believed that Maria had called.’

The two clause types also differ in possible subject placement in structures where an adjunct *wh*-word is extracted from the embedded clause.ⁱ In clauses without a complementizer, subjects must occur in the postverbal position, even for speakers who permit preverbal subjects in clauses without adjunct *wh*-word extraction.

- (5) a. ?* Perché credevi [Gianni avesse telefonato?]
 why you.believed Gianni had.SBJV phoned
 b. Perché credevi [avesse telefonato Gianni?]
 why you.believed had.SBJV phoned Gianni
 ‘Why did you believe Gianni had phoned?’

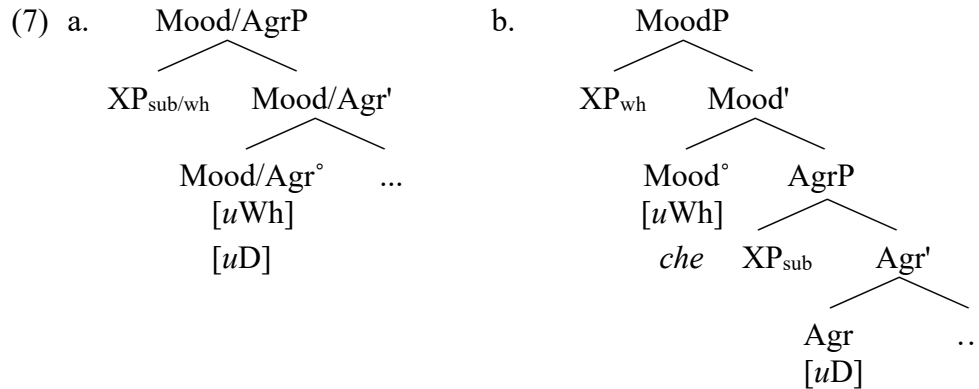
Again, this contrasts with embedded clauses with a complementizer. Here, adjunct *wh*-word extraction is possible from a clause with a preverbal subject.

- (6) Perché credevi [che Gianni fosse partito?]
 why you.believed that Gianni was.SBJV left
 ‘Why did you believe that Gianni had left?’

Giorgi and Pianesi (1997) account for these asymmetries as follows: (i) Complementizer-less subjunctive clauses contain a bundled Mood/AgrP, while clauses with a complementizer contain separate projections MoodP and AgrP. (ii) The highest inflected verb or auxiliary moves to the head that contains [Agr]. (iii) Preverbal subjects are specifiers of the projection containing [Agr]. (iv) Adjunct *wh*-words are first Merged as specifiers of the projection whose head contains [Mood] (cf. Cinque 1990).

Bundled Mood/AgrP (7a) licenses one specifier that serves as either the final position of the embedded subject, or the intermediate landing site of an extracted adjunct *wh*-word. Inter-speaker variation in the acceptability of (3a) results from variation in whether

bundled Mood/Agr° can trigger movement of the subject from its lower postverbal position. An adjunct *wh*-word cannot be extracted from a clause with a preverbal subject because Spec, Mood/AgrP cannot be simultaneously filled by an adjunct *wh*-word and a subject; extraction can occur only if the subject remains below Mood/AgrP.



In contrast, embedded clauses with the complementizer *che* contain two projections AgrP and MoodP, where the *che* is the realization of Mood° (7b). Preverbal subject orders are grammatical for all speakers, as subject movement to Spec, AgrP is permitted.ⁱⁱ Adjunct *wh*-words can be extracted from clauses with a preverbal subject because they are generated in a separate position Spec, MoodP.

Crucially, these patterns are not expected under a strictly cartographic “one feature - one head” alternative (Cinque and Rizzi 2008) in which MoodP and AgrP are always separately projected, and the only difference between clauses with and without *che* is in whether Mood° is pronounced. There is no principled explanation in this view for why Spec, AgrP cannot be filled when Mood° is unpronounced, as would be needed to account for the uniform ungrammaticality of (5a) and for speakers who reject (3a).

Although not always couched in the same terms, Feature Scattering has been applied to variation in subject positions in the IP domain (Poletto 2000), Infl° and Agr° (Iatridou 1990; Ouhalla 1991; Speas 1991; Bobaljik 1995; Thráinsson 1996; Bobaljik and Thráinsson 1998), Voice° and Causative° heads (Pylkkänen 2002), C° and Infl° (Bennett et al. 2012; Erlewine 2018), deixis and reference in the nominal domain (Panagiotidis 2014; Höhn 2016; Hsu and Syed 2019), and the extended complementizer domain (Douglas 2017; Hsu 2017).

In the proposal of Giorgi and Pianesi (1997), languages share a universal inventory of features, but differ in how individual features [F] are packaged onto the lexical items that enter syntactic derivations (a similar proposal is made in Cowper 2005). In other words, the locus of variation is within the *pre-syntactic lexicon*, rather than in syntactic derivations.ⁱⁱⁱ Here, I will focus on their primary claims about the possible distributions of features on lexical items. While Giorgi and Pianesi present Feature Scattering as an operation that scatters the features of initially complex lexical items, I will show that the relevant restrictions can be generated by a bundling operation that further permits a unified analysis with head movement patterns.

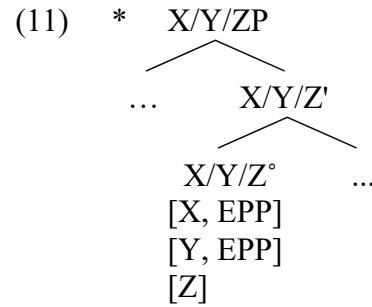
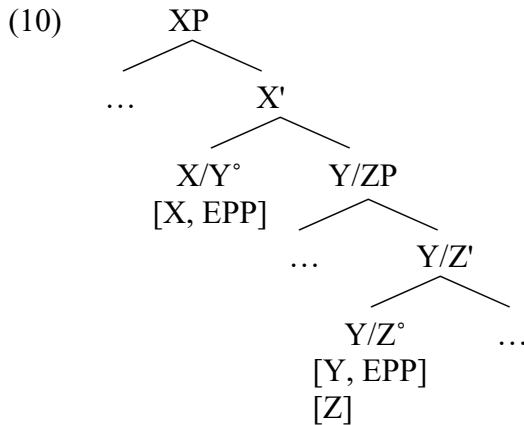
To account for the observation that languages share a universal hierarchy of projections, regardless of the number of realized projections in a given structure, Giorgi and Pianesi propose that languages share a universal feature set, and restrictions on the order in which individual features can be checked (in this case, by external Merge).

- (8) UNIVERSAL ORDERING CONSTRAINT: Features are ordered so that given $F_1 > F_2$, the checking of F_1 does not follow the checking of F_2 . (Giorgi and Pianesi 1997)

This constraint restricts the possible feature bundles that can be found within and across languages. It imposes a *locality restriction* on feature bundling, restricted to apply to features that are contiguous in the universal checking order. To illustrate, the hypothetical ordered features $[Z] > [Y] > [X]$ can be realized only in the configurations in (9). The Universal Ordering Constraint crucially rules out the appearance of $[Z]$ and $[X]$ in a projection that excludes $[Y]$ (9f). In brief, bundling cannot “skip” intervening features.

- (9)
- | | | | | |
|----|-----------|-------|---------|-----|
| a. | $[XP$ | $[YP$ | $[ZP$ | ... |
| b. | $[X/YP$ | | $[ZP$ | ... |
| c. | $[XP$ | | $[Y/ZP$ | ... |
| d. | $[X/Y/ZP$ | | | ... |
| f. | $*[X/ZP$ | | $[YP$ | ... |

In addition, Giorgi and Pianesi (1997: 15) propose a key link between the degree of bundling and the number of specifier positions necessary to license items in the numeration. Under the assumption that each projection admits at most one specifier, a feature that can be checked only if its projection has a filled specifier cannot be bundled with another feature that has the same requirement.^{iv} We illustrate again with the hypothetical ordered features $[Z] > [Y] > [X]$. If both $[X]$ and $[Y]$ must occur in a projection with a filled specifier (indicated in the figures as $[EPP]$ – a different account of EPP features is given in Section 3.3), $[X]$ and $[Y]$ cannot be bundled on a single head (11).^v In contrast, there is no prohibition against the bundling of $[Z]$, which lacks this requirement, with either of the other features.



Broadly speaking, this can be viewed as a contrast between “prominent” features whose projections must host a specifier, and “deficient” ones that lack this requirement.

Furthermore, while “prominent” features cannot be bundled together, there is no prohibition against the bundling of prominent features with deficient features.

While descriptively successful in many domains, the notion that feature bundling takes place in the lexicon raises theoretical issues about the origin of hierarchical ordering among projections. To illustrate, consider how the hierarchy of projections is generally implemented in theories in which each category feature is realized on a separate head. In this view, the order of feature checking reflects the distribution of *c-selection* features (Svenonius 1994; Julien 2002; Matushansky 2006; Di Sciullo and Isac 2008). Each head contains an interpretable category feature, and optionally an uninterpretable category feature. Because Merge takes place to check uninterpretable category features, the order of feature checking arises from selectional properties of the atoms of the syntactic derivation. In this view, *category features* can be defined as features of syntactic objects that trigger external Merge (Di Sciullo and Isac 2008).

On the other hand, if some category features are bundled in the pre-syntactic lexicon, the restriction on the order of feature checking has to exist independently of syntactic objects, for instance as a stipulated metacondition on bundling in the lexicon like the Universal Ordering Constraint. Restrictions on the order of feature checking are thus redundantly specified in separate modules of the grammar. In this work, I argue that this redundancy is dispensed with all bundling of category features takes place during the derivation; the order of feature checking is established uniquely via c-selection features.

2.2 Head movement and M-Merger

Head movement has played a pivotal role in analyses of language-internal and cross-linguistic word order variation – briefly illustrated here with two well-known examples. Within English, subject-auxiliary inversion patterns are commonly accounted for as T-to-C head movement (den Besten 1983). This is particularly clear in counterfactual conditionals, in which auxiliaries move from T to the position otherwise occupied by the complementizer *if* (12).

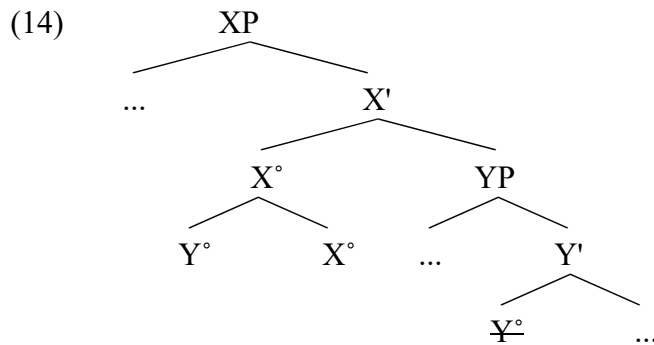
- (12) a. If_C Michael had_T gone to Phoenix, the company would have collapsed.
 b. Had_{C+T} Michael ~~had~~ gone to Phoenix, the company would have collapsed.

Head movement is also instrumental in explaining cross-linguistic variation. As a famous example, English and French differ in the placement of tense-inflected verbs relative to adverbs like ‘often’ (Emonds 1970; 1978; Pollock 1989). Assuming that such adverbs occupy a position between TP and vP in both languages, the difference is accounted for in terms of whether tensed lexical verbs remain in v (English) or move to T (French).

- (13) a. Lucille _T often drinks_v martinis *English*
 b. Lucille boit_{v+T} souvent ~~boit~~ des martinis *French*

Head movement patterns have received numerous approaches in generative syntax (see Dékány 2018 for a recent summary). Broadly speaking, they have been analyzed in terms of various syntactic operations (Travis 1984; Baker 1988; Georgi and Müller 2010), a syntactic process whose output is subject to post-syntactic operations or pronunciation rules (Embick and Noyer 2001; Roberts 2005; Matushansky 2006; Harizanov 2014; Arregi and Pietraszko 2018), or as a purely post-syntactic operation (Chomsky 2000; Boeckx and Stjepanovic 2001; Platzack 2013). Alternatively, it has been argued that some types of head movement occur in syntax, while others occur postsyntactically (Harizanov and Gribanova 2019). In this subsection, I outline key justifications for a syntactic analysis of head movement, and remaining issues that the proposal aims to address.

Prior to the emergence of Minimalism, head movement was largely analyzed as a single movement operation that creates an adjunction structure in which both the moved head and attracting head are dominated by a head-level projection (14). Additional head-internal branching structure can be created by successive steps of head movement (Travis 1984; Baker 1988). Head adjunction structures account well for the observation that head movement often creates morphologically complex forms in which features of the target position appear as affixes on the exponent of the moved head (Baker 1988; Julien 2002).



Key aspects of this “traditional” analysis become suspect in the bottom-up, derivational theory of syntax developed in the Minimalist Program, and the problematic aspects of head movement become particularly apparent when compared with phrasal movement (Chomsky 1994, 2000; Mahajan 2001; see Dékány 2018 for a recent overview). Here, we consider three key objections levied against head movement as a syntactic operation.

First, traditional head movement fails to extend the root node of the tree (Chomsky's 1995 Extension Condition), unlike both external Merge and phrasal movement. Consequently, the moved head does not asymmetrically c-command its lower position, unlike the result of phrasal movement.^{vi}

Second, locality restrictions on head movement do not resemble those of phrasal movement. Although the generalization is subject to debate (e.g. Rivero 1991; Borsley et al. 1996; Roberts 2010 and references therein), head movement from X° to Y° generally cannot skip an intervening head Z° (the Head Movement Constraint of Travis 1984), whereas phrasal movement does not require such locality. Moreover, it has been argued that phrasal movement is *anti-local*; phrasal movement from XP to YP must cross some type of intervening phrase ZP (Grohman 2001, 2002, 2003; Abels 2003; Erlewine 2016). These opposing locality restrictions are shown schematically below with Spec-to-Spec

anti-locality (Erlewine 2016). While this distinction does not necessarily cast doubt on head movement as a syntactic operation, at minimum it suggests a fundamental difference in the types of features that trigger head movement versus phrasal movement.

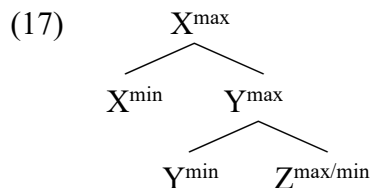
(15) a. $[_{XP} \quad X+Y \quad [_{YP} \quad \cancel{Y} \quad \dots$

b. $*[_{XP} \quad X+Y \quad [_{ZP} \quad Z \quad [_{YP} \quad \cancel{Y} \quad \dots$

(16) a. $*[_{XP} \quad WP \quad X \quad [_{YP} \quad \cancel{WP} \quad Y \quad \dots$

b. $[_{XP} \quad WP \quad X \quad [_{ZP} \quad Z \quad [_{YP} \quad \cancel{WP} \quad Y \quad \dots$

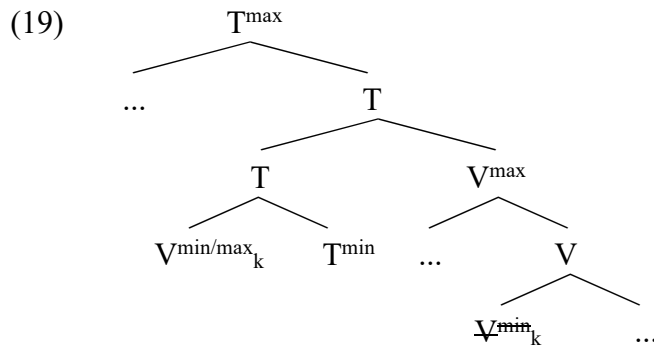
Third, head adjunction structures violate the Chain Uniformity Condition. A key insight of Bare Phrase Structure (Chomsky 1994) is that the phrase structure status of a syntactic object can be determined uniquely by its position within the tree. A syntactic object whose label does not project is a maximal projection (notationally XP or X^{\max}), while an object whose label is not identical to that of a node that it dominates is a minimal projection (X° or X^{\min}). An object with both properties is simultaneously maximal and minimal. The three options are schematized in (17).



Intermediate projections (X') that are neither minimal nor maximal are proposed to be inaccessible to syntactic operations and thus unable to undergo movement. To rule out movements deemed to be impossible, such as head-to-spec movement or XP adjunction to heads, Chomsky (1995: 253) posits a uniformity condition on movement chains:

(18) UNIFORMITY CONDITION
A chain is uniform with respect to phrase structure status

The Uniformity Condition is notably violated by head adjunction structures. Consider the structure commonly used for V° -to- T° movement, where V° adjoins to T° .



The two copies of V differ in their phrase structure status, in violation of the Uniformity Condition. Because the higher copy of V does not project, it is both minimal and maximal, while the lower copy is minimal because it projects. As noted by Harley (2013), successive-cyclic head movement proves even more problematic. For instance, because the complex V - T head created in (19) is neither minimal or maximal, it is predicted to be inaccessible to later syntactic operations (such as T -to- C movement).

Although these observations have been taken to suggest that head movement is a post-syntactic operation, other patterns suggest that it takes place during the syntactic derivation. First, some instances of head movement have semantic effects, often related to scope and the licensing of negative polarity items (Lechner 2006; Matushansky 2006; Roberts 2010; Hartman 2011; Matyiku 2017), which are unexpected if head movement occurs at PF. Here, I focus on a second case: patterns in which head movement is needed to license phrasal movement to the specifier of the target projection (Gallego 2006, 2010; Den Dikken 2007; Kandybowicz 2009). I illustrate with Den Dikken's (2007) account of

Holmberg's Generalization, the observation that in Scandinavian Germanic languages, certain (typically definite) direct objects can precede sentential negation only if verb movement also takes place.

(20) jag kysste henne inte ~~kysste~~ ~~henne~~
 I kissed her not
 'I did not kiss her.'

(21) a. *at jag henne inte kysste ~~henne~~
 that I her not kissed

b. at jag inte kysste henne
 that I not kissed her
 '... that I did not kiss her.'

(den Dikken 2007)

Den Dikken argues that object shift is driven by a functional projection above vP whose head contains a probe that agrees with the object. However, the probe can only trigger object movement if $v+V^\circ$ first head-moves to F° . Abstracting away from the exact label and probing feature, the derivation is shown in (22).^{vii} Verb movement to a higher projection and subject movement in later steps produce the word order in (20).

(22) a. [_{FP} F [_{vP} subject [_{v'} $v+V$ [_{VP} \forall object]]]]
 [PROBE]

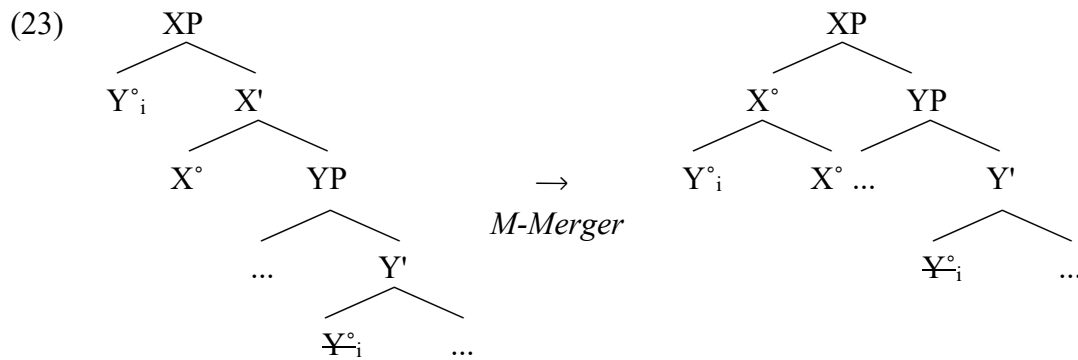
b. [_{FP} $F+v+V$ [_{vP} subject [_{v'} ~~$v+V$~~ [_{VP} \forall object]]]]
 [PROBE]

c. [_{FP} object $F+v+V$ [_{vP} subject [_{v'} v [_{VP} \forall ~~object~~]]]]
 [PROBE]

I will refer to this type of pattern as *delayed gratification*, in the sense that a probe of the target head is only able to attract a phrasal specifier after head movement. While such patterns are not commonly discussed in the literature on head movement (aside from Dékány 2018; 34), they suggest that the displacement of heads takes place in syntax; a

feeding relationship between head movement and phrasal movement is unexpected if head movement is a postsyntactic operation while phrasal movement is not. Such patterns are also problematic for syntactic accounts in which phrasal movement and head movement involve non-overlapping sets of features, and those in which the ability to trigger phrasal movement is an inherent property (i.e. strength) of individual probes (cf. Chomsky 1995). I argue in Section 3 that the ability to license a specifier, the EPP property, is inherited from the lower, moved head.

We now turn to Matushansky's (2006) analysis of head movement, which forms a basis for the proposal in Section 3. Matushansky's analysis aims in particular to address head movement's apparent incompatibilities with the Extension Condition, unique locality restriction, and the prohibition against movement of subparts of a complex head formed by head movement (a.k.a. *excorporation*). The structure produced by traditional head movement (14) takes place in two steps. First, a lower head moves to the specifier of the highest head in the derivation, as occurs in phrasal movement, satisfying the Extension Condition and c-command condition on movement.^{viii} The two heads are then bundled by an operation called M-Merger.



Matushansky proposes that head movement differs from phrasal movement in its featural trigger: phrasal movement is triggered by Agree (feature valuation), while head movement is triggered by c-selectional features of the attracting head. Because c-selection is a local relation between heads and complements, this drives the effects of the Head Movement Constraint.

However, several issues arise in this feature system. First, given the assumption that that c-selection features are checked under adjacency, there is no apparent need for movement to take place. Second, it remains unresolved why c-selection does not always result in head movement. In other words, it is unexplained why languages vary in head-movement paths within extended projections. The problem here reflects the difficulty in identifying a featural trigger in many instances of head movement that is distinct from the need for heads to be bundled (Baker 1988; Julien 2002; Roberts 2005).

Lastly, Matushansky proposes that M-Merger is a PF operation that applies after Spell-out, rather than a syntactic one. Spell-out renders the internal structure of the complex head opaque to later syntactic operations, accounting for the impossibility of excorporation. This also leaves the resulting head-adjunction structure immune to the Extension Condition and c-command condition on movement, which remain satisfied in the narrow syntax. However, the proposal requires an architecture of grammar in which constituents created by the morphological component remain accessible to later syntactic operations. To preview the main proposal, I will maintain Matushansky's two-step derivation of head movement as movement followed by bundling. However, the

impossibility of excorporation and compliance with the Extension Condition is accounted for with the claim that no internal branching structure is created when heads are bundled.

2.3 Structural resemblances

Feature Scattering and M-Merger are similar bundling mechanisms that have been posited to occur in the presyntactic lexicon and postsyntactic morphology, respectively. While it may be that similar operations can take place in different components of the grammar, the pursuit of a theory that minimizes the complexity of grammatical derivations motivates a unified analysis. I argue that key structural similarities between Feature Scattering and M-Merger suggest that both types of bundling can be attributed to a single operation that applies in one component of the grammar, the syntactic derivation.

First, consider the locality conditions that constrain both types of bundling. Under the Feature Scattering Hypothesis, a pair of features that are adjacent in the universal checking order must be realized either in immediately adjacent projections or bundled in a single projection. Similarly, M-Merger applies to adjacent heads in an asymmetric c-command relation, with no intervening specifier. In this sense, both types of head-bundling are restricted by the same condition of head adjacency (defined in Section 3.1). The upshot is that it is possible to generate both patterns from a bundling operation that applies during the syntactic derivation to structurally adjacent heads.

Second, both Feature Scattering and M-Merger patterns involve an interplay between “prominent” and “deficient” features. In Giorgi and Pianesi (1997), the number of projections that instantiates a set of category features is determined by the number of prominent features that must project; all other features are bundled. Head movement

requires a similar relationship between two types of features: a deficient feature in the target head that must be bundled (Julien 2002, Roberts 2005), and a lower head with a prominent feature that enables bundling. Furthermore, the lower head must first move to a position where it c-commands the target head in order for bundling to occur.

Third, in both patterns, phrasal specifiers are only licensed by heads that contain a prominent feature. This is explicitly argued by Giorgi and Pianesi (1997) for Feature Scattering. Although this observation is less obvious for head movement, it successfully characterizes delayed gratification patterns; the probe of a target projection obtains its ability to license a specifier from the moved head, which also enables head bundling.

In summary, head bundling in both feature scattering and head movement serves to license “deficient” category features, which cannot host phrasal specifiers or be realized with a standalone head, by associating them structurally with a “prominent” feature that can.^{ix} In essence, this is a type of *prominence based licensing* pattern found in a range of phonological domains (Itô 1988; Goldsmith 1990; Steriade 1995; Walker 2011; a.o.).

Given these similarities, I propose that both bundling patterns can be unified as the result of one syntactic operation. To preview the proposals in Section 3, all category features enter the derivation on distinct heads, which can be bundled by a syntactic operation called *Coalescence*. Coalescence preceded by head displacement generates the effects of M-Merger, while the effects of Feature Scattering are generated by applications of Coalescence that are not immediately preceded by movement.

With few exceptions (Giorgi and Pianesi 1997; Bobaljik and Thráinsson 1998), the two types of bundling are not generally discussed together in individual analyses. It is

possible in many cases to investigate variation in the bundling of functional projections independently of the mechanics of head movement, and vice versa. However, there are empirical advantages to a unified approach. Coalescence allows a parsimonious analysis of patterns that involve both types of bundling; for instance, where the target projection of head movement contains a bundle of probes that are associated with distinct category features. Section 4.1 discusses one such case in the context of verb second patterns.

3 Coalescence

This section defines the bundling operation Coalescence and the proposed feature system that determines when it applies. I assume a bottom-to-top, derivational theory of syntax (Chomsky 1995, 2000, 2008) in which each derivation begins with a selection of lexical items to be manipulated, a.k.a. the numeration. Lexical items are then assembled into a hierarchical constituent structure by internal or external Merge. Syntactic operations are triggered by features of lexical items, and variation arises from differences in the featural properties of lexical items (Borer 1984; Chomsky 1995). Based on the desiderata made in Section 2.1, I assume that all lexical items contain exactly one interpretable category feature (i.e. a feature whose selection triggers external Merge), and that each category feature can be associated with other interpretable and uninterpretable features (e.g. the lexical item containing category feature [Tense] can also contain [$u\phi$], [pres]).

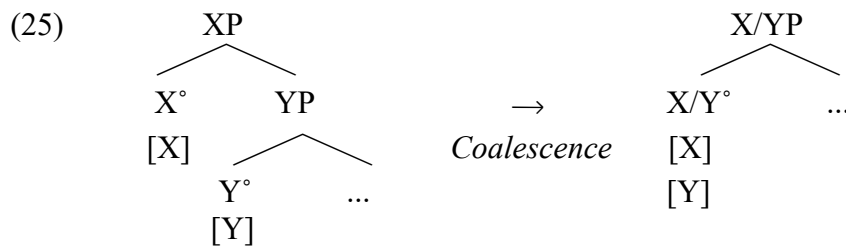
3.1 Defining Coalescence

We first consider the structural definition of this bundling operation. I propose that Coalescence, similar to M-Merger, applies to structurally adjacent heads: heads in an

asymmetric c-command relationship that are not separated by a specifier. I will use the following definition of head adjacency:^x

- (24) α and β are head-adjacent iff:
- (i) α and β are minimal projections (i.e. heads),
 - (ii) α asymmetrically c-commands β ,
 - (iii) there is no node κ that asymmetrically c-commands β and is asymmetrically c-commanded by α .

In this configuration, Coalescence creates a single node that contains all features associated with the individual heads.^{xi}



I depart from M-Merger and prior approaches to head adjunction by proposing that no branching structure is present in the newly formed head (I leave it an open question as to whether feature bundles are subject to other ordering relations; cf. Cowper 2005). There are two primary motives for the elimination of internal branching structure: First, the absence of branching accounts for the impossibility of excorporation from bundled heads, without having to assume that bundling triggers Spell-out, as Matushansky (2006) proposes for M-Merger. Second, this representation obviates incompatibilities of traditional adjunction structures with the Uniformity Condition (discussed in greater depth in Section 3.3). Section 5 addresses to the question of how affix ordering generalizations are captured in the absence of branching structure.

3.2 Dominance and Recession

We now turn to defining the structural environment that triggers Coalescence that can account for the patterns attributed to Feature Scattering and M-Merger. First, I propose a binary distinction between *dominant* and *recessive* category features.

(26) A category feature is either *dominant* [F_D] or *recessive* [F_R]

Informally, recessive features are those that must occur on a bundled head (e.g. [Asp] obligatorily realized with [T] in T/AspP, or [T] that must appear on V). Dominant features do not need to be bundled, and are potential “hosts” for recessive features. Finally, only heads with a dominant feature can license a phrasal specifier (cf. Section 3.4).

I assume that this is a distinction among formal features visible to syntactic operations, and not directly predictable from other factors. For instance, although there is a tendency for heads that undergo bundling to have null or affixal exponents, it is not possible to state a phonological characterization of the phonologically dependent or affixal items that must undergo head bundling to the exclusion of similar forms that do not. For example, while undergoers of bundling are often affixal in the sense that their exponents must be linearly adjacent to particular types of morphemes, the same generalization applies to bound roots (ex. Spanish *habl-* ‘speak’) that pattern as dominant heads.^{xii} Furthermore, not all syntactic heads that have affixal exponents undergo bundling (cf. Section 4.2).

In addition to the distinction between dominant and recessive features, a distinction is required between *dominant* versus *recessive heads*. While category features are lexically specified as being dominant or recessive, the status of a head is determined by its featural composition. A head that contains at least one dominant feature is dominant (27),

whereas a head that contains only recessive features is recessive (28). In all following examples, dominance is indicated with subscript _D, recessiveness with _R.

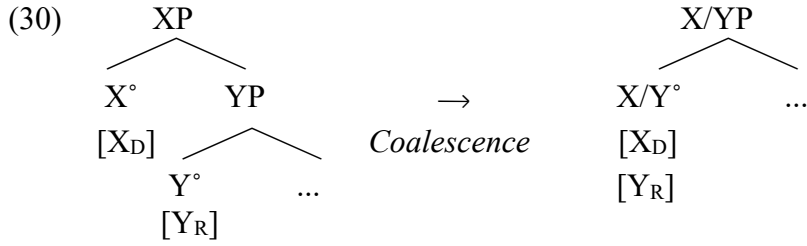
$$(27) \quad \begin{array}{c} X/Y^{\circ}_D \\ [X_D] \\ [Y_R] \end{array}$$

$$(28) \quad \begin{array}{c} X^{\circ}_R \\ [X_R] \end{array}$$

I assume that the grammaticality of surface forms is determined not only by principles of the syntax proper, but by well-formedness requirements of the Conceptual-Intentional (LF) and Articulatory-Perceptual (PF) interfaces (Chomsky 1993, 1995, 2000). I propose that the recessive heads are not legitimate PF objects, and that Coalescence applies to ensure that no recessive heads remain at the end of the derivation. I refer to this inviolable restriction as the Dominance Condition.

- (29) DOMINANCE CONDITION
All terminal nodes of the syntactic representation contain a dominant feature.

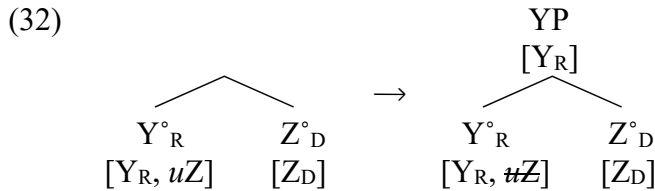
To which structural configurations of dominant and recessive heads does Coalescence apply? The key patterns to be accounted for here are that (i) dominant heads do not undergo Coalescence with each other and that (ii) bundling in head movement requires a lower dominant head to first move to the specifier of the target projection. To account for these restrictions, I propose that Coalescence only applies in a head-adjacency configuration where a dominant head asymmetrically c-commands a recessive one (30). This ensures that pairs of dominant heads cannot be bundled, and that a recessive head cannot trigger Coalescence with a lower head.^{xiii}



Here, I illustrate the derivation of a bundled X/Y° head. I assume that the order of projections arises from the distribution of uninterpretable c-selectional features (Svenonius 1994; Julien 2002; Matushansky 2006; Di Sciullo and Isac 2008). I further assume that the inventory of category features and their c-selectional properties do not depend on whether they are dominant or recessive.

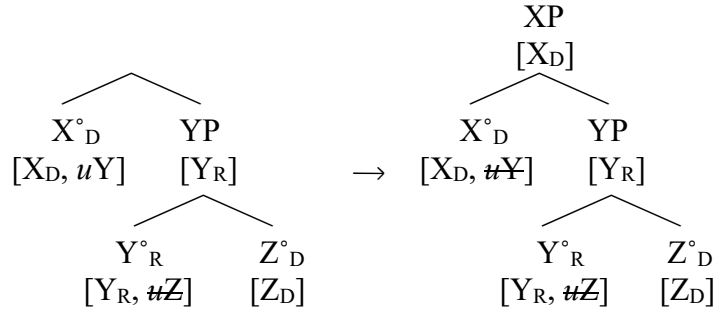
- (31) *Numeration:*
 $\text{X} [\text{X}_\text{D}, u\text{Y}]$
 $\text{Y} [\text{Y}_\text{R}, u\text{Z}]$
 $\text{Z} [\text{Z}_\text{D}]$

In the first step, Y° c-selects Z° . Z° is dominant, as it contains a dominant category feature. Y° is recessive as it contains only recessive category features. Y° Merges with Z° to check $[u\text{Z}]$ on Y° . Assuming that the c-selecting head projects its features, the newly formed root node is a projection of $[\text{Y}]$. Coalescence cannot bundle Y°_R and Z°_D because Z°_D does not asymmetrically c-command Y°_R .



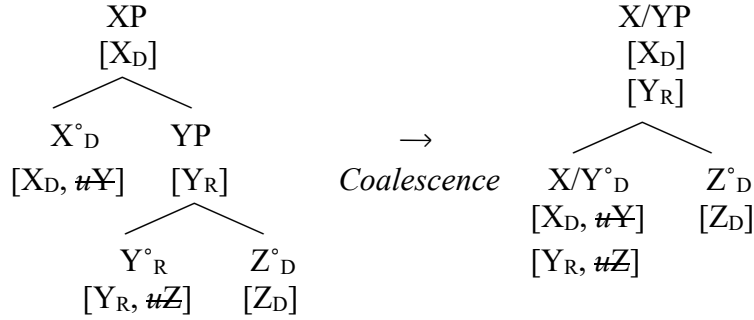
X° , whose category feature is dominant, then enters the derivation. Merge applies to check its $[u\text{Y}]$ feature and creates a new phrase headed by X .

(33)



Because this step creates a head adjacency configuration where the lower head is recessive, Coalescence applies. The bundled X/Y_D head is dominant because it contains the dominant feature $[X_D]$, and no recessive heads remain in the workspace.

(34)



At this point, the question arises of how the label of the highest node is determined.

Maintaining the assumption that all features of the selecting head project, both $[X]$ and

$[Y]$ project to the root node. This appears to run afoul of the expectation that syntactic

operations do not alter the labels of items that they manipulate, the No Tampering

Condition (Chomsky 2008).^{xiv} I propose that the problem can be avoided by adopting

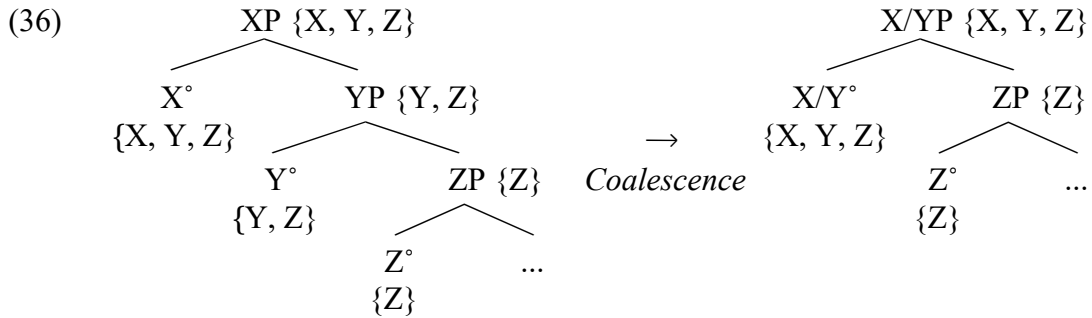
Category Percolation (Keine 2019), proposed independently to explain key properties of

extended projections (van Riemsdijk 1988, 1998; Grimshaw 1991, 2000).

(35) CATEGORY PERCOLATION

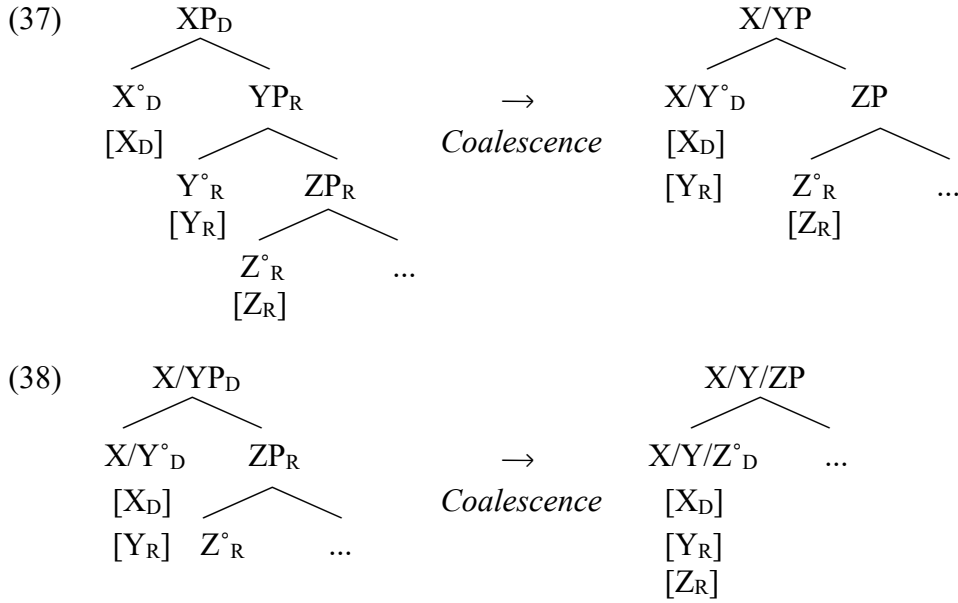
Given an extended projection $\Phi = \{\Pi_n > \Pi_{n-1} > \dots \Pi_1\}$, the categorial features of Π_m percolate to Π_{m+1} .

Specifically, I adopt a variant of this proposal in which each head contains the category subfeatures of all lower heads in the same extended projection (Keine 2019; 38, fn. 21). This is shown schematically in (36) for a structure in which X° , Y° , and Z° are in the same extended projection, with subfeatures shown in curly brackets. Note that the root node contains the same subfeatures both before and after Coalescence of X° and Y° , because the set of subfeatures in $YP \{Y, Z\}$, is a subset of those in $XP \{X, Y, Z\}$. Coalescence within an extended projection cannot create head or root node labels distinct from those of non-bundled heads, in compliance with the No Tampering Condition.



For consistency of presentation, I will continue to indicate bundling on head and phrase node labels (e.g. X/Y° , X/YP), while noting that these are not substantively distinct from their non-bundled counterparts due to Category Percolation. Category subfeatures are omitted in the remainder of the paper, unless otherwise noted.

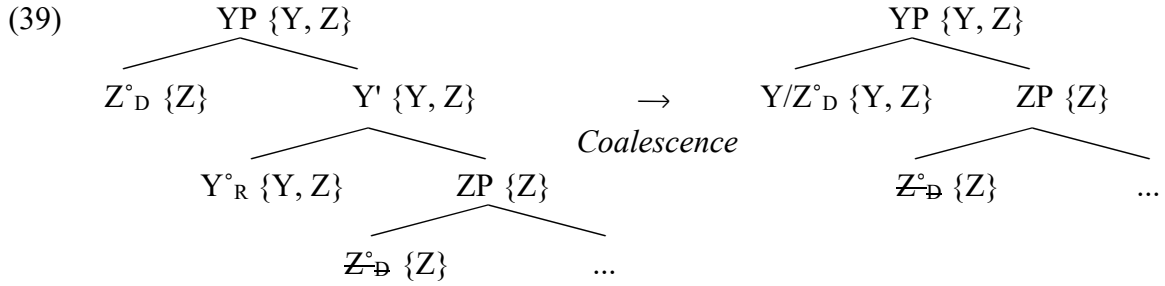
Lastly, complex heads that contain more than one recessive category feature are generated by successive “top-down” applications of Coalescence. At each step in (37) and (38), the topmost dominant head is bundled with a head-adjacent recessive head. For representational simplicity, c-selection features and features of phrasal nodes are omitted in the remainder of the paper.



Having illustrated the application of Coalescence following external Merge of a dominant head, the next subsection turns to Coalescence fed by internal Merge.

3.3 Coalescence and head movement

Like Matushansky (2006), I propose that head movement first consists of movement of a lower head to the specifier of the target projection. However, I claim that the movement is not triggered by a probe or selecting property of the target head. Rather, it is a Last Resort operation that ensures that the Dominance Condition is satisfied. The recessive target head attracts the closest dominant head that contains a subset of its category features (maintaining Category Percolation), creating the head-adjacency configuration that permits Coalescence. This is schematically shown in (39), in which Z°_D first moves to the specifier of YP. Again, Category Percolation ensures that the root node following Coalescence is featurally identical to the root node prior to movement.



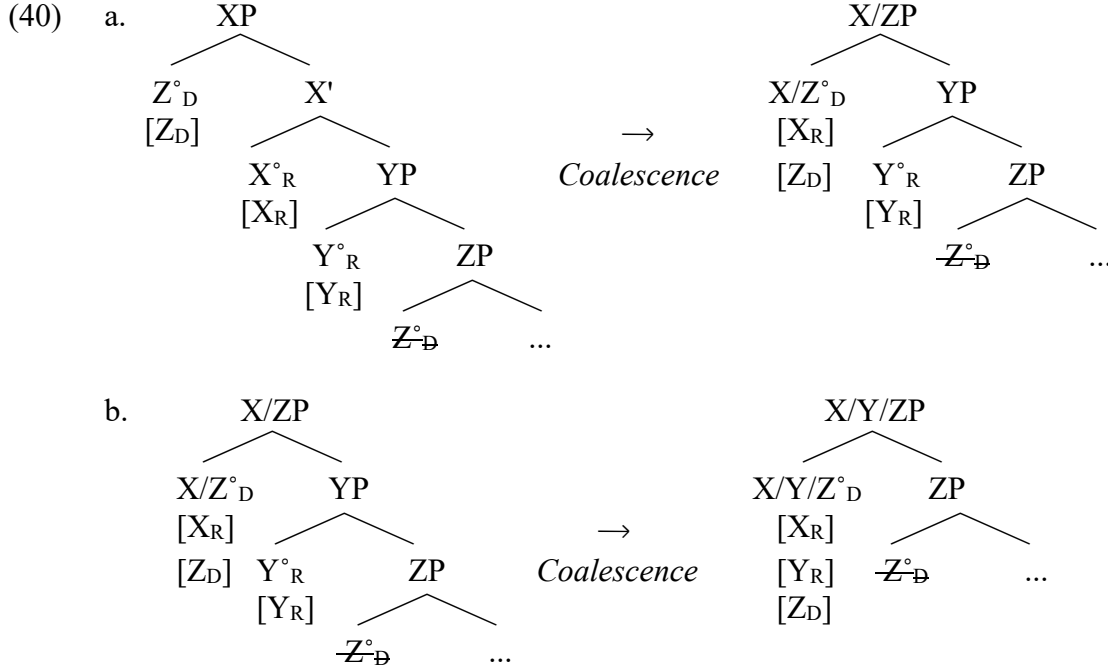
Note that Category Percolation and the absence of head-internal branching ensure that head movement satisfies the Uniformity Condition. After Coalescence, the bundled Y/Z°_D head is a minimal projection of $\{Y\}$ and $\{Z\}$. Furthermore, the two links of the head movement chain are minimal projections of $\{Z\}$, and the higher head c-commands its lower position. Similarly, later movement of the bundled head produces only minimal projections, and there is no need to posit movement of intermediate projections.^{xv}

Category Percolation, in concert with the No Tampering Condition, also accounts for the generalization that lexical heads can move to functional heads, but not vice versa (Li 1990; Delsing 1993; Baker 1996, 2003). It is possible for V° to move to C° and undergo Coalescence, as the subfeatures of $C^\circ \{C, T, V\}$ include those of $V^\circ \{V\}$. The features of the root node remain unchanged. However, C° cannot undergo Coalescence with a higher V° because V° does not contain the subfeatures $\{C, T\}$. Coalescence would result in the addition of those features to the root node, in violation of the No Tampering Condition.

Another key consequence of the proposal is that there is no need for the grammar to include a locality restriction on head movement like the Head Movement Constraint (Travis 1984). In structures where one or more recessive heads intervene between the source and target positions when movement takes place, iterative application of Coalescence ensures that the two positions of the moved head will be head-adjacent at the

end of the derivation.^{xvi} This is illustrated in (40) for a structure in which the dominant head Z°_D closest to the target X°_R crosses an intervening projection headed by Y°_R .

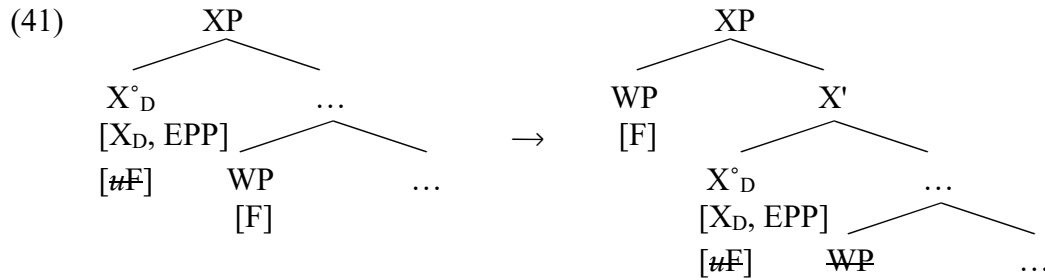
Coalescence first bundles Z°_D and X°_R , before bundling X/Z°_D and Y°_R .



3.4 Coalescence and phrasal movement

Like Matushansky (2006), I propose that both head movement and phrasal movement involve movement to the specifier of the target projection, in accordance with the Extension Condition. However, they differ in key featural properties of both the head of the target projection and of the moved item. In head movement, the moved head has a subset of the category features of the target projection (due to Category Percolation), whereas this is not the case in phrasal movement. Furthermore, phrasal movement is dependent on agreement between the attracting head and the moved phrase. Following Matushansky, I maintain that phrasal movement requires a $[uF]$ probe on the target head to Agree (subject to locality restrictions) with a constituent that it c-commands.

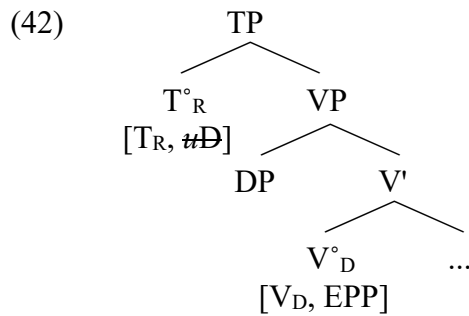
In addition, I propose that phrasal movement occurs only if the attracting head has the EPP property, defined as the ability to trigger Merge of an item that does not contain a subset of its category features. Specifically, uninterpretable [μ F] probes are checked if they c-command a suitable goal, but only trigger movement of the goal if the probing head also contains an [EPP] feature (Chomsky 2001).^{xvii} To account for the generalization that only dominant heads license phrasal specifiers, I propose that only lexical items that contain a dominant category feature can carry [EPP]. The necessary configuration for phrasal movement is shown in (41).



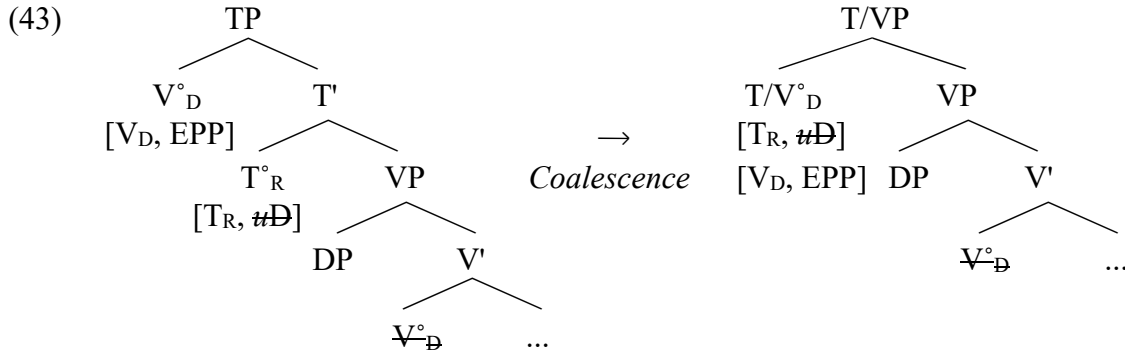
In addition, I claim that delayed gratification patterns motivate a reconception of the EPP property that diverges from most prior approaches. In these patterns, a probe associated with a (recessive) target head can trigger phrasal movement only after it has undergone Coalescence with a moved, dominant head. I account for this as follows: First, while the presence of [EPP] on a head enables its projection to have a phrasal specifier, no specifier is Merged if the head does not also contain a relevant [μ F] probe. Second, [EPP] features are not checked or disabled during the derivation. Third, because [EPP] is associated with dominant category features rather than individual probes, it can be associated with more than one probe during the derivation. Finally, I make the auxiliary assumption that checked uninterpretable features are not immediately deleted from the derivation

(Pesetsky and Torrego 2000). To preview the analysis of delayed gratification patterns, a [μ F] probe on a recessive head is first checked by c-commanding its goal. A dominant head with [EPP] then moves and undergoes Coalescence with the recessive head. This then enables the checked [$\#$ F] probe to trigger phrasal movement in concert with [EPP].

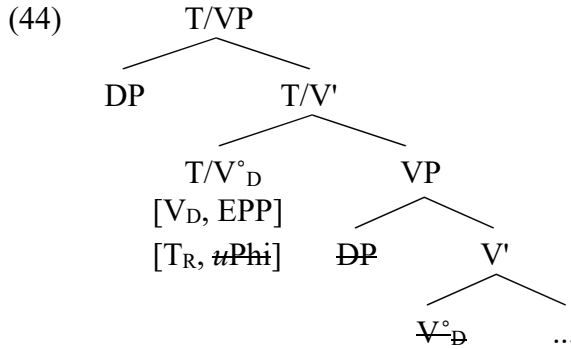
The basic workings of head movement, Coalescence, agreement, and phrasal movement are illustrated with a schematic example of Romance-style V-to-T head movement. In this and subsequent examples, I assume that all V° heads are dominant heads with [EPP], a reinterpretation of Baker's (2003) claim that the defining property of verbs as a lexical category is their ability to license specifiers.^{xviii} In the first step (42), a recessive T°_R with a [μ D] probe is Merged upon the completion of VP, which has a dominant V°_D and a subject DP in Spec, VP. At this point, [μ D] on T° is checked by agreement with the subject. However, phrasal movement of the subject does not occur because T° lacks [EPP].



Here, the Dominance Condition is satisfied by moving V°_D to Spec, TP (43). As dominant V°_D now c-commands recessive T°_R , Coalescence creates a bundled T/V° head.



The bundling of V° and T° into T/V° then enables the checked $[\#D]$ probe to make use of $[EPP]$ associated with $[V_D]$, triggering phrasal movement of the subject to Spec, TP.



In Section 4.1, inheritance of $[EPP]$ on V°_D by probes in the clausal left periphery is used to account for verb second patterns, and the flexible discourse interpretations of sentence-initial items in verb-second clauses (Fanselow 2009; Fanselow and Lenertová 2010).

Before concluding this section, I note that this analysis of head movement does not extend to proposed cases of *long head movement* where head movement skips a potential intervener $[X^\circ_i \dots [Y^\circ \dots [X^\circ_i \dots]]]$ (Rivero 1991; Borsley et al. 1996; Roberts 2010 and references therein). These patterns potentially arise under conditions where the head of the target projection has $[EPP]$ and a probe whose goal can be a head. For example, in Breton either a head or a phrase can move to the clause-initial position (Rivero 1991; Borsley et al. 1996; Roberts 2010).

- (45) a. Al levr en deus lennet Tom (Breton: Borsley et al. 1996)
the book 3SG.M has read Tom
‘Tom has read the book.’
- b. Lennet en deus Yann al levr
read 3SG.M has Yann the book
‘Yann has read the book.’

In the present proposal, Dominance Condition driven movement attracts the closest dominant head within the extended projection, and must be followed by Coalescence. There is indeed no evidence in these cases that the moved and target heads (e.g. the moved participle + auxiliary) can move together later in the derivation, as would be expected if they were bundled after movement.^{xix}

4 Case studies

This section illustrates key aspects of the proposals on Coalescence and movement, in the context of functional projections in the inflectional and complementizer domains. It focuses on cases in which languages permit a given head to have both dominant and recessive variants: the Kashmiri V2/V3 alternation, the analytic vs. synthetic past perfect in Catalan, and the realization of sentential negation in English.

4.1 *Relaxed Verb Second Effects*

To illustrate how Coalescence accounts for Feature Scattering analyses, I present a reinterpretation of Hsu’s (2017) analysis of verb second (V2) effects, which aims to generate a range of attested ‘strict’ and ‘relaxed’ V2 patterns. In well-known ‘strict’ V2 patterns like that of Standard German and Dutch, finite verbs move to a C-domain projection, and are preceded by a single phrase. Traditionally, these patterns are analyzed

as the result of verb movement from T° to C°, followed by movement of exactly one phrase to Spec, CP (den Besten 1983).

(46) [CP XP V-C° [TP ... V° ... ~~XP~~ ...

However, research in the cartographic program has produced a range of evidence that the traditional “CP” contains a series of projections associated with clause type and information structure features, as illustrated with the “core” structure of Rizzi (1997) in (47). If the expanded inventory of left-peripheral features is universally present, then the generation of strict V2 languages requires language-specific restrictions on the number of features that can be simultaneously instantiated in phrase structure.

(47) [ForceP ... [TopicP ... [FocusP ... [FinitenessP ... [TP

In addition, there are a variety of ‘relaxed’ V2 languages that show verb movement to the C-domain, but allow more than one phrase to precede the verb (either optionally or obligatorily). In these languages, the ordering of pre-verbal constituents is restricted based on discourse properties, consistent with the order in (47) (Poletto 2002; 2014; Benincà and Poletto 2004; Walkden 2017; Wolfe 2019). For example, Ingush, whose V2 pattern otherwise resembles that of Standard German (Nichols 2011), permits V3 orders in which a topic XP and focused XP precede the verb, in the order *Topic + Focus + V*.

(48) Ingush: Topic + Focus + V
 [Jurta jistie] [joaqa sag] ull cymogazh jolazh.
 town.GEN nearby AGR.old person lie.PRS sick.CVB.SIM AGR.PROG.CVB.SIM
 'In the next town an old woman is sick.' (Nichols 2011: 683)

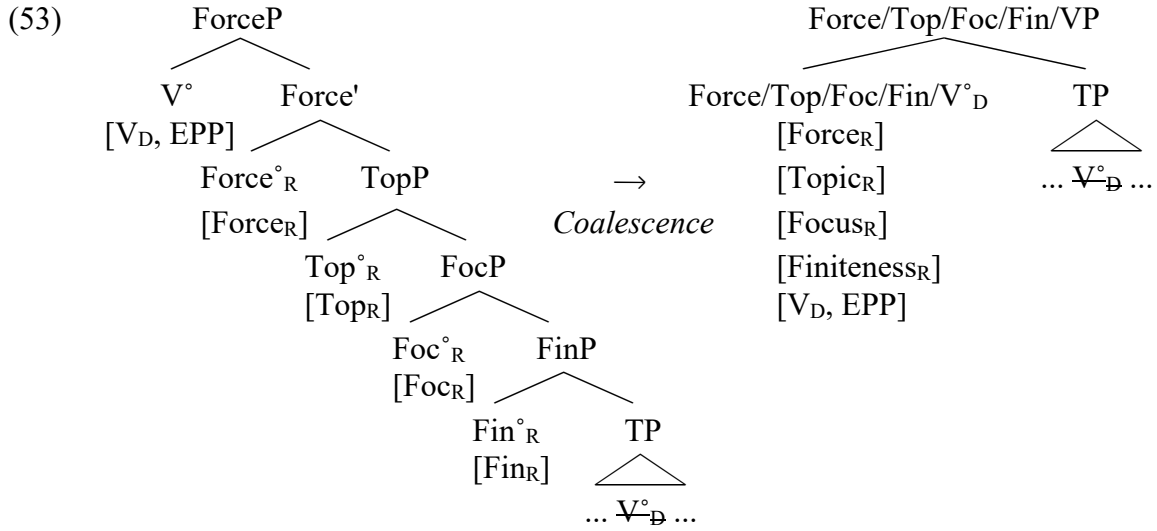
A further articulated structure is found in languages like Old Italian, which additionally permits frame-setting adverbials in the highest, clause-initial position.

- (49) Old Italian: Frame-setting adverb + topic + focus + V
 [e per volontà de le Virtudi] [tutta questa roba] [tra' poveri]
 and by will of the virtues all this stuff among poor.PL
 dispense.
 distribute
 'And according to the will of the virtues, distributed all these goods among the
 poor.' (Poletto 2014: 16)

Hsu (2017) proposes that the primary parameter that drives variation in the ‘strictness’ of V2 patterns is in the number of heads on which left-peripheral features are distributed. Strict V2 patterns are generated when all left-peripheral features are bundled on a single head, which attracts a single specifier (50). Relaxed V2 patterns (51)-(52) are generated when left-peripheral features are distributed across multiple heads, whose hierarchical order remains consistent with those proposed in strictly cartographic analyses.

- (50) V2 (German)
 $[_{\text{Force/TopFocus/FinP}} \text{XP} \text{ V-Force/Top/Foc/Fin}^\circ \text{ } [_{\text{TP}} \dots \text{V}^\circ]$
- (51) V3 *Topic Focus V* (Ingush)
 $[_{\text{Force/TopP}} \text{XP}_{\text{top}} \text{ } [_{\text{Foc/FinP}} (\text{XP}_{\text{foc}}) \text{ V-Foc/Fin}^\circ \text{ } [_{\text{TP}} \dots \text{V}^\circ]$
- (52) V4 *Frame-setter Topic Focus V* (Old Italian)
 $[_{\text{ForceP}} \text{XP}_{\text{frame}} \text{ } [_{\text{TopP}} (\text{XP}_{\text{foc}}) \text{ } [_{\text{Foc/FinP}} (\text{XP}_{\text{top}}) \text{ V-Foc/Fin}^\circ \text{ } [_{\text{TP}} \dots \text{V}^\circ]$

I now present a derivational analysis of these patterns in terms of Coalescence. To generate the idealized strict V2 pattern, suppose that all left-peripheral category features $[\text{Force}_R]$, $[\text{Topic}_R]$, $[\text{Focus}_R]$, $[\text{Finiteness}_R]$ are recessive. The only way to satisfy the Dominance Condition is to move V°_D to ForceP (53). This allows Coalescence to apply iteratively until all recessive features are bundled into a dominant head.

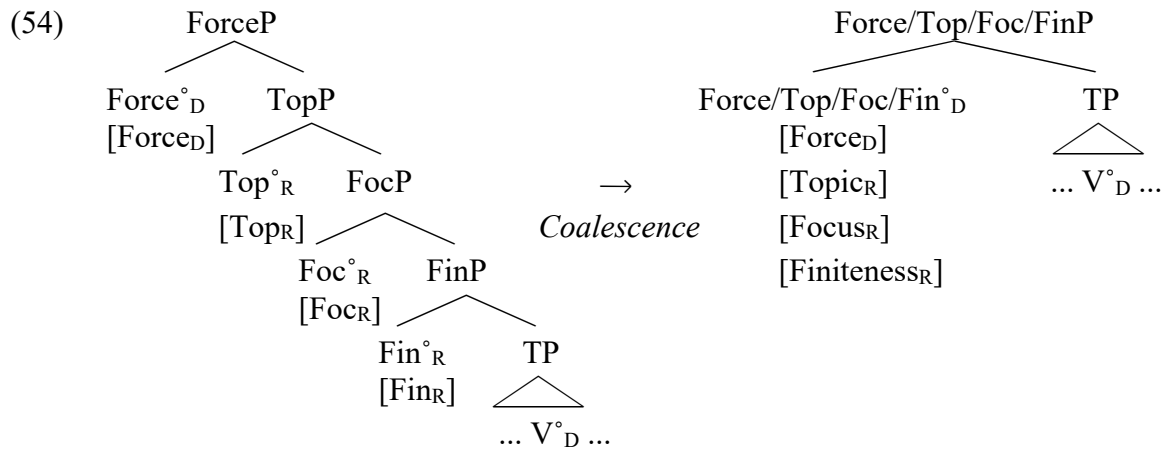


The association of the $[EPP]$ property with $[V_D]$, rather than a left-peripheral feature, accounts for the fact that subsequent phrasal movement to Spec, *ForceP* has several possible pragmatic functions. In Standard German, the first position can be occupied by either a given information topic, contrastive focus, or pragmatically unmarked subject (Mohr 2009). Suppose that German permits $[Topic_R]$ to be associated with a $[uTopic]$ probe, $[Focus_R]$ with $[uContrast]$, and $[Finiteness_R]$ with $[uD]$, and that each probe can be checked via Agree with a constituent lower in the clause.^{xx} However, phrasal movement occurs only if the probe is found on a dominant head with $[EPP]$, a situation that arises only due to verb movement. Assuming that there are no priority restrictions in German on which probe triggers movement in concert with $[EPP]$, either focus, topicalization, or subject movement can take place in the final step, thus deriving the “unrestricted edge feature” property of verb second clauses (Fanselow 2009; Fanselow and Lenertová 2010).

As a brief aside, this analysis of movement to first position as the result of $[EPP]$ on *V°* makes an additional prediction about word order in nominal extended projections. The

translation of Baker’s (2003) argument that only verbs license specifiers into the claim that V°_D heads have [EPP] also implies that N° heads lack [EPP], even if they are dominant. Consequently, even if N-to- D° movement is possible, as is likely the case for languages with noun-initial DP order, we do not expect to find “noun second” patterns in which nouns are always preceded by exactly one phrase within DP, because N° never has [EPP]. To my knowledge, such patterns are unattested, as predicted.

Turning back to verb second structures, suppose that rather than moving V°_D , a dominant Force°_D head can be externally Merged (54). This allows Coalescence to apply, obviating verb movement out of TP. Even if [Topic_R], [Focus_R], and [Finiteness_R] are associated with the same probes, agreement cannot trigger phrasal movement, because [Force_D] is not associated with [EPP].

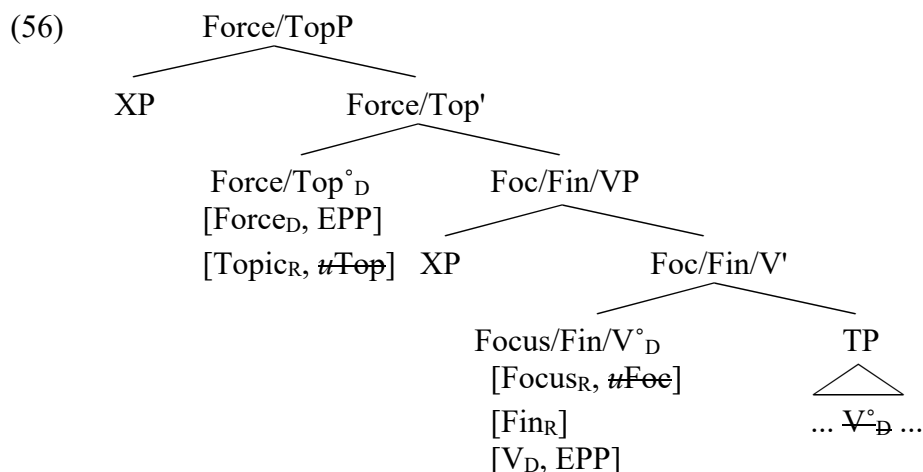


The availability of both internal Merge and external Merge options to satisfy the Dominance Condition accounts for asymmetric V2 patterns in which embedded clauses can either contain an overt complementizer or show V2 order. For example, in Standard German *dass* realizes a dominant, externally Merged Force°_D head that lacks [EPP]. V2 clauses are generated by movement of V°_D , with subsequent phrasal movement to first

position permitted by that head's [EPP] feature.^{xxi} V2 patterns thus represent a delayed gratification effect in that C-domain probes are only able to trigger phrasal movement after head movement has taken place.

- (55) a. Er sagte [dass_C er morgen kommt_T]
 He said that he tomorrow comes
- b. Er sagte [er kommt_{T+C} er morgen ~~kommt~~]
 He said he comes tomorrow
 'He said that he is coming tomorrow.'
- c. *Er sagte [dass er kommt morgen]
 He said that he comes tomorrow
- d. *Er sagte [er dass kommt morgen]
 He said he that comes tomorrow
- (Holmberg 2015 after den Besten 1983)

As the number of dominant category features increases, fewer applications of Coalescence take place, leaving a greater number of heads in the left periphery. The Ingush V3 pattern in which topics and foci can simultaneously precede the verb arises when Force[°]_D is dominant, and [Topic] and [Focus] are realized on separate heads.



Note that verb movement occurs before Topic_R is Merged, and the question arises as to why the derivation cannot wait until Force_D is Merged before Coalescence bundles all left-peripheral heads, as is possible in German. As a preliminary explanation, suppose that the Dominance Condition is enforced on projections that constitute phasal spell-out domains (cf. Section 5.1). Further assuming that languages can vary as to whether a given projection is a phase (Abels 2003), the projection containing [Topic] is phasal in Ingush, but not in German. This subjects FocusP to the Dominance Condition in Ingush, compelling Last Resort verb movement. The proposal remains tentative, however, given remaining uncertainty on how phasehood is best defined and diagnosed.

Further support for the distinction between dominant and recessive heads is found in languages that contain both dominant and recessive variants of the same feature. This accounts for an otherwise puzzling V2/V3 alternation in Kashmiri. Kashmiri has a strict V2 requirement in declarative main clauses, in which no more than one preverbal phrase is permitted (Bhatt 1999; Munshi and Bhatt 2009; Manetta 2011).

- (57) a. laRk-as dyut rameshan raath kalam
 boy gave Ramesh yesterday pen
 'It was a boy to whom Ramesh gave a pen yesterday'
- b. *tem raath dyut akh laRk-as kalam
 he yesterday gave one boy pen (Bhatt 1999)

There is some inconsistency in descriptions of the information structure characteristics of Kashmiri V2. According to Bhatt (1999) and Manetta (2011), non-subjects in the first position of declarative clauses must be focused, not topicalized. On the other hand, Munshi and Bhatt (2009) report that first-position non-subjects can be either topics or foci, though the interpretations must be distinguished by intonational contour.

In light of these patterns, the word order of *wh*-questions in Kashmiri is initially puzzling. As is typical of languages with V2, *wh*-phrases obligatorily precede finite verbs. However, Bhatt (1999) and Manetta (2011) report that if the clause contains eligible topics, one of them preferably precedes the *wh*-phrase. This is unexpected both because it appears to be an obligatory deviation from strict V2, and because topicalization to a preverbal position is not possible in declarative clauses in their descriptions.

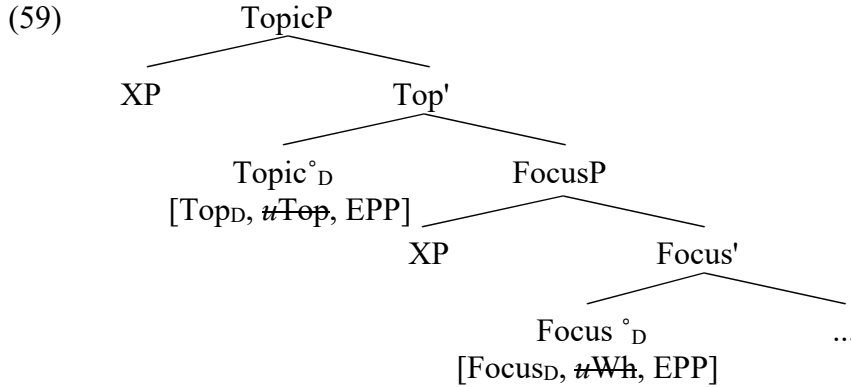
- (58) a. tse kyaa dyutnay Rameshan
 you what gave Ramesh
 'As for you, what is it that Ramesh gave?'
 b. ?kyaa dyutnay Rameshan tse
 what gave Ramesh you (Bhatt 1999)

Munshi and Bhatt (2009) similarly note the acceptability of [XP_{topic} XP_{wh} V ...] order in questions, but describe topicalization to first position as optional, rather than obligatory. While I cannot explain the source of disagreement between these descriptions, I proceed with a working assumption that they are correct descriptions of related but distinct grammars, and describe how each system can be accounted for in the present proposal.

At first glance, the Kashmiri V2/V3 alternation is puzzling in that the realization of the high topic position depends on the presence of a *wh*-phrase. However, it is straightforwardly accounted for if Kashmiri has two types of Focus heads – a dominant version with a [uWh] probe in interrogative clauses, and a recessive version with a [$uContrast$] probe in declarative clauses. In contrast, for languages like Ingush, where V3 is available in both declaratives and interrogatives, [Focus] is always dominant.

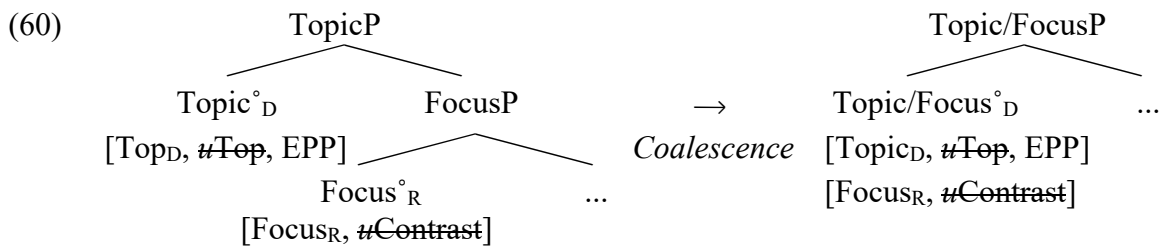
Restricting our attention to the structural realization of topic and focus, consider how the derivation proceeds if the Numeration contains the dominant [uWh] Focus° head.

Since both Topic° and Focus° are dominant as Merged, Coalescence cannot apply, leaving $[\text{Topic}]$ and $[\text{Focus}]$ in separate projections. Because both heads contain $[\text{EPP}]$, this creates the V3 word order in interrogative clauses.



The distinction between the pattern described by Bhatt (1999) and Manetta (2011) versus that of Munshi and Bhatt (2009) can be understood as variation in whether a $[\text{uTop}]$ probe is always associated with the $[\text{Top}_D]$ categorial feature. In the first pattern, $[\text{uTop}]$ is always present, triggering phrasal movement in concert with EPP. Alternatively, topic movement is optional in varieties in which $[\text{uTop}]$ can be omitted.

In Kashmiri declarative clauses, the Focus head with a $[\text{uContrast}]$ probe is recessive. Coalescence applies once Topic°_D head is Merged. This bundles the $[\text{Topic}]$ and $[\text{Focus}]$ features into a single projection, leaving only one position available for movement.



In this case, the contrast between varieties that permit either a topic or a focus in first position and those that permit only a focus in first position results from a difference in the

priority with which probing features in the bundled head can trigger phrasal movement in conjunction with [EPP]. In the former case, either [*u*Top] or [*u*Contrast] can trigger phrasal movement, whereas in the latter variety [*u*Contrast] must take precedence over [*u*Top], even if [EPP] is originally associated with [Topic_D].

4.2 English negative contraction

English negation is another case where a head has both dominant and recessive varieties. English has a 'full' negative morpheme (orthographic *not*) and a contracted form (orthographic *n't*). In many contexts, the two forms appear to be in free variation, with the contracted form apparently derived by optional phonological reduction.

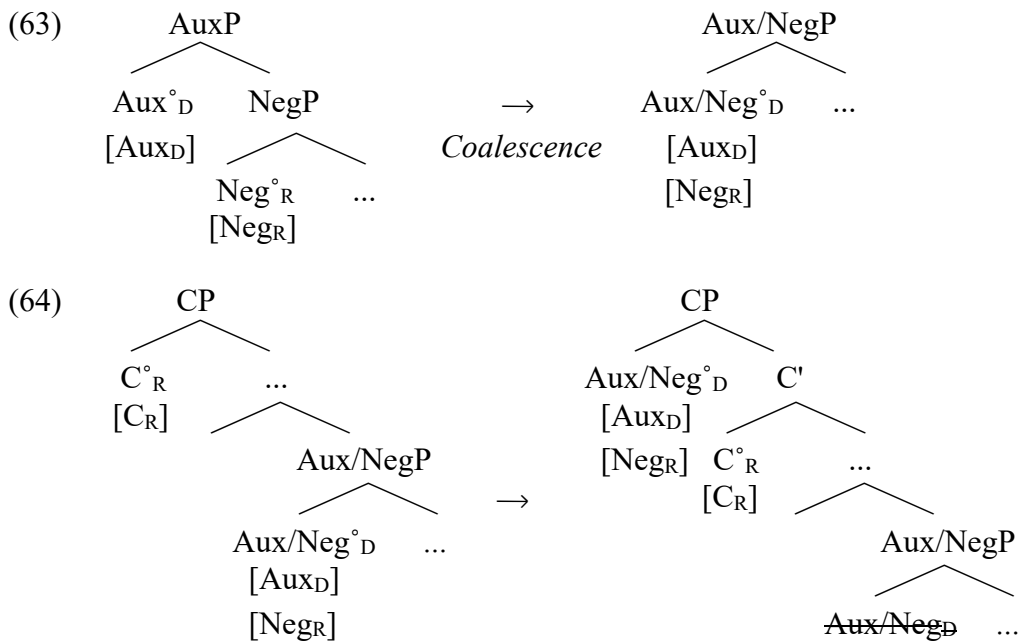
- (61) a. Michael did not make a mistake.
b. Michael didn't make a mistake.

However, the distribution of the two forms is constrained by syntactic factors, and the use of a particular form is obligatory in certain contexts (Zwicky and Pullum 1983). For example, consider the case of negative inversion. In English, auxiliary verbs raise to a pre-subject position in interrogative contexts. If the negation morpheme raises along with the auxiliary, use of the affixal form is obligatory (62a). This gives the effect of contraction feeding raising. On the other hand, only the full form is possible if the negative remains in a post-subject position (62b). Under an approach where the affixal form is derived by an operation that applies after the syntactic derivation, the obligatory use of the affixal negative when it raises with the auxiliary is unexpected.

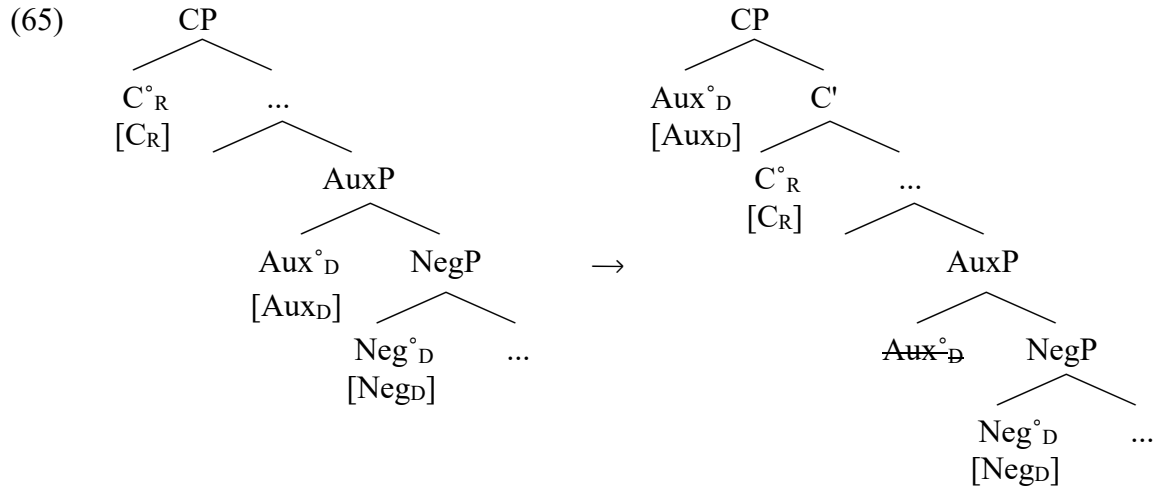
- (62) a. Didn't Lindsay host the gala? (cf. *Did not Lindsay host the gala?)
b. Did Lindsay not host the gala? (cf. *Did Lindsay n't host the gala?)

Matushansky (2006) makes the key observation that the distribution of the full and contracted forms is explained if contracted negation is formed by head-bundling during the derivation. Specifically, she proposes that Neg° and Aux° optionally undergo M-Merger once they are Merged, and that M-Merged Neg° corresponds to *n't*. If Aux° and Neg° are M-Merged, both negation and the auxiliary undergo movement together when Aux° is attracted to C° . If M-Merger does not apply, the auxiliary moves alone.^{xxii}

This analysis can be reframed using Coalescence as follows. I first illustrate the basic analysis with a structure in which Aux-to-T movement has already applied, and Aux° immediately c-commands Neg° . The different distribution of the full and contracted forms is accounted for if the full form enters the derivation with a dominant category feature $[\text{Neg}_D]$, while the contracted form is first Merged with a recessive feature $[\text{Neg}_R]$. Recessive Neg°_R undergoes Coalescence with the dominant auxiliary (63). Thus, when recessive C° attracts the closest dominant head, it attracts bundled $\text{Aux}/\text{Neg}^\circ$ (64).



On the other hand, dominant Neg°_D does not undergo Coalescence. In inversion contexts, interrogative C° attracts Aux_D° , leaving Neg_D° in its first-Merge position (65).^{xxiii}



It is noteworthy that negative contraction differs in key ways from auxiliary reduction (Zwicky 1970; Anderson 2008), which affects auxiliary forms *is*, *has*, *would*, *had*, *have*, *am*, *are*, and *will*. First, reduced auxiliaries do not form syntactic units with their hosts; they do not move with elements that they are affixed to.

- (66) a. Who's going to Phoenix?
b. *Who's do you think ~~who's~~ going to Phoenix?

Whereas contracted negation only follows tensed auxiliaries, reduced auxiliaries *is*, *has*, *would*, *had*, are less restricted by the category or phrase structure status of items that precede them, e.g. *no touching's allowed by the guards* / *the role that he auditioned for's been written out*. On the basis of these differences, Zwicky and Pullum (1983) categorize reduced auxiliaries as *simple clitics*, phonologically reduced variants that occur in the same locations as corresponding full forms, and *n't* as an *inflectional affix* that forms a morpho-syntactic constituent with its host.

In the present proposal, the affix-like properties (in Zwicky and Pullum's terms) of contracted negation, a high degree of selection and morpho-syntactic grouping with its host, follow from being the product of Coalescence. Contracted negation raises with auxiliaries because the corresponding heads are bundled during the derivation, and its restriction to a tensed auxiliary host results from the head-adjacent configuration of Neg_R° and $\text{Aux}/\text{T}_D^\circ$. In contrast, I analyze auxiliary reduction as an optional post-syntactic process, which accounts for its low degree of host selection and lack of evidence for morpho-syntactic constituency with its host.

Finally, these contrasts between contracted negation and auxiliary reduction highlight the impossibility of relying on surface exponence alone to determine the dominance or recessiveness of a category feature. However, syntactic diagnostics suggest that contracted auxiliaries correspond to dominant heads. First, the highest auxiliary in the clause head-moves to TP, and enables subject movement. Furthermore, the possibility of quantifier float with all auxiliaries in a sequence (Sportiche 1988), even when reduced, indicates that all auxiliary heads carry [EPP], a unique property of dominant heads.

4.3 The Catalan Perfect

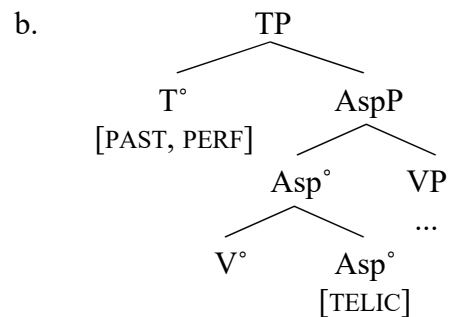
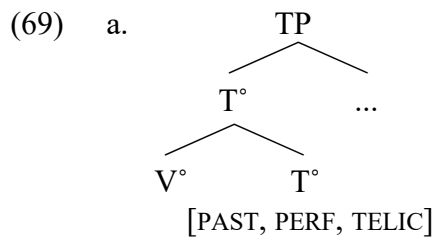
Here, we consider another case in which a language varies in whether Coalescence is fed by external Merge or movement. As described by Oltra-Massuet (2013), some dialects of Catalan express the past perfect either in a synthetic form where subject agreement, tense and aspect are realized as suffixes to a lexical verb (67), or an analytic form where subject agreement is realized on an auxiliary *anar* 'to go' followed by a participle that resembles the infinitive form (68).

(67) purific-ares
purify-2SG.PST.PERF
'you purified'

(68) vas purificar
AUX.2SG purify
'you purified'

An unusual and important property about the alternation is that there is no apparent semantic difference between the two ways of forming the past perfect. According to Oltra-Massuet (2013; 1), "these forms do not express different lexical or truth-conditional semantics, nor do they show different morpho-syntactic functions, and individual speakers use some subset of them without distinction." Variation both within and across speakers depends on the lexical items and conjugations used; speakers do not probabilistically use both forms for any given verb and conjugation pair.

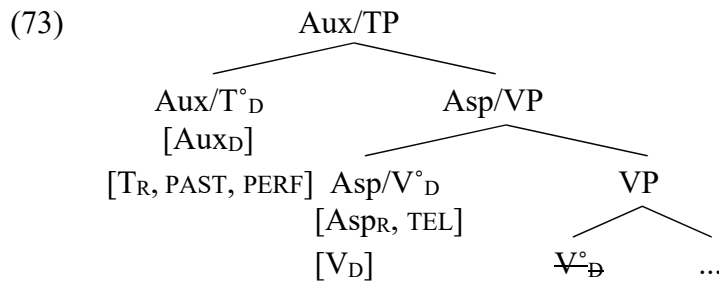
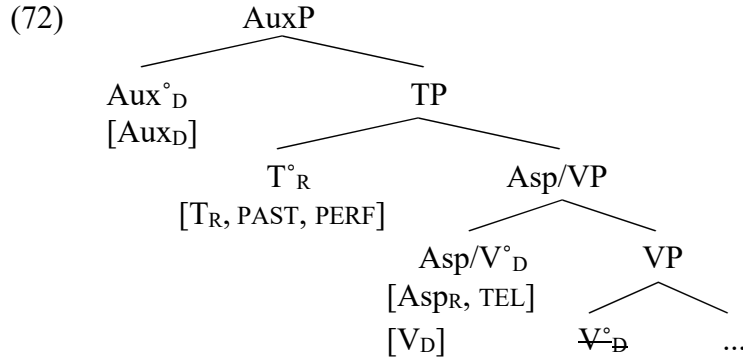
Oltra-Massuet's analysis appeals to both head movement and bundling distinction. In brief, synthetic forms (ex. *purificares*) are generated by verb movement to a T° head that carries specifications for past tense, perfective aspect, and telicity (69a). In derivations with analytic forms (ex. *vas purificar*), the features [PAST, PERF] are associated with T° , but [TELIC] is on a separate Aspect $^\circ$ head. Verb movement stops at Asp° , and V- Asp° is pronounced as the participle (69b). T° is obligatorily a suffix, thus triggering the insertion of an *anar* auxiliary that supports tense and subject inflection.



In my proposal, the distinction between the synthetic and analytic forms need not result from the presyntactic packaging of tense and aspect features. Rather, category features [Tense] and [Asp] are separately Merged on recessive heads in all derivations. The synthetic and analytic forms differ in whether Coalescence is fed by external or internal Merge. In the derivation for the synthetic form, V°_D moves to T°_R , and Coalescence bundles V°_D with T°_R and Asp°_R .^{xxiv}

Suppose that in these dialects of Catalan, verb movement can also proceed only as far as Spec, AspP. At this point, Coalescence bundles V°_D and Asp°_R . Note that even though V°_D has [EPP], subject movement to Asp/VP will not take place if [Asp_R] lacks a subject agreement probe. Indeed, there is no agreement morphology on the participle to suggest , and *auxiliary-subject-infinitive participle* orders are ungrammatical.

In the next steps, T°_R is Merged, followed by a dominant auxiliary head corresponding to *anar* (72). I will remain agnostic as to the categorial status of the auxiliary, labeling it simply as [Aux]. Coalescence then bundles Aux°_D and T°_R (73).



The analysis raises the question of how the derivation “knows” which option is used after T°_R is Merged, given that the choice is uniquely determined by the verbal predicate. As a tentative solution, I propose that individual predicates can potentially introduce more than one corresponding object in the numeration. Concretely, predicates in these dialects are specified for whether or not they license an auxiliary head if the numeration also contains $[T_R, PAST, PERF]$. The selection of an “analytic” predicate leads to the inclusion of V°_D and Aux°_D in the numeration, while selection of a “synthetic” predicate results in the inclusion of V°_D alone. Under the assumption that Aux°_D must be Merged if it is available, its selection obviates the movement of Asp/V°_D .^{xxv}

5 PF transfer and morphological realization

This section discusses remaining issues related to the transfer of structures created by Coalescence, primarily in head movement, to the PF interface. Section 5.1 addresses the question of predicting when Coalescence must take place within a derivation, before turning to the morphological realization of the proposed structures. Section 5.2 considers the determination of affix ordering in the absence of branching structure within featurally complex heads. Section 5.3 discusses cases where postsyntactic rules on lexical insertion can result in the non-realization of heads, or the apparent displacement of their exponents.

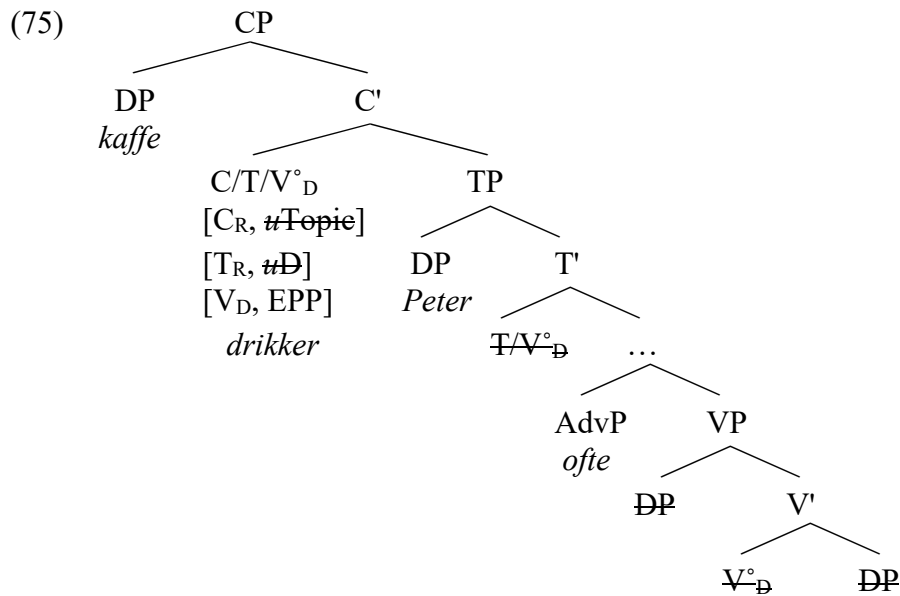
5.1 The timing of Coalescence

I have proposed that the Dominance Condition is a well-formedness requirement on syntactic structures that is evaluated when they are transferred to PF. In the context of phase theory (Chomsky 2000), this makes clear predictions about when the Dominance Condition should apply. Assuming that the complements of phase heads are transferred to PF upon completion of the phasal projection, we predict that all complement projections of phase heads must contain only dominant heads. As a concrete illustration, under common assumptions, C° is a phase head that spells out its complement, TP.^{xxvi} We thus expect TP (or the highest head in the inflectional domain) to obligatorily have a dominant head. This prediction is borne out in the English and Catalan cases – recessive T°_R triggers Last Resort head movement of a dominant head.

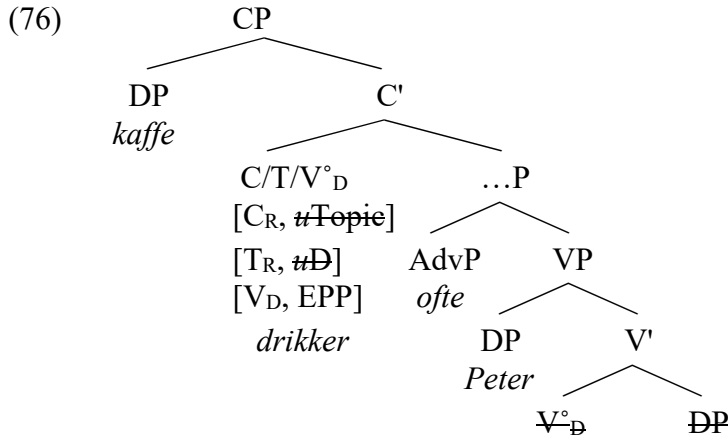
Here, we consider another case where phasal spell-out makes desirable predictions on the timing of head movement and Coalescence. Consider the Danish verb-second clause below, returning to a simplified structure where all left-peripheral features are in a single

C° head (74). Both CP and TP have specifiers (topic *kaffe* and subject *Peter* respectively), and the verb precedes the subject. As expected from the previous proposals, this is accounted for if V°_D first moves to and Coalesces with T°_R, licensing movement of the subject to Spec, TP. Bundled T/V°_D then moves to C°_R and undergoes Coalescence, enabling the [μ Topic] probe to trigger phrasal movement of the topic *kaffe*.

- (74) [_{CP} Kaffe drikker [_{TP} Peter ofte [_{VP} om morgonen]]
 Coffee drinks Peter often in morning
 ‘Coffee, Peter often drinks in the morning’



However, consider an alternative derivation in which head movement of V°_D takes place after T°_R and C°_R have been Merged, and Coalescence creates a single head with features of T°_R and C°_R. Assuming that [EPP] can trigger phrasal movement in concert with either the [μ D] or [μ Topic] probes, we predict the generation of structures like (76), in which the topic moves to Spec, C/TP while the subject *Peter* remains within VP. However, placement of subjects below VP adverbials is ungrammatical in Danish.



This derivation can be ruled out as follows. Suppose that spell-out applies to the complement of the head that contains [C] when it ceases to project. There is then no prohibition against the creation of a $C/T/V^\circ_D$ head, as long as its complement projection has a dominant head. Rather, the impossibility of moving a topic to Spec, CP while stranding the subject in VP is due to a separate restriction that requires [$\#D$] to take precedence over [$\#Topic$] in associating with [EPP] (recall that the Kashmiri V3 analysis requires a similar relative priority of probes). Topicalization is thus only possible if C and T are realized as separate projections, each the target of verb movement. This recasts in some ways the idea that subject-initial V2 clauses contain fewer projections than V2 clauses with non-subjects in first position (Travis 1984; Zwart 1997). While further pursuit of this hypothesis is outside of the scope of this paper, it is a promising avenue for further examination. I thank an anonymous reviewer for inspiring this discussion.

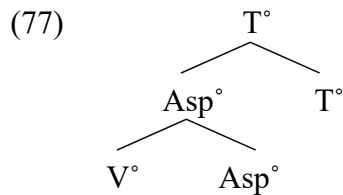
5.2 Affix ordering

A key empirical strength of “classic” head adjunction structures with head-internal branching is its ability to account for affix ordering. It permits a simple mapping from syntactic structure to its morphological realization: the association of phonological content

with syntactic representations, *lexical insertion*, targets terminal nodes of the tree.

Furthermore, the generation of head adjunction structures during the syntactic derivation derives the Mirror Principle Generalization: that affix order reflects the hierarchical orderings of their corresponding projections (Baker 1985).

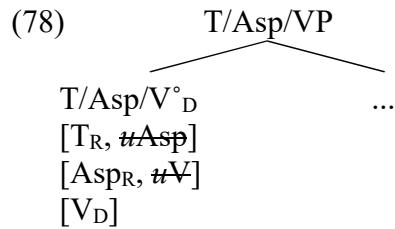
As an illustration of a standard analysis of head movement, cyclic head movement of V° through the Aspect° and Tense° heads produces the complex head (77), with the projecting head at each level on the right (Williams 1981).



If vocabulary items are inserted directly the terminal nodes, both aspect and tense are realized morphologically as verbal suffixes. Prefixes can be generated by positing a postsyntactic operation like Local Dislocation (Embick and Noyer 2001), which alters the linear order of sister nodes prior to vocabulary insertion. If Local Dislocation applies between sister nodes of complex heads from the bottom up, $V/\text{Asp}/\text{T}^\circ$ can be realized as $[[V-\text{Asp}]-\text{T}]$, $[[\text{Asp}-V]-\text{T}]$, $[\text{T}-[V-\text{Asp}]]$, $[\text{T}-[\text{Asp}-V]]$. It is not possible to generate Mirror Principle-violating orders $*[\text{Asp}-\text{T}-V]$ and $*[V-\text{T}-\text{Asp}]$ (Harley 2013).^{xxvii}

I have proposed that Coalescence produces a single terminal node that contains the features of each of the bundled heads. While it is possible to formulate vocabulary insertion rules that target bundles of features (Anderson 1982; 1992), this view nonetheless requires an alternative way of predicting the Mirror Principle generalization on affix order, no longer directly accessible from the syntactic constituent structure.

I propose that the Mirror Principle can be understood as a preference in the PF grammar for affix order to reflect the derivational history of a head. Alternatively put, the order of affixes reflects the order in which their corresponding heads enter the derivation via external Merge. This is possible if c-selection features are not deleted from the syntactic representation upon checking, and thus accessible to affix linearization operations. This is consistent with the analysis of delayed EPP patterns in Section 4, which also requires checked probes to be accessible at later stages of the derivation. I illustrate briefly with the simplified clause structure $[_{TP} [_{AspP} [_{VP} \dots]]]$. External Merge of each functional head is triggered by c-selectional features $[uV]$ on Asp° and $[uAsp]$ on T° . Subsequent head movement of V° and Coalescence creates the bundled head in (78).



The Mirror Principle can thus be stated as a preference for linearization to reflect c-selection relations among features of a head, such that the exponent of a feature must be equidistant or closer to the word edge than the exponent of a feature that it c-selects.

The key here is that the absence of branching structure does not preclude an account of the Mirror Principle generalization. However, there are reasonable concerns with the approach. Although there is an independent conceptual necessity for linearization procedures to refer to branching structure, it is less clear why c-selection relations must also be accessible (though it is plausibly driven by a preference for transparent scope relations). My proposal also does not adhere as strictly to the desideratum of

informational encapsulation, in that one type of feature is visible to two modules of the grammar. While these issues must ultimately be addressed, they must also be weighed against the aforementioned theoretical benefits of eliminating head-internal branching from syntactic representations.

Finally, it is worth noting that I have focused on movement and bundling patterns in the complementizer and inflectional domains of the clause. Important questions remain about whether the approach can be extended to syntactic approaches to derivational morphology, in which acategorical roots combine with category-defining heads (Marantz 2007; Embick and Marantz 2008). In particular, Coalescence might not be expected to apply to structures with more than one category-defining head like *gloriousness* [*n* [*a* [$\sqrt{\text{GLORY}}$]]] (Embick and Marantz 2008), which patterns in syntax as one head noun despite containing category features [*n*], [*a*] that do not seem to be in a subset relation.^{xxviii}

5.3 Lexical insertion in extended projections

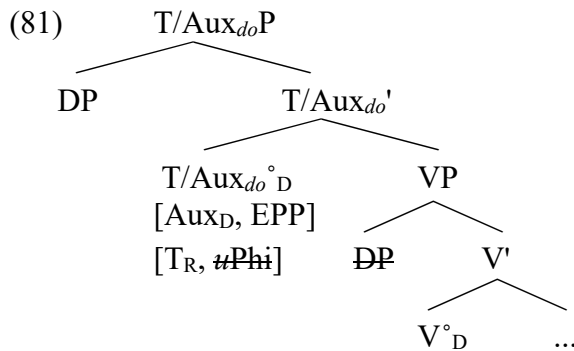
There are strong arguments that the linear order of morphemes is not determined by syntactic derivations alone, and that postsyntactic PF operations can affect word order in restricted ways (Marantz 1984; Embick and Noyer 2001; Svenonius 2016; but cf. Kayne 2005b; Koopman 2017). While this adds some complexity to analyses, particularly when “lowering” or “affix hopping” affects a head with EPP, the main proposals of this paper can be maintained. To illustrate, consider the patterning of English clauses that lack a modal, aspectual, or passive auxiliary. In declarative clauses without negation, verbs remain in VP and carry tense features as suffixes (79). In *do*-support contexts (sentential negation, questions, ellipsis), an auxiliary *do* appears in T with tense suffixes (80).

(79) [TP Lucille [often [VP sees George]

(80) [TP Lucille does [NegP not [often [VP see George]

This raises several questions in context of this paper. First, in order for subjects to be in Spec, TP, T° must be a dominant head with [EPP] in all clauses. It then remains to be explained why T° is sometimes unpronounced, and sometimes realized as *do*. Second, why are tense features pronounced on a lower head when *do* is not present?

These issues can be accounted for by positing that the structures created by syntax are subject to language-particular pronunciation and linearization principles. To illustrate with the English case, I claim that (79) and (80) both contain a dominant auxiliary (labeled Aux_{do}°) as the head of TP (cf. Emonds 1970; Pollock 1989; Wilder and Ćavar 1994 for similar null Aux or null *do* proposals). I posit that like other auxiliaries, Aux_{do}° first head-moves and undergoes Coalescence with T°_R (81).



I make use of the insight that the realization of English tense is conditioned by the structural relation between T° and VP (Embick and Noyer 2001; Adger 2003); *do*-support occurs uniquely in contexts in which VP is not the sister of T° . Assuming that in *do*-support contexts Aux_{do}° moves to T° and acquires tense features, I will use the following informally stated pronunciation rules for Aux_{do} :

- (82) If VP is the complement of Aux_{do}° , pronounce all features of Aux_{do}° on V.
 If VP is not the complement of Aux_{do}° , pronounce Aux° as *do*, with all of its included features as suffixes.

I note two advantages of this analysis of *do* as a sometimes-null auxiliary head. First, in Embick and Noyer (2001), *do* is the realization of a head that is Merged onto T° if T° lacks a VP complement and no auxiliary moves to T° . However, this triggering mechanism is unorthodox and difficult to implement given the guiding assumption in Minimalism that Merge is driven by features. This issue is avoided if Aux_{do}° is present in all clauses. Second, the absence of *do* in clauses with auxiliaries is accounted for if Aux_{do}° is Merged below other auxiliaries; Aux_{do}° does not move to TP in clauses that contain another (dominant) auxiliary, and is unpronounced due to adjacency with VP. As a further empirical advantage, this accounts for dialects where *do*-support occurs with auxiliaries in ellipsis, e.g. *George will have escaped to Mexico, and Buster will have done too*. This placement of *do* below auxiliaries is not predicted if *do* is inserted in T° .

In a theory in which postsyntactic processes can lead to the non-realization or linear displacement of syntactic heads, it is a non-trivial task to determine which kinds of displacements are generated by syntax, versus postsyntactic operations. I have discussed two types of patterns, feeding relations between head movement and phrasal movement, and the existence of heads that contain probes associated with multiple category features, as key evidence for a syntactic approach to head movement and head bundling.

In this context, it is worth discussing an emerging view of the lexical insertion of heads: Spanning (Svenonius 2012; 2016). In Spanning theory, lexical insertion targets the full sequence of heads in an extended projection (rather than terminal nodes, as assumed

in Distributed Morphology), also called *spans*.^{xxix} Word order variation arises from two parameters: the number of lexical items inserted in a given span, and where they are linearized relative to specifiers, shown by diacritic @ (Svenonius 2016; based on Brody 2000). Specifiers of projections higher than @ are linearized to the left of the head, while those lower than @ are linearized to the right. To illustrate, French and English do not differ in head movement paths of verbs in clauses without auxiliaries – rather, the verbal span is linearized in T in French (T@-v-V), but v in English (T-v@-V).

Although Spanning is presented as a means of eliminating head movement as a syntactic operation, it faces two key empirical challenges. First, because @ is only a diacritic for linearization, the approach cannot directly account for instances of head movement with semantic effects. Svenonius (2016; 213) suggests that these semantic properties can be attributed to @; alternatively, semantic features could restrict possible placement of @. Second, it is difficult to account for delayed gratification patterns, in which the placement of specifiers is dependent on the positions of heads. This is because @ determines where heads are linearized relative to specifiers in span, which is an independent parameter of whether individual projections license specifiers. Accounting for delayed gratification patterns would require ad hoc restriction that some projections only have @ if they can host a specifier. In contrast, these patterns are generated in syntax by the proposed feature system without such stipulation.

6 Conclusion

In this paper, I have argued that a variety of bundling processes that affect heads should be understood as the result of a single syntactic operation, Coalescence. Although it

presents a non-trivial addition to the set of syntactic operations assumed in standard Minimalism, it permits a unified analysis of bundling patterns in Feature Scattering phenomena and head movement, which in prior approaches have been attributed to separate operations in the lexicon and in a postsyntactic module. I have argued that this accounts for key structural similarities between the two patterns, and for derivations in which there is a feeding relationship between head bundling and phrasal movement. Furthermore, application of Coalescence to head movement avoids the main problems posed by traditional head-adjunction models in Minimalism, including violations of the Extension and Uniformity Conditions, and defining a featural trigger for head movement.

References

- Abels, Klaus. 2003. Successive cyclicity, anti-locality, and adposition stranding. Doctoral dissertation, University of Connecticut.
- Aboh, Enoch Oladé. 2006. Complementation in Saramaccan and Gungbe: the case of C-type modal particles. *Natural Language and Linguistic Theory* 24: 1–55.
- Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford: Oxford University Press.
- Anderson, Stephen R. 1982. Where's morphology? *Linguistic Inquiry* 13: 571–612.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Anderson, Stephen R. 2008. English reduced auxiliaries really are simple clitics. *Lingue e Linguaggio* 7: 169–186.
- Arnott, D. W. 1970. *The nominal and verbal systems of Fula*. Oxford: Clarendon Press.
- Arregi, Karlos, and Asia Pietraszko. 2018. Generalized head movement. *Proceedings of the Linguistic Society of America* 3: 5:1–15.
- Baker, Mark C. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16: 373–415.
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Baker, Mark C. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.
- Baker, Mark C. 2003. *Lexical categories*. Cambridge: Cambridge University Press.
- Baunaz, Laura, Liliane Haegeman, Karen De Clercq, and Eric Lander, ed. 2018. *Exploring Nanosyntax*. Oxford: Oxford University Press.
- Benincà, Paola, and Cecilia Poletto. 2004. Topic, focus, and V2. In *The Structure of CP and IP*, ed. by Luigi Rizzi, 52–75. Oxford: Oxford University Press.

- Bennett, Wm. G, Akinbiyi Akinlabi, and Bruce Connell. 2012. Two subject asymmetries in Defaka focus constructions. In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, ed. by Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman, 294–302. Somerville, MA: Cascadilla Proceedings Project.
- den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. In *On the formal syntax of the Westgermania*, ed. by Werner Abraham, 47–131. Amsterdam: John Benjamins.
- Bhatt, Rakesh. 1999. *Verb movement and the syntax of Kashmiri*. Dordrecht: Kluwer.
- Bobaljik, Jonathan David. 1995. Morphosyntax: The syntax of verbal inflection. Doctoral dissertation, Massachusetts Institute of Technology.
- Bobaljik, Jonathan David, and Höskuldur Thráinsson. 1998. Two heads aren't always better than one. *Syntax* 1: 37–71.
- Boeckx, Cedric, and Sandra Stjepanovic. 2001. Head-ing toward PF. *Linguistic Inquiry* 32: 345–355.
- Borer, Hagit. 1984. *Parametric syntax: case studies in Semitic and Romance languages*. Dordrecht: Foris.
- Borsley, Robert D., Maria-Luisa Rivero, and Janig Stephens. 1996. Long head movement in Breton. In *The syntax of the Celtic languages: a comparative perspective*, ed. by Robert Borsley and Ian G. Roberts, 53–74. Cambridge: Cambridge University Press.
- Brody, Michael. 2000. Word order, restructuring and Mirror Theory. In *Derivation of VO and OV*, ed. by Peter Svenonius, 27–43. Amsterdam/Philadelphia: John Benjamins.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20*, ed. by Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1994. Bare phrase structure. *MIT Occasional Papers in Linguistics* 5.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by Step: essays on Minimalist syntax in honour of Howard Lasnik*, ed. by Roger Martin, David Michaels, Juan Uriagereka, and Samuel Jay Keyser, 89–115. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory*, ed. by Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1990. *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford: Oxford University Press.
- Cinque, Guglielmo, and Luigi Rizzi. 2009. The cartography of syntactic structures. In *The Oxford handbook of linguistic analysis*, ed. by Bernd Heine and Heiko Narrog, 1–21. Oxford: Oxford University Press.
- Coon, Jessica, and Alan Bale. 2014. The interaction of person and number in Mi'gmaq. *Nordlyd* 40: 85–101.
- Cowper, Elizabeth. 2005. The geometry of interpretable features: Infl in English and Spanish. *Language* 81: 10–46.

- Dékány, Éva. 2018. Approaches to head movement: a critical assessment. *Glossa* 3: 65.
- Delsing, Lars-Olof. 1993. The internal structure of noun phrases in the Scandinavian languages: a comparative study. Doctoral dissertation, Lunds University.
- Diesing, Molly. 2005. The upper functional domain in Yiddish. In *Focus on Germanic typology*, ed. by Werner Abraham, 195–209. Berlin: Akademierverlag.
- Den Dikken, Marcel. 2007. Phase extension: contours of a theory of the role of head movement in phrasal extraction. *Theoretical Linguistics* 33: 1–41.
- Douglas, Jamie. 2017. Unifying the that-trace and anti-that-trace effects. *Glossa* 2: 60.
- Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39: 1–53.
- Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32: 555–595.
- Emonds, Joseph. 1978. The verbal complex V'-V in French. *Linguistic Inquiry* 9: 151–175.
- Emonds, Joseph. 1970. Root and structure-preserving transformations. Doctoral dissertation, Massachusetts Institute of Technology.
- Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language and Linguistic Theory* 34: 429–479.
- Erlewine, Michael Yoshitaka. 2018. Extraction and licensing in Toba Batak. *Language* 94: 662–697.
- Fanselow, Gisbert. 2003. Münchhausen-style head movement and the analysis of verb second. *UCLA Working Papers in Linguistics*: 40–76.
- Fanselow, Gisbert. 2009. Bootstrapping verb movement and the clausal architecture of German (and other languages). In *Advances in comparative Germanic syntax*, ed. by Artemis Alexiadou, Jorge Hankamer, Thomas McFadden, Justin Nüger, and Florian Schäfer, 85–118. Amsterdam/Philadelphia: John Benjamins.
- Fanselow, Gisbert, and Denisa Lenertová. 2010. Left peripheral focus: mismatches between syntax and information structure. *Natural Language and Linguistic Theory* 29: 169–209.
- Gallego, Ángel J. 2006. *Phase sliding*. Ms, Universitat Autònoma de Barcelona and University of Maryland.
- Gallego, Ángel J. 2010. *Phase theory*. Amsterdam/Philadelphia: John Benjamins.
- Georgi, Doreen, and Gereon Müller. 2010. Noun phrase structure by reprojection. *Syntax* 13: 1–36.
- Georgi, Alessandra, and Fabio Pianesi. 1997. *Tense and aspect: from semantics to morphosyntax*. New York: Oxford University Press.
- Georgi, Alessandra, and Fabio Pianesi. 2004. Complementizer deletion in Italian. In *The structure of CP and IP: The cartography of syntactic structures, Volume 2*, ed. by Luigi Rizzi, 190–210. Oxford: Oxford University Press.
- Giusti, Giuliana. 2002. The functional structure of noun phrases: A bare phrase structure approach. In *Functional structure in DP and IP: The cartography of syntactic structures, Volume 1*, ed. by Guglielmo Cinque, 54–90. Oxford: Oxford University Press.
- Goldsmith, John. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell.

- Good, Jeffrey C. 2003. Strong linearity: Three cases studies towards a theory of morphosyntactic templatic constructions. University of California, Berkeley.
- Grimshaw, Jane. 1991. Extended projections. Ms., Brandeis University.
- Grimshaw, Jane. 2000. Extended projection and locality. In *Lexical Specification and Insertion*, ed. by Peter Coopmans, Martin Everaert, and Jane Grimshaw, 115–133. Amsterdam: John Benjamins.
- Grohman, Kleanthes K. 2001. Grohman, Kleanthes K. 2001. On predication, derivation and anti-locality. *ZAS Papers in Linguistics* 26. 87–112. *ZAS papers in linguistics* 26: 87–112.
- Grohmann, Kleanthes K. 2002. Anti-locality and clause types. *Theoretical Linguistics* 28: 43–72.
- Grohmann, Kleanthes K. 2003. *Prolific domains: On the anti-locality of movement dependencies*. Amsterdam: John Benjamins.
- Haider, Hubert. 1988. Matching projections. In *Constituent structure: papers from the 1987 GLOW Conference*, ed. by Anna Cardinaletti, Guglielmo Cinque, and Giuliana Giusti, 101–121. Dordrecht: Foris.
- Harizanov, Boris. 2014. Clitic doubling at the syntax-morphophonology interface. *Natural Language and Linguistic Theory* 32: 1033–1088.
- Harizanov, Boris, and Vera Gribanova. 2019. Whither head movement ? *Natural Language and Linguistic Theory* 37: 461–522.
- Harley, Heidi. 2013. Getting morphemes in order: Merger, Affixation and Head Movement. In *Diagnosing syntax*, ed. by Lisa L. Cheng and Norbert Corver, 44–74. Oxford: Oxford University Press.
- Hartman, Jeremy. 2011. The semantic uniformity of traces: evidence from ellipsis parallelism. *Linguistic Inquiry* 42: 367–388.
- Höhn, Georg F.K. 2016. Unagreement is an illusion: Apparent person mismatches and nominal structure. *Natural Language and Linguistic Theory* 34: 543–592.
- Hsu, Brian. 2017. Verb second and its deviations: An argument for Feature Scattering in the left periphery. *Glossa* 2: 35.
- Hsu, Brian, and Saurov Syed. 2019. Variation in the co-occurrence of indexical elements : evidence for split indexical projections in DPs. In *Proceedings of the 36th West Coast Conference on Formal Linguistics*, ed. by Richard Stockwell, Maura O’Leary, Zhongshi Xu, and Z.L. Zhou, 188–197.
- Hyman, Larry M. 2003. Suffix ordering in Bantu: a morphocentric approach. In *Yearbook of Morphology 2002*, ed. by Geert Booij and Jaap van Marle, 245–281. Dordrecht: Springer.
- Iatridou, Sabine. 1990. About Agr(P). *Linguistic Inquiry* 21: 551–577.
- Itô, Junko. 1988. *Syllable theory in Prosodic Phonology*. New York: Garland Publishing.
- Julien, Marit. 2002. *Syntactic heads and word formation: a study of verbal inflection*. New York: Oxford University Press, USA.
- Julien, Marit. 2005. *Nominal phrases from a Scandinavian perspective*. Amsterdam: John Benjamins.
- Kandybowicz, Jason. 2009. Embracing edges: syntactic and phono-syntactic edge sensitivity in Nupe. *Natural Language and Linguistic Theory* 27: 305–344.

- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kayne, Richard S. 2005a. Some notes on comparative syntax, with special reference to English and French. In *The Oxford handbook of comparative syntax*, ed. by Guglielmo Cinque and Richard S. Kayne, 3–69. New York: Oxford University Press.
- Kayne, Richard S. 2005b. *Movement and silence*. Oxford: Oxford University Press.
- Keine, Stefan. 2019. Selective opacity. *Linguistic Inquiry* 50: 13–62.
- Koopman, Hilda. 1996. The Spec Head configuration. In *Syntax at Sunset, UCLA Working papers in syntax and semantics Volume 1*, ed. by Edward Garrett and Felicia Lee, 37–64. Los Angeles: University of California, Los Angeles.
- Koopman, Hilda. 2017. A note on Huave morpheme ordering : Local Dislocation or generalized U20? In *Perspectives on the architecture and acquisition of syntax: essays in honour of R. Amritavalli*, ed. by Gautam Sengupta, Shruti Sircar, Gayatri Raman, and Rahul Balusu, 23–47. Dordrecht: Springer.
- Lechner, Winfried. 2006. An interpretive effect of head movement. In *Phases of Interpretation*, ed. by Mara Frascarelli, 45–71. Berlin: Mouton de Gruyter.
- Ledgeway, Adam. 2010. Subject licensing in CP. In *Mapping the Left Periphery*, ed. by Paola Benincà and Nicola Munaro, 257–296. Oxford: Oxford University Press.
- Leu, Thomas. 2015. Generalized x-to-C in Germanic. *Studia Linguistica* 69: 272–303.
- Li, Yafei. 1990. X⁰-binding and verb incorporation. *Linguistic Inquiry* 21: 399–426.
- Lundquist, Björn. 2008. Nominalizations and participles in Swedish. Doctoral dissertation, University of Tromsø.
- Mahajan, Anoop Kumar. 2001. *Word order and remnant VP movement*. In *Word order and scrambling (Explaining Linguistics 4)*, ed. by Simin Karimi, 217–237. Malden, MA: Blackwell.
- Manetta, Emily. 2011. *Peripheries in Kashmiri and Hindi-Urdu: The syntax of discourse-driven movement*. Amsterdam/Philadelphia: John Benjamins.
- Manova, Stela, and Mark Aronoff. 2010. Modeling affix order. *Morphology* 20: 109–131.
- Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.
- Marantz, Alec. 2007. Phases and words. In *Phases in the theory of grammar*, ed. by S. H. Choe, 191–222. Seoul: Dong-In Publishing Co.
- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37: 69–109.
- Matyiku, Sabina Maria. 2017. Semantic effects of head movement : Evidence from negative auxiliary inversion. Doctoral dissertation, Yale University.
- McCarthy, John J, and Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology*: 79–153.
- Mohr, Sabine. 2009. V2 as a single-edge phenomenon. In *Selected Papers from the 2006 Cyprus syntaxfest*, ed. by Kleanthes K. Grohmann and Phoevos Panagiotidis, 141–159. Cambridge: Cambridge Scholars Publishing.
- Müller, Gereon. 2004. Verb-second as vP-first. *Journal of Comparative Germanic Linguistics* 7:179–234.
- Munshi, Sadaf, and Rajesh Bhatt. 2009. Two locations for negation: Evidence from

- Kashmiri. *Linguistic Variation Yearbook* 9: 205–205–239.
- Newell, Heather, and Maire Noonan. 2018. A re-portage on spanning; feature portaging and non-terminal spell-out. In *Heading in the right direction: Linguistic treats for Lisa Travis*, ed. by Ileana Paul, Laura Kalin, and Jozina van der Klok, 289–303. Montréal: McGill Working Papers in Linguistics.
- Oltra-Massuet, Isabel. 2013. Variability and allomorphy in the morphosyntax of Catalan past perfective. In *Distributed Morphology Today: Morphemes for Morris Halle*, ed. by Ora Matushansky and Alec Marantz, 1–20. Cambridge, MA: MIT Press.
- Ouhalla, Jamal. 1991. *Functional categories and parametric variation*. London: Routledge.
- Panagiotidis, Phoevos. 2014. *Categorial features: a generative theory of word class categories*. Cambridge: Cambridge University Press.
- Paster, Mary. 2005. Pulaar verbal extensions and phonologically driven affix ordering. *Yearbook of Morphology* 2005: 155–199.
- Pesetsky, David. 2013. *Russian case morphology and the syntactic categories*. Cambridge, MA: MIT Press.
- Platzack, Christer. 2013. Head movement as a phonological operation. In *Diagnosing syntax*, ed. by Lisa Lai-Shen Cheng and Norbert Corver, 21–43. Oxford: Oxford University Press.
- Poletto, Cecilia. 2000. *The higher functional field*. New York: Oxford University Press, USA.
- Poletto, Cecilia. 2002. The left-periphery of V2-Rhaetoromance dialects: a new view on V2 and V3. In *Syntactic microvariation*, ed. by Sjef Barbiers, Leonie Cornips, and Susanne van der Kleij, 214–242. Amsterdam: Meertens.
- Poletto, Cecilia. 2014. *Word order in Old Italian*. Oxford: Oxford University Press.
- Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20: 365–424.
- Pylkkänen, Liina. 2002. Introducing arguments. Doctoral dissertation, Massachusetts Institute of Technology.
- Rice, Keren. 2011. Principles of affix ordering: An overview. *Word Structure* 4: 169–200.
- Richards, Norvin. 2001. *Movement in language*. Oxford: Oxford University Press.
- van Riemsdijk, Henk. 1988. The representation of syntactic categories. *Proceedings of the Conference on the Basque Language, Basque World Congress* 1: 104–116.
- van Riemsdijk, Henk. 1998. Categorial feature magnetism: The endocentricity and distribution of projections. *Journal of Comparative Germanic Linguistics* 2: 1–48.
- Rivero, Maria-Luisa. 1991. Long head movement and negation: Serbo-Croatian vs. Slovak and Czech. *The Linguistic Review* 8: 319–351.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar*, ed. by Liliane Haegeman, 281–337. Dordrecht: Kluwer.
- Roberts, Ian G. 2005. *Principles and parameters in a VSO language: a case study in Welsh*. Oxford: Oxford University Press.
- Roberts, Ian G. 2010. *Agreement and head movement: clitics, incorporation, and defective goals*. Cambridge, MA: MIT Press.

- Di Sciullo, Anna Maria, and Daniela Isac. 2008. The asymmetry of Merge. *Biolinguistics* 2: 260–290.
- Speas, Margaret. 1991. *Phrase structure in natural language*. Dordrecht: Kluwer.
- Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19: 425–449.
- Sportiche, Dominique. 2005. Division of labor between Merge and Move : strict locality of selection and apparent reconstruction paradoxes. In *Proceedings of the workshop “Division of Linguistic Labor” : The la Bretesche workshop*, 159–262.
- Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36: 1–6.
- Steriade, Donca. 1995. Underspecification and markedness. In *The Handbook of Phonological Theory*, ed. by John Goldsmith, 114–174. Oxford: Blackwell.
- Surányi, Balázs. 2005. Head movement and reprojection. In *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae. Sectio Linguistica. Tomus XXVI*, 313–342. Budapest: ELTE.
- Svenonius, Peter. 1994. C-selection as feature checking. *Studia Linguistica* 48: 133– 155.
- Svenonius, Peter. 2012. *Spanning*. Ms., University of Tromsø.
- Svenonius, Peter. 2016. Spans and words. In *Morphological metatheory*, ed. by Heidi Harley and Daniel Siddiqi, 199–220. Amsterdam: John Benjamins.
- Thráinsson, Höskuldur. 1996. On the (non)-universality of functional projections. In *Minimal ideas*, ed. by Werner Abraham, Samuel David Epstein, Hoskuldur Thráinsson, and C. Jan-Wouter Zwart, 253–81. Amsterdam: John Benjamins.
- Travis, Lisa. 1984. Parameters and effects of word order variation. Doctoral dissertation, Massachusetts Institute of Technology.
- Ussishkin, Adam. 2007. Morpheme position. In *The Cambridge handbook of phonology*, ed. by Paul de Lacy, 457–472. Cambridge: Cambridge University Press.
- Vangsnes, Øystein Alexander. 1999. The identification of functional architecture. Doctoral dissertation, University of Bergen.
- Walkden, George. 2017. Language contact and V3 in Germanic varieties new and old. *Journal of Comparative Germanic Linguistics* 20: 49–81.
- Walker, Rachel. 2011. *Vowel patterns in language*. New York: Cambridge University Press.
- Wilder, Chris, and Damir Ćavar. 1994. Word order variation, verb movement, and economy principles. *Studia Linguistica* 48: 46–86.
- Williams, Edwin. 1981. On the notions “lexically related” and “head of a word.” *Linguistic Inquiry* 12: 245–274.
- Wolfe, Sam. 2019. *Verb second in Medieval Romance*. Oxford: Oxford University Press.
- Zwart, C. Jan-Wouter. 1997. *Morphosyntax of verb movement: A minimalist approach to the syntax of Dutch*. Dordrecht: Kluwer.
- Zwicky, Arnold M. 1970. Auxiliary reduction in English. *Linguistic Inquiry* 1: 323–336.
- Zwicky, Arnold M., and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English N’T. *Language* 59: 502–513.

Department of Linguistics
University of North Carolina at Chapel Hill
104A Smith Building, CB# 3155
Chapel Hill, NC 27599-3155
hsub@email.unc.edu

This work is greatly indebted to helpful discussion at various stages with Nico Baier, Theresa Biberauer, Michael Yoshitaka Erlewine, Randall Hendrick, Nicholas LaCara, Elliott Moreton, Katya Pertsova, Jennifer Smith, Mike Terry, and two anonymous *LI* reviewers. I would like to thank in particular Andrew Simpson and Roumyana Pancheva for inspiring the early stages of this project. I am also grateful to audiences at NELS 46 at Concordia University, the Rethinking Verb Second workshop at Cambridge University, the Parameters Workshop in Honour of Lisa Travis at McGill University, and a UNC Linguistics Friday Colloquium. This research has been supported by the Carolina Postdoctoral Program for Faculty Diversity at UNC-CH. All errors are my own.

ⁱ This asymmetry is not found in the extraction of embedded arguments, which appears insensitive to the presence of a complementizer. Giorgi and Piansesi (1997: 247-253) posit that this is because arguments do not need to move successive cyclically through the embedded clause edge (cf. Cinque 1990).

ⁱⁱ Giorgi and Piansesi leave aside the issue of why the ability of the subject probe on [Agr] to attract a specifier depends on whether it is realized on a standalone head. Section 4.1 discusses a similar pattern in Kashmiri in which bundling “suppresses” a probe.

ⁱⁱⁱ In addition to pre-syntactic bundling, “Scattering B,” Giorgi and Piansesi hypothesize the existence of “Scattering A,” a syntactic operation that takes a head with a feature bundle and splits one of its features onto a separate head. I will restrict my attention to Scattering B, the more commonly assumed approach to feature bundling.

^{iv} Giorgi and Piansesi (1997: 15) make the stronger claim that a bundle of features can project if and only if they require a filled specifier: “A bundle of features [...] can be projected by means of more than one head (i.e., scattered) only if extra *Spec* positions are required to locate other bundles of features contained in the initial array. This implies that if there are no Specifiers to be projected from the initial array, no extra head

can be created by scattering.” As detailed in Sections 3 and 4, my analysis departs from this: I claim that while only dominant heads can license specifiers, not all dominant heads have this ability.

^v While I focus on bundling patterns where probes compete to trigger phrasal movement, bundled heads in some cases appear to carry a *composite probe* that seeks a goal that checks multiple targeted features (Coon & Bale 2014; Erlewine 2018). I must leave unexplored the question of when composite probes can arise from bundling.

^{vi} The latter issue does not arise within Kayne’s (1994) revised definition of c-command that distinguishes between segments and categories.

^{vii} In Den Dikken’s proposal, the probe cannot initially agree with the object because *vP* is a phase. Head movement of *v+V* to *F* extends the phase upwards so that object becomes an accessible goal. Kandybowicz (2009) posits that some movement-triggering features are ‘dormant,’ and require activation via agreement with and attraction of a lower head. My proposal provides an alternative explanation: agreement between *F* and the object is always possible, but *F* must inherit its ability to have a specifier from head-moved *v+V*.

^{viii} This structural characterization resembles the one posited in reprojection theories of head movement (Fanselow 2003; Surányi 2005; Georgi and Müller 2010), which also aim to make head movement compliant with the Uniformity Condition. In this view, moved heads project their labels to the root node, thus making both positions of the head minimal projections. As discussed in Section 3.2, reprojection is obviated if Category Percolation (Keine 2019) is adopted.

^{ix} This resembles previous claims that functional projections are only licensed if they have a filled specifier or an overt head (Koopman 1996; Vangsnes 1999; Giusti 2002; Julien 2005). However, in Feature Scattering these criteria do not affect whether functional category features are present in extended projections (assumed to always have the same features), but in how they are mapped to phrase structure.

^x This structural definition, which precludes Coalescence from applying to two heads with an intervening specifier, often resembles linear adjacency. The two definitions would differ in a context where a head-initial projection dominates a head-final one with no specifier, [_{zP} Z [_{YP} [_{XP} ...] Y], satisfying structural but not linear adjacency. While I know of no cases of bundling in this configuration, this may be explained if true head-final structures are excluded from syntactic representations (Kayne 1994).

^{xi} The output of Coalescence is identical to that of Matching Projections in Haider (1988). However, the contexts in which it applies are not the same. Haider proposes that functional projections whose heads have no phonetic realization are superimposed onto another projection within the same extended projection. In the present proposal, there is no requirement for all projections to have a pronounced head.

^{xii} I thank an anonymous reviewer for pointing out the relevance of such examples.

^{xiii} I cannot yet provide a deeper explanation for why bundling requires a dominant head to c-command a recessive head, but not vice versa. Potentially, it may follow from restrictions on the elements that can trigger syntactic operations, or on their search space. However, a fuller explanation must await future work.

^{xiv} In Chomsky's (2008) definition, Merge of syntactic objects X and Y leaves both unchanged. Although Coalescence is not a variety of Merge, one might consider this as a broader restriction on syntactic operations.

^{xv} The Uniformity Condition and Extension Condition could each turn out to be incorrect. For example, certain patterns have been taken to require violations of the Extension Condition as commonly defined (Richards 2001; Sportiche 2005; Pesetsky 2013). A detailed evaluation of these conditions and where they (don't) apply must await another occasion.

^{xvi} Tentatively, we may explore the idea that the movement of dominant heads is *always* anti-local, crossing at least two projections before Coalescence. Given the empirical arguments made in the cartographic program for a highly articulated hierarchy of features, it could be that the apparently local movement patterns used to justify the Head Movement Constraint in fact require movement across intervening heads. However, this discussion remains speculative, given the range of proposals on how anti-locality restrictions are best defined, and the absence of strong, widely accepted diagnostics to precisely determine how many category features are contained within a given projection.

^{xvii} The approach resembles the system of Chomsky (2001) in the sense that [EPP] is distinct from [μ F] probes on a lexical item, and Agree only triggers movement if the probing item contains [EPP]. As to be discussed, my proposal differs in that [EPP] is not 'satisfied' or checked once a specifier is Merged.

^{xviii} This requires a slight modification in a theory where roots obtain categorial properties from a categorizing functional head, e.g. ‘nominalizer’ *n* and ‘verbalizer’ *v* (Marantz 1997; Embick and Marantz 2008). Under this view, [EPP] is a unique property of categorizing head *v*.

^{xix} This has some similarities to the analysis of Harizanov and Gribanova (2019), who argue that only long head movement involves syntactic displacement of heads, while “short” head movement is postsyntactic morphological amalgamation. Although this also accounts for the differences in locality and bundling between long and “short” head movement, it does not explain delayed gratification patterns in which apparently local head movement affects the licensing of phrasal movement to the target projection.

^{xx} For arguments that a low C-domain feature like Finiteness is associated with subject probes, see Poletto (2000) Aboh (2006), Ledgeway (2010).

^{xxi} In a similar vein, Leu (2015) proposes that the overt complementizer vs. verb-second alternation reflects two strategies used to activate the highest clausal projection in German. In brief, the projection can be activated either by movement of a remnant *v*P (Müller 2004) or by movement of a lower head “*d-*”, in which case the target head is realized as *-ass*.

^{xxii} For Matushansky, sentential negation is Merged as specifier of AuxP that is both maximal and minimal. Neg and Aux can undergo M-Merger because both are minimal projections; Because TP and AuxP are adjacent, Aux-to-T movement satisfies the locality requirement on head movement that she assumes. However, this approach leaves unresolved the status of negation in clauses without auxiliaries, and why only the highest auxiliary projection in a series can host negation in its specifier.

^{xxiii} Some modifications are needed in light of a standard analysis of English clause structure in which auxiliary heads are first Merged below NegP, and the highest auxiliary moves to T: [TP T° [NegP Neg° [AuxP Aux° ...]]]. In this structure, Aux-to-T movement followed by Coalescence is generated if both T°_R and Neg°_R are recessive, and dominant Aux°_D moves to TP. However, in derivations that contain a dominant Neg°_D (as argued for *not*), we incorrectly predict that Neg°_D will move to TP, leaving the auxiliary below negation. I propose a tentative solution appealing to Category Percolation. Suppose that the dominant and recessive Neg° heads differ in their featural content, such that Neg°_R contains a subset of T°’s subfeatures, but Neg°_D does not. In other words, Neg°_D is not part of the verbal extended projection, and

thus cannot undergo Coalescence with T°_R , as it would violate the No Tampering Condition. The only option is to move Aux°_D .

^{xxiv} I do not exclude the possibility that V° moves first through AspP before raising to TP. However, this is not critical, since the two possible derivations result in the same bundled head structure.

^{xxv} A reviewer suggests an alternative in which all clauses contain Aux, but verbs select either Aux_D *anar* or an Aux_R null auxiliary. This permits a uniform set of functional heads in all clauses, and a more restrictive way of determining the content of the numeration. It also comes with potentially significant implications. If auxiliary selection arises from Agree between Aux and V, it suggests that dominance vs. recessiveness is not always an inherent property of heads when they are first Merged, but can be contextually determined by agreement. I must leave detailed consideration of this hypothesis to future work.

^{xxvi} It remains an important question how the phasal properties of C° should be defined, if the left periphery contains an articulated series of functional projections. See Douglas (2017) for a recent approach.

^{xxvii} The Mirror Principle is not exceptionless; languages may show fixed affix orders that contradict scopal or derivational relations (Hyman 2003; Good 2003). The Mirror Principle can also be violated in favor of phonological restrictions like relative sonority (Arnott 1970; Paster 2005) or prosodic properties of stems (McCarthy and Prince 1993; Ussishkin 2007). Nonetheless, head-internal branching in syntactic representations has explanatory power to the extent that it predicts one of several fundamental preferences on affix ordering (Manova and Aronoff 2010, Rice 2011).

^{xxviii} It could be the case that lexical category features are in fact organized in subset relations (Lundquist 2008), akin to my implementation of Category Percolation. For example, if adjectival heads *a* contain a subset of features of noun heads *n*, *gloriousness* poses no issues for Coalescence. However, it is currently unclear how this can account for category changes in the opposite direction (e.g. $N > A$, $A > N$). A second possibility possibility is that the c-selection of *n* by another head renders all other categorial heads “inert,” such that $[n [a [\sqrt{GLORY}]]]$ patterns like a minimal projection of *n* in the remainder of the derivation. The nature of such a restriction, however, remains to be explained.

^{xxix} Spanning bears key resemblances to the approach to lexical insertion in Nanosyntax (Starke 2009).

While a detailed comparison among these theories and my proposal is outside the scope of the paper, given

major differences in the assumed architecture of grammar, I refer the reader to recent discussion in Baunaz et al. (2018) and Newell and Noonan (2018).