Coalescence: A unification of bundling operations in syntax

Brian Hsu (hsub@email.unc.edu)
University of North Carolina at Chapel Hill

Abstract

This article revisits the status of two proposed bundling operations that affect heads: Feature Scattering (Georgi & 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 called 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 a single head compete to trigger phrasal movement (Fanselow & Lenertová 2010).

This paper first discusses two types of bundling patterns in syntax: the creation of featurally complex heads as the result of head movement, with a focus on the M-Merger bundling operation of Matushansky (2006), and variation in the number of heads on which a set functional features is distributed even in the absence of movement, in particular Giorgi & Pianesi's (1997) Feature Scattering Hypothesis. 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" features 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 patterns involving heads 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 (Travis 1984) with anti-locality restrictions on movement (Abels 2003; Erlewine 2016). 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 adjunction, 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 phasal spell-out, and implications and compatibilites of the approach with theories of affix order. 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, the two features [X] and [Y] can either occur on separate heads X˚, Y˚ (1), or bundled on a single head X/Y˚ (2).

(1)　　　　XP　　　　　　　　　　　(2)　　　　X/YP
　　　　　／＼　　　　　　　　　　　　　　　／＼
　　　　X˚　　YP　　　　　　　　　　　X/Y˚　　...
　　　　　　　／＼
　　　　　　Y˚　　...
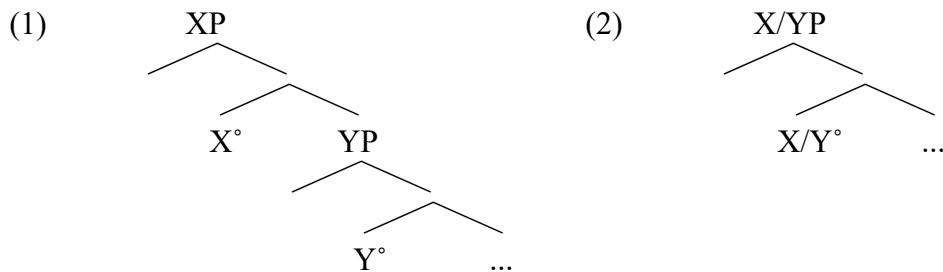
This contrasts with standard "cartographic" analyses in which each feature occurs on a single head in a universal hierarchy of projections (Cinque 1999; R. S. Kayne 2005a; Cinque and Rizzi 2009). On this view, (1) is the only possible configuration of the features [X], [Y]. While both approaches are 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 & Pianesi's 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 latter group, the only grammatical subject position in these clauses is a post-verbal position (3b).

(3) a. Credo   [Maria   sia   partita]
       I.believe   Maria   is.SBJV   left
       'I believe that Maria left'

    b. Credeva   [fosse   arrivato   Gianni]
       I.believe   was.SBJV   arrived   Gianni
       'I believed that Gianni had arrived'

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

(4) Credo   [che   Maria sia   partita]
    I.believe   that   Maria is.SBJV   left
    'I believe that Maria left'

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

5

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
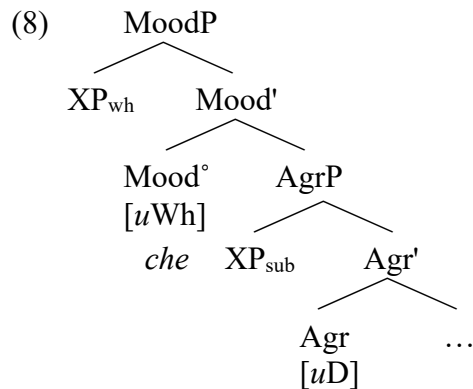   'Why did you believe that Gianni had left?'

Giorgi & Pianesi (1997) argue that both asymmetries are accounted for with the following claims. (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 occupy the specifier of the projection containing [Agr]. (iv) Adjunct *wh*-words are first Merged as specifiers of the projection whose head contains [Mood] (cf. Cinque 1990).

The bundled Mood/AgrP (7) licenses a single specifier position 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) is accounted for as variation in whether a bundled Mood/Agr° head is able to trigger movement of the subject from a lower postverbal position. The ungrammaticality of adjunct *wh*-word

extraction from a clause with a preverbal subject arises because Spec, Mood/AgrP cannot be simultaneously filled by an adjunct *wh*-word and a subject; extraction is only possible if the subject remains in a postverbal position below Mood/AgrP.

(7) Mood/AgrP

    XP$_{sub/wh}$ Mood/Agr'

           Mood/Agr°   ...
             [uWh]
             [uD]
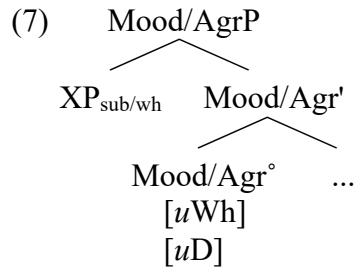
In contrast, embedded clauses with a complementizer contain two projections AgrP and MoodP, where the complementizer is the realization of Mood° (8). Preverbal subject orders are grammatical for all speakers if they uniformly allow subject movement to Spec, AgrP.[ii] Adjunct *wh*-words can be extracted from clauses with a preverbal subject because they are generated in a separate position Spec, MoodP.

(8) MoodP

    XP$_{wh}$ Mood'

        Mood°   AgrP
        [uWh]
         che XP$_{sub}$ Agr'

                      Agr   …
                      [uD]

Crucially, these differences between clauses with and without a complementizer are not expected under a strictly Cartographic 'one feature - one head' alternative (Cinque & Rizzi 2008) in which MoodP and AgrP are always separately projected, and the only difference between clauses with and without a complementizer is in whether the Mood°

head is pronounced. In essence, there is no principled explanation in this view for why Spec, AgrP cannot be filled when Mood° is not pronounced, as would be necessary 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 the realization of Tense, Aspect, and Mood projections (Giorgi & Pianesi 1997), 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 *to appear*), and positions in the extended complementizer domain (Douglas 2017; Hsu 2017).
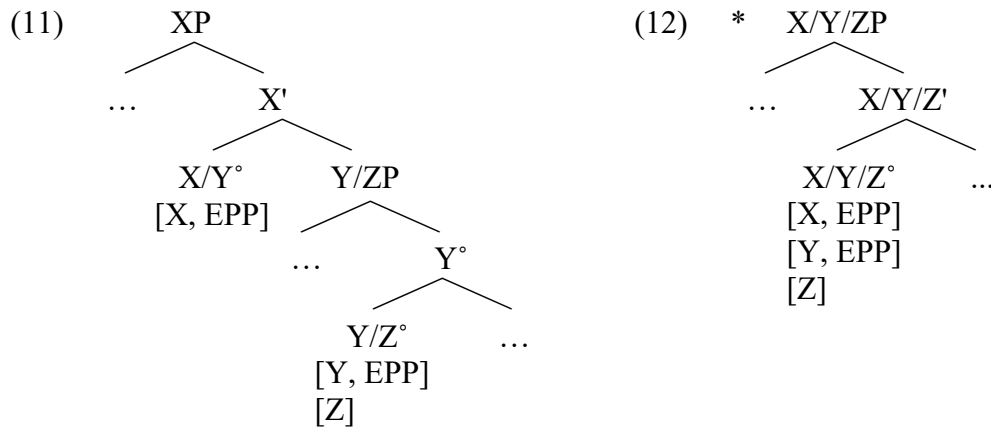
In the proposal of Giorgi & Pianesi (1997), languages share a universal inventory of features, but differ in how individual features [F] are packaged onto the lexical items that enter syntactic derivations (a similar proposal is articulated in Cowper 2005). In other words, the relevant locus of variation is within the *pre-syntactic lexicon*, rather than in syntactic derivations.[iii] To account for the observation that languages maintain a universal hierarchy of projections, regardless of the number of realized projections in a given structure, Giorgi and Pianesi (1997) propose that languages share a hierarchically ordered set of features, and that the order in which they are checked (in this case, by external Merge) must reflect this hierarchical order.

(9)   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 & Pianesi 1997)

This constraint restricts the possible feature bundles that can be found within and across languages. In effect, it enforces 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 instantiated in the structural configurations in (10). The Universal Ordering Constraint crucially rules out the appearance of [Z] and [X] within a projection that excludes [Y] (10f); in other words, bundling cannot 'skip' intervening features.

(10)   a.   [XP   [YP   [ZP   …
        b.   [X/YP   [ZP   …
        c.   [XP   [Y/ZP   …
        d.   [X/Y/ZP   …
        f.   *[X/ZP   [YP   …

In addition, Giorgi & Pianesi (1997: 15) propose an intimate 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 (12). In contrast, there is no prohibition against the bundling of [Z], which lacks this requirement, with either of the other features.

(11)
```
              XP
           ╱      ╲
        …          X'
                ╱      ╲
          X/Y°          Y/ZP
          [X, EPP]    ╱      ╲
                    …          Y°
                            ╱      ╲
                       Y/Z°          …
                       [Y, EPP]
                       [Z]
```

(12)   *
```
              X/Y/ZP
           ╱        ╲
        …            X/Y/Z'
                  ╱        ╲
            X/Y/Z°            …
            [X, EPP]
            [Y, EPP]
            [Z]
```

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). In this approach, each head contains an interpretable category feature, and optionally an uninterpretable category feature. Merge is triggered in order to check uninterpretable category features, which can be checked under sisterhood (13). Uninterpretable category features enable heads to select the categories of their complements. Consequently, the universal order of feature checking is encoded in the atoms of the syntactic derivation and

their selectional properties. In this view, *category features* can be defined as features of syntactic objects that trigger external Merge (Di Sciullo and Isac 2008).

(13)
$$
\begin{array}{ccc}
& & \text{XP} \\
& & [X] \\
\diagup \quad \diagdown & \longrightarrow & \diagup \quad \diagdown \\
\text{X} \qquad \text{Y} & & \text{X} \qquad \text{Y} \\
[\text{X}, u\text{Y}] \quad [\text{Y}] & & [\text{X}, \text{\textit{u}Y}] \quad [\text{\textbf{Y}}]
\end{array}
$$
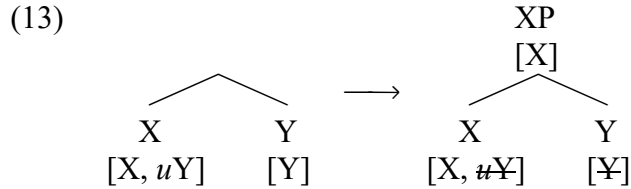
On the other hand, if some category features are in the lexicon, prior to numeration of the syntactic derivation, the restriction on the order of feature checking has to exist independently of lexical entries, for instance as a stipulated metacondition on bundling in the lexicon like the Universal Ordering Constraint. Restrictions on the order of feature checking are thus duplicated both in the syntax and in the lexicon, a redundant specification in separate modules of the grammar. In this work, I argue that this redundancy is dispensed with if the 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 both language-internal and cross-linguistic word order variation – briefly illustrated here with two well-known examples. Within English, subject-auxiliary inversion observed in a range of clause types is commonly accounted for in terms of 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* (14).

(14)   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.
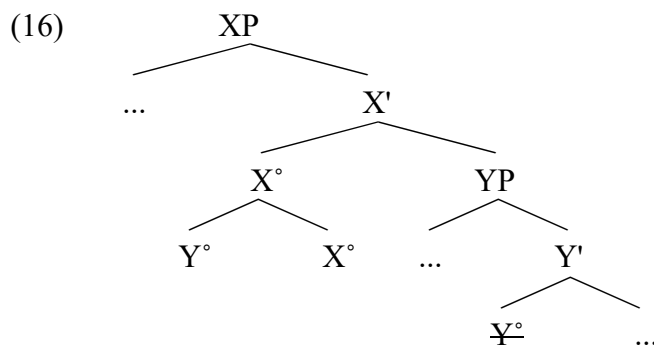
Head movement is similarly instrumental in accounting for cross-linguistic variation. As a famous example, English and French differ in the placement of tense-inflected verbs relative to adverbs like 'often' (Emonds 1970; 1978; Pollock 1989). Assuming that such adverbs occupy the same position between TP and VP in both languages, the difference can be accounted for in terms of head movement. Tensed lexical verbs remain in V (or *v*) in English, but move to T in French.

(15) a. Lucille $_T$ often   drinks$_V$ martinis   *English*

b. Lucille   boit$_{V+T}$ souvent   ~~boit~~   des martinis   *French*

Head movement has received numerous approaches in generative syntax (see also Dékány 2018 for a recent summary). Broadly speaking, head movement patterns have been analyzed in terms of various syntactic operations (Travis 1984; M. Baker 1988; Pollock 1989; 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 or phonological operation (Chomsky 2000; Boeckx and Stjepanovic 2001; Platzack 2013). Alternatively, it has been argued that some types of head displacement take place in syntax, while others are generated postsyntactically (Harizanov and Gribanova 2018). In this subsection, I outline key justifications for a syntactic analysis of head movement, and remaining issues that the proposal will aim 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 (16). Additional branching structure within complex heads can be created by successive steps of head movement (Travis 1984; Baker 1988). Head adjunction structures are used to account for the observation that head movement tends to create morphologically complex forms in which features of the movement target position appear as affixes on the exponent of the moved head (M. Baker 1988; Julien 2002).

(16)

```
                    XP
              ┌──────┴──────┐
             ...            X'
                      ┌──────┴──────┐
                     X°            YP
                  ┌───┴───┐    ┌────┴────┐
                 Y°       X°  ...        Y'
                                    ┌─────┴─────┐
                                   Y̶°          ...
```

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.[v]

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˚ (Travis's 1984 Head Movement Constraint), whereas phrasal movement does not require this kind of locality. Moreover, it has been argued in recent works that phrasal movement is *anti-local*; phrasal movement from XP to YP must cross an intervening phrase ZP (Grohman 2001, 2002, 2003; Abels 2003; Erlewine 2016). These opposing locality restrictions are shown schematically below. While this distinction does not necessarily cast doubt on head movement as a syntactic operation, at minimum it suggests that there is fundamental difference in the types of features that trigger head movement versus phrasal movement.

(17) a. [$_{XP}$ X+Y [$_{YP}$ Y̶ …
      b. *[$_{XP}$ X+Y [$_{ZP}$ Z   [$_{YP}$ Y̶ …

(18) a. *[$_{XP}$ WP X   [$_{YP}$ W̶P̶   Y …
      b. [$_{XP}$ WP X   [$_{ZP}$ Z   [$_{YP}$ W̶P̶   Y …

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 (head vs. phrasal) is determined uniquely by its position within the syntactic 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 (19).

(19)

X^{max}
X^{min} Y^{max}
Y^{min} $Z^{max/min}$
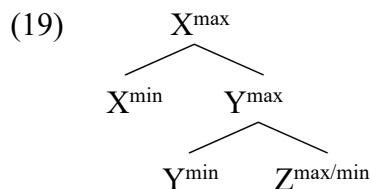
Projections that are neither minimal nor maximal, corresponding to X' levels in X-bar Theory, are proposed to be inaccessible to syntactic operations and thus unable to undergo movement. To rule out movements deemed to be impossible, such as head-to-specifier movement or XP adjunction to heads, Chomsky (1995: p. 253) posits a uniformity condition on movement chains:

(20) SMALL CAPS UNIFORMITY CONDITION
      A chain is uniform with respect to phrase structure status

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

(21)

$$
\begin{array}{l}
\text{I}^{max} \\
\quad \diagup \diagdown \\
\ldots \qquad \text{I} \\
\qquad \diagup \diagdown \\
\quad \text{I} \qquad \text{V}^{max} \\
\diagup \diagdown \qquad \diagup \diagdown \\
\text{V}^{min/max}_k \quad \text{I}^{min} \quad \ldots \quad \text{V} \\
\qquad\qquad\qquad \diagup \diagdown \\
\qquad\qquad \text{V}^{min}_k \quad \ldots
\end{array}
$$
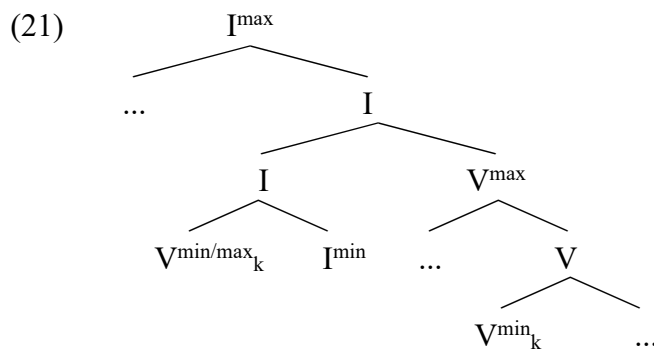
The two copies of V differ in their phrase structure status, in violation of the Uniformity Condition. Because the higher copy of V does not project, it is both minimal and maximal, while the lower copy is minimal because it projects. As observed by Harley (2013), successive-cyclic head movement proves even more problematic. For instance, because the complex V-Infl head created in (21) is neither minimal or maximal, it is predicted to be inaccessible to later syntactic operations (such as Infl-to-C movement).

    Although the aforementioned properties of head movement 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 been argued to have clear 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 necessary 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 languages, objects can move above sentential negation only if verb movement also takes place.

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

(23) 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° first head-moves to F°. Abstracting away from the exact label and probe of this head, the derivation is shown in (24).[vi] Verb movement to a higher projection (with movement of the subject) in a later step produces the word order in (22).

(24) a. [FP   F   [vP   *subject*   [$_{v'}$ v [VP V *object*]]]]
                 [PROBE]

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

16

c. [$_{FP}$ *object* F+*v*+V [$_{v}$P *subject* [$_{v'}$ ~~*v*~~ [$_{VP}$ ~~V~~ ~~*object*~~]]]]
   [~~PROBE~~]

A similar dependence of phrasal movement on head movement characterizes 'asymmetric' verb second patterns like that of Standard German (den Besten 1983). Finite embedded clauses can contain either an overt complementizer *dass* (25a) or show verb second word order (25b). The complementizer *dass* cannot precede a verb second clause (25c) or be preceded by another phrase (25d).

(25) 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)

The pattern suggests that verb second order arises via movement of a tensed verb or auxiliary to the position of the complementizer, C. However, phrasal movement to Spec, CP occurs uniquely when C is filled by head movement. A more detailed analysis of verb second patterns is presented in Section 4.1.

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 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 will argue in Section 3 that the ability to license a specifier, the EPP property, is itself inherited from the lower, moved head.

We now turn to Matushansky's (2006) analysis of head movement, which will form 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 (16) is argued to be built in two steps. First, a lower head is attracted 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.[vii] The two heads are subsequently bundled into a complex head by an operation called M-Merger.

(26)
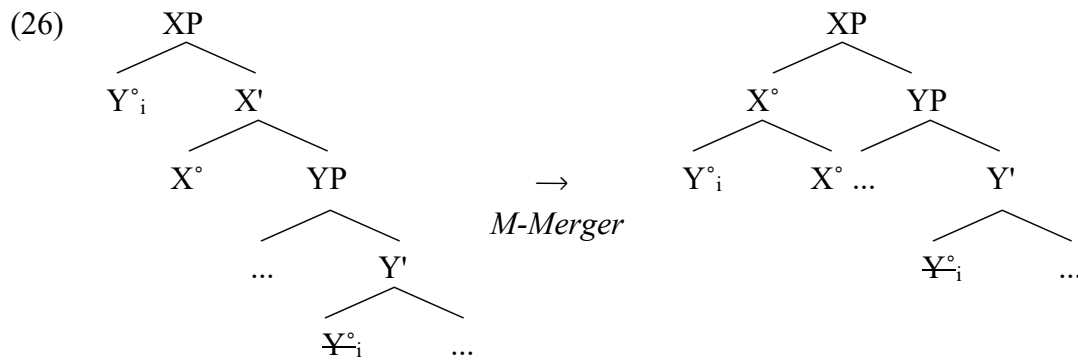
Matushansky proposes that head movement differs from phrasal movement in its featural trigger: phrasal movement is triggered by Agree (feature valuation), while head movement is triggered by c-selectional features of the attracting head. Because c-selection is a local relation between heads and their 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 can 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 (M. Baker 1988; Julien 2002; Roberts 2005).

Lastly, Matushansky proposes that M-Merger is a PF morphological operation that applies after Spell-out, rather than a syntactic one. The key motivation for this claim is to account for the impossibility of excorporation; Spell-out is expected to render the internal structure of the complex head opaque to later syntactic operations. This 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 the grammar in which syntactic and morphological operations are interleaved throughout the derivation, and in which constituents created by the morphological component $[X^o [Y^o X^o]]$ remain accessible to later syntactic

operations. To preview the proposals of Section 3, I will maintain Matushansky's two-step derivation of head movement as movement followed by bundling. However, the impossibility of incorporation 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 operate in the presyntactic lexicon and postsyntactic morphology, respectively. While it may turn out that similar operations can take place in various components of the grammar, the pursuit of a theory that minimizes the complexity of grammatical derivations motivates a more unified analysis. Here, 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 bundling in both Feature Scattering and M-Merger. 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 (to be defined in Section 3.1). The upshot of this 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 phenomena involve an interplay between "prominent" and "deficient" features. In Giorgi & Pianesi's (1997) Feature Scattering proposal, the number of functional projections that instantiates a given set of category features is determined by the number of prominent features that obligatorily project; all other features must be bundled. Head movement also requires a relationship between two types of features: a deficient feature in the target projection that must be bundled (Julien 2002, Roberts 2005), and a lower head with a prominent feature that enables bundling. Note as well that 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 & Pianesi (1997) for Feature Scattering. Although this is a less obvious observation 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.[viii] In essence, this can be viewed as a type of *prominence based licensing* pattern found in a range of phonological domains (Itô 1988; Goldsmith 1990; Steriade 1995; Walker 2011; among others).

Given these similarities, and the identical structural constraints on bundling, I propose that both bundling patterns can be unified as the result of one syntactic operation. To

preview the basic system to be presented in Section 3, all category features enter the derivation as distinct heads, but can be concatenated into complex heads as the result of a syntactic operation called *Coalescence*. Coalescence preceded by head displacement generates the effects of M-Merger, while the effects of Feature Scattering are determined by applications of Coalescence that are not immediately preceded by movement.

With a few exceptions (Giorgi and Pianesi 1997; Bobaljik and Thráinsson 1998), the two types of bundling are not generally discussed together in individual syntactic analyses. It is possible in many instances to investigate variation in the bundling of features independently of the mechanics of head movement, and vice versa. However, there are empirical advantages to a unified approach. Coalescence provides a parsimonious analysis of patterns that necessarily involve both types of bundling; for instance, where the target projection of head movement contains a bundle of probes that are each associated with distinct category features. I present an analysis of one such case in the context of verb second patterns in Section 4.1.

3   Coalescence

The proposal is framed within a bottom-to-top, derivational theory of syntax (Chomsky 1995, 2000, 2008). Each derivation begins with a selection of lexical items to be manipulated, a.k.a. the numeration. Lexical items are then combined into a hierarchical constituent structure by internal or external Merge. Syntactic operations are triggered by features of lexical items, and syntactic 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 propose that all lexical items contain exactly one

interpretable category feature [F] (i.e. features whose selection triggers external Merge), and assume 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 [$u\varphi$], [pres]). In this section, I define a bundling operation Coalescence and a feature system that determines when the operation applies.

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 definitions of c-command and head adjacency:[ix]

(27)   α c-commands β iff
        (i) α does not dominate β, and
        (ii) every node that dominates α also dominates β.

(28)   α asymmetrically c-commands β iff
        (i) α c-commands β, but
        (ii) β does not c-command α.

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

In this configuration, Coalescence creates a single node that contains all features associated with the individual heads.[x]

(30)

```
        XP                              X/YP
       /  \                            /    \
     X°    YP           →         X/Y°       ...
     [X]   /  \     Coalescence    [X]
         Y°    ...                 [Y]
         [Y]
```
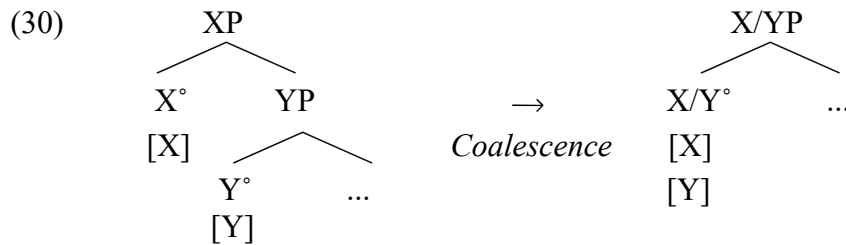
I depart from M-Merger and prior approaches to head adjunction by proposing that no branching structure is present in the newly formed head (I will leave it an open question as to whether feature bundles are subject to other ordering relations; cf. Cowper 2005). There are two primary motives for the elimination of internal branching structure: First, the absence of branching directly accounts for the impossibility of excorporation from bundled heads, without having to assume that bundling triggers Spell-Out to PF, as Matushansky proposes for M-Merger. Second, this representation obviates incompatibilities of traditional adjunction structures with the Uniformity Condition on chains (a point illustrated in more depth in Section 3.3). Section 5 returns 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.

(31)   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 are those that 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 (we will return to this point in 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 representations in surface phonology, it does not appear possible to state a phonological characterization on the types of phonologically dependent or affixal items that trigger head bundling to the exclusion of reduced forms that do not. For example, while triggers 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 throughout the derivation.[xi] Furthermore, not all syntactic heads that have affixal exponents undergo head bundling (Section 4.2 returns to this point in contrasts between contracted negation and reduced auxiliaries in English).

In addition to the distinction between dominant and recessive category features, a distinction is required between *dominant* versus *recessive heads*. While category features are lexically specified as to whether they are 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 (32), whereas a head that contains only recessive features is recessive (33). In all following examples, dominance is indicated with subscript $_D$, recessiveness with $_R$.

(32) X/Y°_D
    $[X_D]$
    $[Y_R]$

(33) X°_R
    $[X_R]$

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 Coalescence applies to ensure that no recessive heads remain at the end of the syntactic derivation. I will refer to this inviolable restriction as the Dominance Condition.

(34)   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 observations 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 domimant head to first move to the specifier of the target projection. To account for these restrictions, I propose that Coalescence applies uniquely in the configuration of head-adjacency in (35), where a dominant head asymmetrically c-commands a recessive one. This ensures that pairs of dominant heads cannot be bundled, and that a recessive head cannot trigger Coalescence with lower heads.[xii]

(35)   XP   X/YP
          /   \   /   \
        X°   YP   →   X/Y°   ...
       [X_D]   /   \   *Coalescence* [X_D]
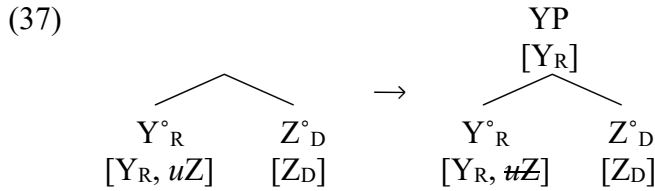   Y°   ...   [Y_R]
             [Y_R]
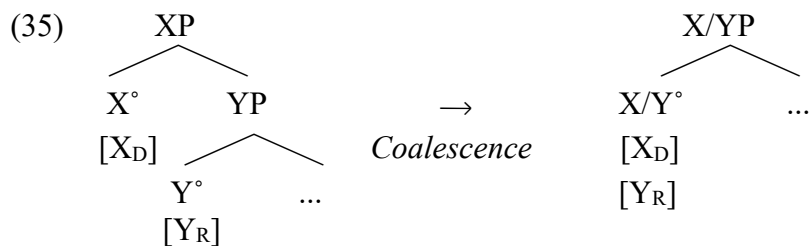
Here, I illustrate the derivation of a bundled X/Y° head. I assume that the order of projections arises from the distribution of uninterpretable c-selectional features (Svenonius 1994; Julien 2002; Matushansky 2006; Di Sciullo and Isac 2008). I further

assume that the inventory of category features and their c-selectional properties do not depend on whether they are dominant or recessive.

(36)   *Numeration:*
       X [X_D, uY]
       Y [Y_R, uZ]
       Z [Z_D]

In the first step, Y° c-selects Z°. Z° is dominant, as it contains a dominant category feature. Y° is a recessive head as it contains only recessive category features. Y° Merges with Z° to check [uZ] on Y°. I assume that the c-selecting head projects its features to the newly formed root node (Matushansky 2006). The root node thus becomes a projection of [Y]. Note that Coalescence cannot bundle Y_R° and Z_D° because Z_D does not asymmetrically c-command Y_R°.

(37)

$$
\begin{array}{ccc}
& & \text{YP} \\
& & [Y_R] \\
\overbrace{\qquad\qquad} & \rightarrow & \overbrace{\qquad\qquad} \\
Y°_R \qquad Z°_D & & Y°_R \qquad Z°_D \\
[Y_R, uZ] \quad [Z_D] & & [Y_R, \cancel{uZ}] \quad [Z_D]
\end{array}
$$

X°, whose category feature is dominant, then enters the derivation. Merge applies to check its [uY] feature and creates a new phrase headed by X.

(38)

$$
\begin{array}{ccc}
& & \text{XP}_D \\
& & [X_D] \\
\overbrace{\qquad\qquad} & & \overbrace{\qquad\qquad} \\
X°_D \qquad \text{YP}_R & & X°_D \qquad \text{YP}_R \\
[X_D, uY] \quad [Y_R] & \rightarrow & [X_D, \cancel{uY}] \quad [Y_R] \\
\quad \overbrace{\qquad} & & \qquad \overbrace{\qquad} \\
\quad Y°_R \qquad Z°_D & & \quad Y°_R \qquad Z_D \\
\quad [Y_R, \cancel{uZ}] \quad [Z_D] & & \quad [Y_R, \cancel{uZ}] \quad [Z_D]
\end{array}
$$
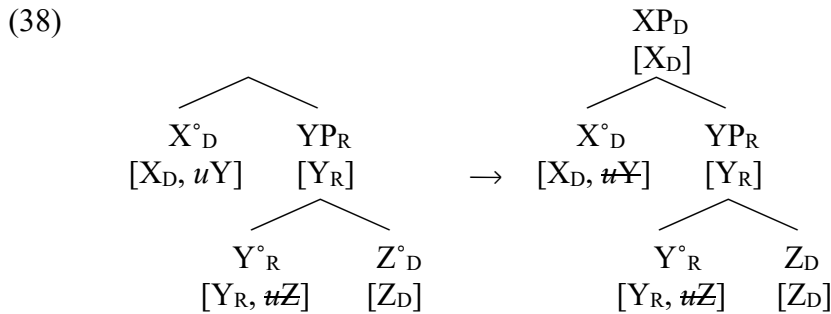
Because this step creates a head adjacency configuration where the lower head is recessive, Coalescence applies. The bundled X/Y°$_D$ head is dominant because it contains the dominant feature [X$_D$], and no recessive heads remain in the workspace.

(39)

$$
\begin{array}{ccc}
\text{XP}_D & & \text{X/YP}_D \\
[\text{X}_D] & & [\text{X}_D] \\
& & [\text{Y}_R] \\
\end{array}
$$

XP$_D$
[X$_D$]
   X°$_D$ YP$_R$
[X$_D$, *uY*] [Y$_R$]

        Y°$_R$ Z°$_D$
    [Y$_R$, *uZ*] [Z$_D$]

$\rightarrow$

Coalescence

X/YP$_D$
[X$_D$]
[Y$_R$]
   X/Y°$_D$ Z°$_D$
   [X$_D$, *uY*] [Z$_D$]
   [Y$_R$, *uZ*]
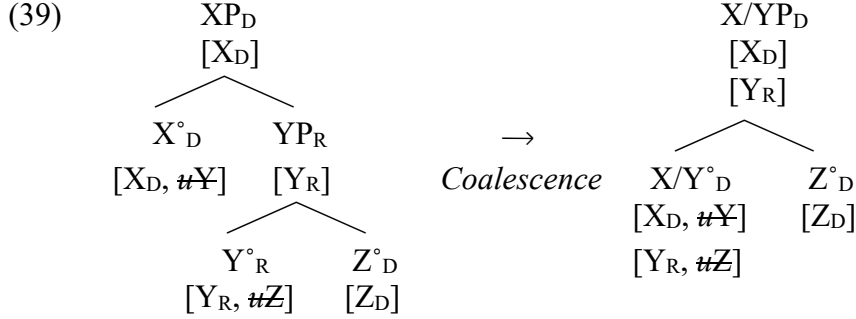
At this point, the question arises of how the label of the highest node is determined. Maintaining the assumption that all features of the selecting head project, both [X] and [Y] will project to the root node. One concern that arises here is that the structure runs afoul of the expectation that syntactic operations do not alter the labels of syntactic objects that they manipulate, the No Tampering Condition (Chomsky 2008).[xiii] 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).

(40)   CATEGORY PERCOLATION
        Given an extended projection $\Phi = \{\Pi_n > \Pi_{n-1} > ... \Pi_1\}$, 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 features of all lower heads within the same extended projection (Keine 2019; 38, fn. 21).

This is schematically represented in (41) for a structure in which the features [X], [Y], and [Z] are in the same extended projection; subfeatures are shown in curly brackets.

(41)
```
              XP {X, Y, Z}
            /            \
        X°               YP {Y, Z}
      {X, Y, Z}         /         \
                    Y°            ZP {Z}
                  {Y, Z}         /      \
                             Z°         ...
                            {Z}
```
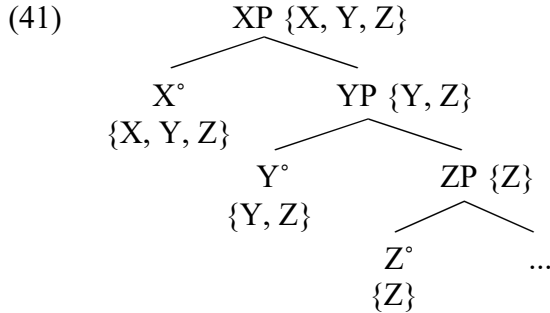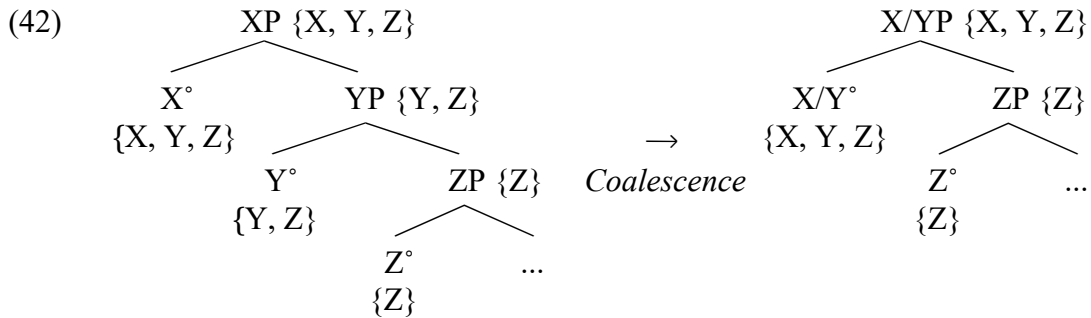
What is key for this analysis is that Coalescence of heads in an extended projection cannot produce head or root node labels distinct from those that would exist if no bundling applied, in compliance with the No Tampering Condition. This is shown in (42), where the highest phrasal node contains features [X, Y, Z]; The highest node contains the same features even if Coalescence applies to X° and Y°, because the set of features in "YP", {Y, Z}, is a subset of those contained in "XP", {X, Y, Z}.

(42)
```
        XP {X, Y, Z}                                      X/YP {X, Y, Z}
       /          \                                      /           \
   X°             YP {Y, Z}                          X/Y°            ZP {Z}
 {X, Y, Z}       /        \            →           {X, Y, Z}        /      \
             Y°           ZP {Z}    Coalescence                 Z°        ...
           {Y, Z}        /      \                              {Z}
                      Z°        ...
                     {Z}
```

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.

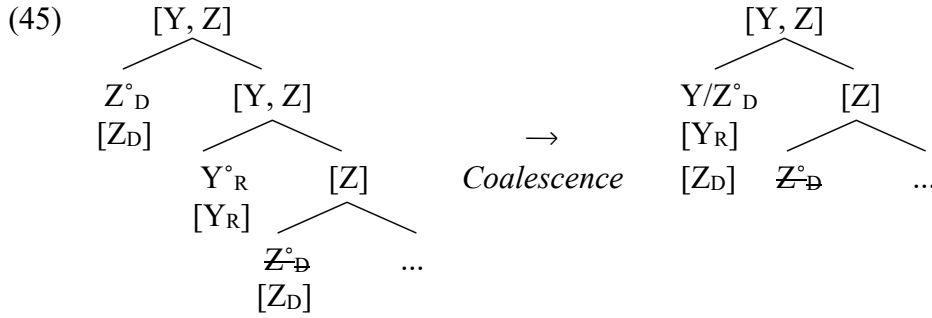Lastly, complex heads that contain more than one recessive category feature are generated by successive "top-down" applications of Coalescence. At each step in (43) and (44), Coalescence applies to the topmost dominant head the immediately lower recessive head. For representational simplicity, checked c-selection features and features of phrasal nodes are omitted in the remainder of the paper.

(43)

XP_D → (Coalescence) → X/YP

X°_D YP_R $\quad\rightarrow\quad$ X/Y°_D ZP
$\qquad\qquad$ *Coalescence* $\quad [\text{X}_D]$
$\text{Y}^{\circ}_R \quad \text{ZP}_R \qquad\qquad\qquad [\text{Y}_R] \quad \text{Z}^{\circ}_R \qquad ...$
$\qquad \text{Z}^{\circ}_R \qquad ...$

(44)

X/YP_D $\quad\rightarrow\quad$ X/Y/ZP

$\text{X/Y}^{\circ}_D \quad \text{ZP}_R \qquad\qquad \text{X/Y/Z}^{\circ}_D \qquad ...$
$[\text{X}_D] \qquad\qquad$ *Coalescence* $\quad [\text{X}_D]$
$[\text{Y}_R] \quad \text{Z}^{\circ}_R \qquad ... \qquad\qquad\qquad [\text{Y}_R]$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [\text{Z}_R]$
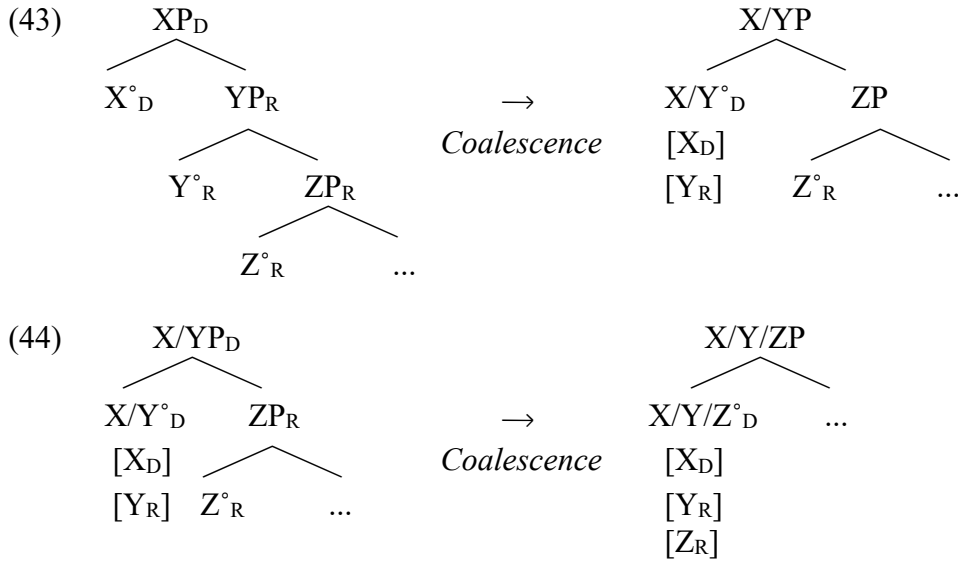
3.3 Coalescence and head movement

The previous discussion shows the application of Coalescence following external Merge of a dominant head. The same head-adjacency configuration can be created by internal Merge of a lower dominant head. This is schematically shown in (45), in which Z°_D first moves to the specifier of YP, providing the head-adjacency configuration that triggers Coalescence. Again, Category Percolation ensures that the root node following Coalescence is featurally identical to the root node prior to movement.

(45)

```
        [Y, Z]                                      [Y, Z]
        /    \                                      /    \
    Z°_D    [Y, Z]                           Y/Z°_D      [Z]
    [Z_D]   /    \              →            [Y_R]      /   \
         Y°_R    [Z]        Coalescence      [Z_D]   Z̶°̶_̶D̶    ...
         [Y_R]  /   \
              Z̶°̶_̶D̶   ...
              [Z_D]
```

Section 4 discusses several cases where languages vary in whether the application of Coalescence to a recessive head is fed by internal or external Merge.

Like Matushansky (2006), I propose that head movement first requires the movement of a lower head to the specifier of the target projection. However, I propose that movement of a dominant head to the specifier of a recessive one is not triggered by a probe or selecting property of the target head, but 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). A key consequence of the proposal is that there is no need for the grammar to include a locality requirement on head movement, such as the Head Movement Constraint (Travis 1984). As further discussed in Section 5.1, if the Dominance Condition only applies upon spell-out of a phase, we predict that Last Resort head movement need not take place to a projection that is not the highest one in its spell-out domain. Crucially however, even in structures in which 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 in adjacent projections at the end of the derivation.[xiv] This is illustrated in (46) for a
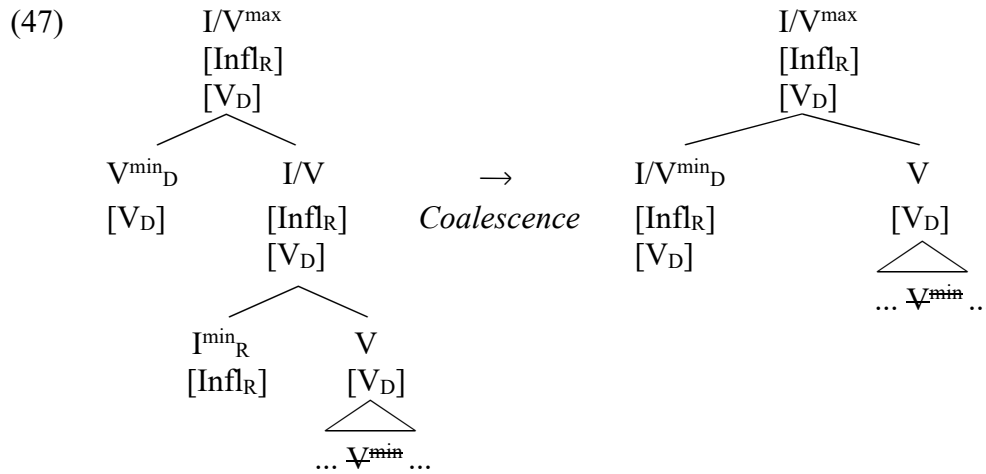
structure in which the dominant head $Z°_D$ closest to the target $X°_R$ crosses an intervening projection headed by $Y°_R$. Coalescence first bundles $Z°_D$ and $X°_R$, before bundling $X/Z°_D$ and $Y°_R$.

(46) a.

```
        XP                                              X/ZP
      /    \                                          /      \
   Z°_D     X'                      →          X/Z°_D         YP
   [Z_D]   /  \               Coalescence      [X_R]        /    \
        X°_R   YP                              [Z_D]    Y°_R      ZP
        [X_R]  /  \                                     [Y_R]    /  \
           Y°_R    ZP                                        Z̶°̶_̶D̶     ...
           [Y_R]   /  \
                Z̶°̶_̶D̶    ...
```

b.

```
        X/ZP                                             XP
       /     \                                         /    \
  X/Z°_D      YP                      →         X/Y/Z°_D      ZP
  [X_R]      /  \               Coalescence     [X_R]       /  \
  [Z_D]  Y°_R    ZP                             [Y_R]    Z̶°̶_̶D̶    ...
         [Y_R]   /  \                           [Z_D]
              Z̶°̶_̶D̶    ...
```
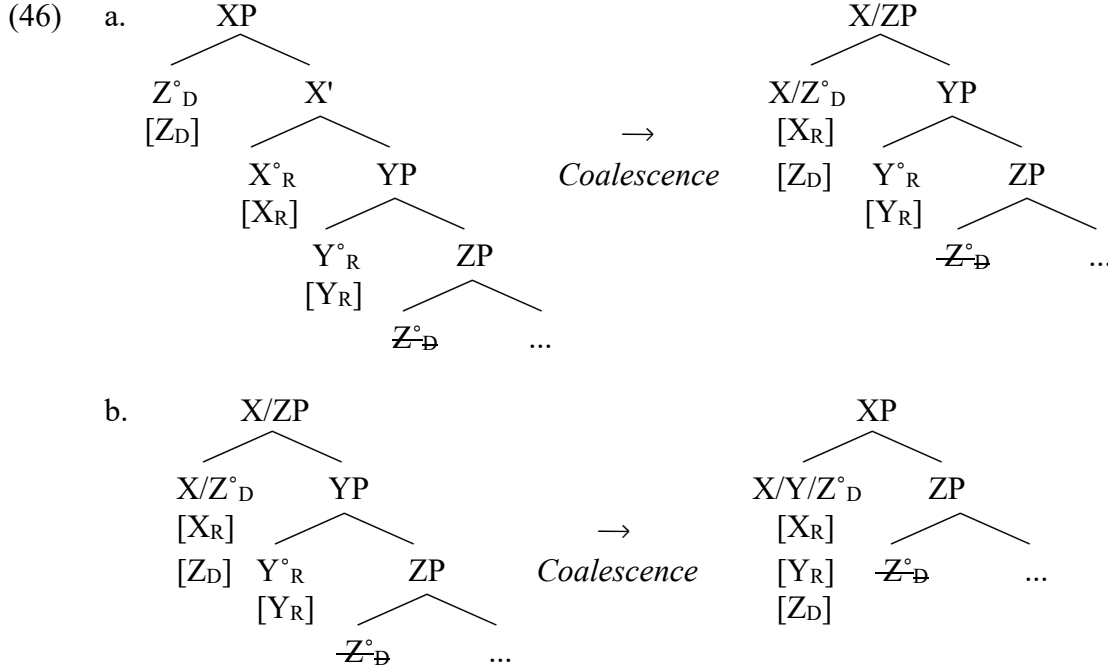
Category Percolation, in concert with the No Tampering Condition, accounts for the generalization that lexical heads can move to functional heads, but functional heads cannot move to lexical heads (Li 1990; Delsing 1993; M. C. Baker 1996; M. C. Baker 2003). For example, it is possible for V° to move to C° and undergo Coalescence; as the features of C° ([C, T, V]) include [V], the features contained in the root node remain unchanged. On the other hand, C° cannot move to a higher head V° because V° does not contain the features [C, T]; Coalescence would result in the addition of features to the root node, in violation of the No Tampering Condition.

Finally, we can address why the proposed structure for head movement satisfies the Uniformity Condition. This rests on two claims: First, that root nodes inherit the category

features of lower heads within their extended projection (Category Percolation); Second, that complex heads formed by Coalescence include no internal branching structure. To illustrate, consider a simplified derivation of a bundled V/Infl head in (47), shown with relational definitions of phrase labels. In the first step, dominant V^{min} moves to the specifier of I^{max}. Because the highest node contains both [V] and [Infl] features in accordance with Category Percolation, it is a maximal projection of both categories (I^{max} and V^{max}). This remains unchanged after Coalescence applies. However, the bundled head is now a minimal projection of both [V] and [Infl]. The two links of the movement chain are minimal projections of [V], in accordance with the Uniformity Condition, and the bundled head c-commands its lower position.

(47)

$$
\begin{array}{cc}
I/V^{max} & I/V^{max} \\
[Infl_R] & [Infl_R] \\
[V_D] & [V_D]
\end{array}
$$

Left tree:
- I/V^{max} [Infl$_R$] [V$_D$]
  - $V^{min}{}_D$ [V$_D$]
  - I/V [Infl$_R$] [V$_D$]
    - $I^{min}{}_R$ [Infl$_R$]
    - V [V$_D$]
      - ... $\overline{V^{min}}$ ...

$\rightarrow$ *Coalescence*

Right tree:
- I/V^{max} [Infl$_R$] [V$_D$]
  - $I/V^{min}{}_D$ [Infl$_R$] [V$_D$]
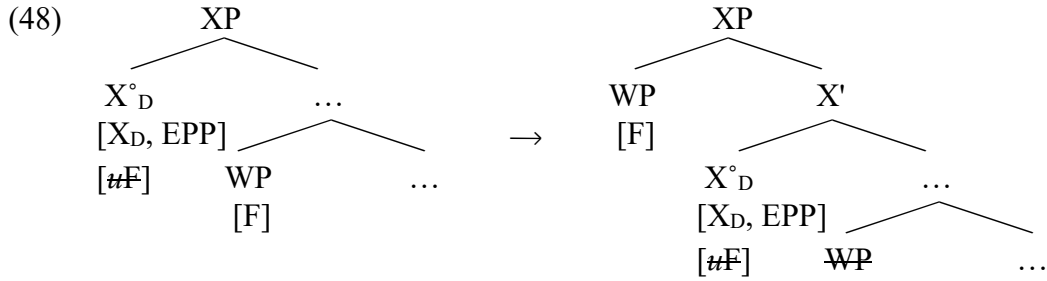 - V [V$_D$]
    - ... $\overline{V^{min}}$ ...

Later head movement of the complex head will similarly produce only minimal projections, and there is no need to posit movement of intermediate-level projections. Coalescence and Category Percolation thus provide a means to generate head movement in accordance with the Extension and Uniformity Conditions.[xv]

3.4 Coalescence and phrasal movement

Like Matushansky (2006), I propose that both head movement and phrasal movement involve displacement of the moved item 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 (2006), 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.
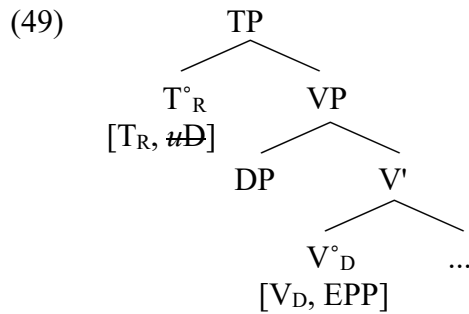
In addition, I propose that phrasal movement takes place only if the attracting head has the EPP property, which can be defined as the ability to trigger Merge of an item that does not contain a subset of its category features. Specifically, uninterpretable [uF] probes can be checked if they c-command a suitable goal, but will only trigger movement of the goal if the probing head also contains an [EPP] feature (Chomsky 2001).[xvi] 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 also carry an [EPP] feature. The necessary configuration for phrasal movement is shown in (48) below.
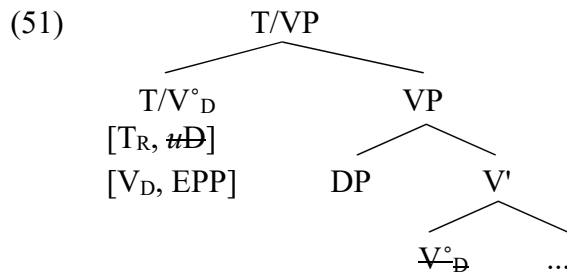
(48)

```
        XP                                    XP
       /  \                                  /  \
   X°_D    …                              WP      X'
  [X_D, EPP]                             [F]      / \
   [u̶F̶]   WP        …        →                 X°_D      …
          [F]                              [X_D, EPP]   / \
                                             [u̶F̶]    W̶P̶      …
```
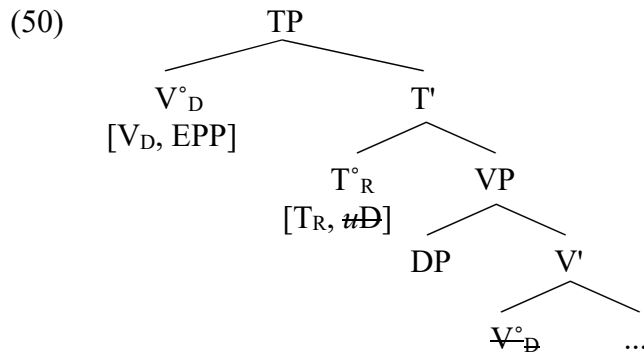
Here, I claim that delayed gratification patterns motivate a reconception of the EPP property that diverges from most prior approaches. In these patterns, a probe associated with a (recessive) target head is only able to trigger phrasal movement after it has undergone Coalescence with a moved, dominant head. I account for this with the following claims: First, while the presence of [EPP] enables its projection to have a phrasal specifier, no specifier is Merged if the head does not also contain a relevant [uF] 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 need not be immediately deleted from the derivation (Pesetsky & Torrego 2000). To preview the analysis of delayed gratification patterns, a [uF] probe on a recessive head is first checked by c-commanding an appropriate goal. A dominant head with [EPP] then moves and undergoes Coalescence with the recessive head, then enabling the checked [u̶F̶] probe to trigger phrasal movement in concert with [EPP].

The basic workings of head movement, Coalescence, agreement, and phrasal movement can be succinctly illustrated with a schematic example of Romance-style V-to-T head movement. In this and all subsequent examples, I assume that all V heads are
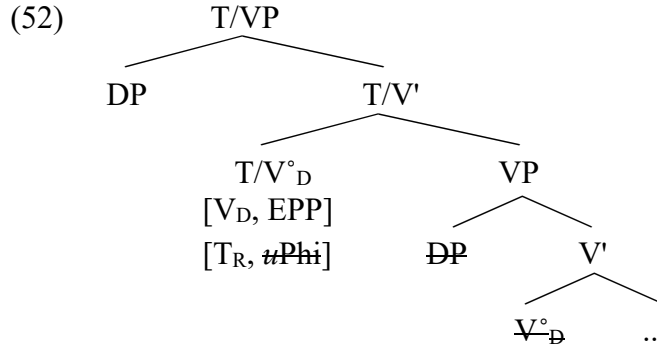
dominant and that they have the EPP property, a reinterpretation of Baker's (2003) claim that the defining property of verbs as a lexical category is their ability to license specifiers.[xvii] In the first step (49), a recessive T° with a [uD] probe is Merged upon the completion of VP, which includes a dominant V° and a subject DP in Spec, VP. At this point, [uD] on T° is checked by agreement with the subject DP. However, phrasal movement of the subject does not take place since T° lacks [EPP].

(49)

```
              TP
           /      \
        T°_R       VP
      [T_R, u̶D̶]   /  \
              DP    V'
                   /  \
                V°_D   ...
              [V_D, EPP]
```

Here, the Dominance Condition is satisfied by moving V° to Spec, TP (50). As dominant V°_D now c-commands recessive T°_R, Coalescence creates a bundled T/V° head (51).

(50)

```
                 TP
              /      \
          V°_D        T'
        [V_D, EPP]   /   \
                  T°_R    VP
                [T_R, u̶D̶] /  \
                        DP    V'
                             /  \
                           V̶°̶_̶D̶   ...
```

(51)

```
                 T/VP
              /        \
         T/V°_D         VP
        [T_R, u̶D̶]      /  \
        [V_D, EPP]   DP    V'
                          /  \
                        V̶°̶_̶D̶   ...
```

At this point, the bundling of V° and T° into a single head T/V° enables the checked [~~uD~~]

probe to make use of the EPP feature associated with [V$_D$], triggering phrasal movement

of the subject to Spec, TP.

(52)

```
                    T/VP
              ┌──────────┴──────────┐
             DP                    T/V'
                            ┌────────┴────────┐
                        T/V°_D                VP
                       [V_D, EPP]        ┌─────┴─────┐
                       [T_R, uPhi]      DP          V'
                                              ┌──────┴──────┐
                                            V°_D           ...
```

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).

It should be noted that the Last Resort analysis of head movement does not extend to

proposed cases of "long head movement" where head movement appears to skip over a

potential intervener [X°$_i$ ... [Y° ... [X°$_i$ ...]]] (Rivero 1991; Borsley et al. 1996; Roberts

2010 and references therein). However, these patterns can potentially arise under

conditions where the target projection has a probing dominant head with an EPP feature

whose goal can be a head. In the best-known cases of participle fronting in many Slavic

languages (Lema and Rivero 1989; Wilder and Ćavar 1994; Bošković 1995; Embick and

Izvorski 1995; Migdalski 2006, a.o.) and Breton (Rivero 1991; Borsley et al. 1996;

Roberts 2010), the target position can be filled by either a head or a phrase.

(53) a. Az sŭm   čel   kniga-ta (*Bulgarian*: Migdalski 2006)
          I   be.AUX.PRES.1SG   read.PART.F.SG book-the

b. Čel   sŭm   kniga-ta
   read.PART.F.SG be.AUX.PRES.1SG   book-the
   'I have read the book'

(54) 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.'

It is worth noting that under the present proposal, long head movement cannot take place to satisfy the Dominance Condition, and must be triggered by a dominant target head with [EPP]. This is because Dominance Condition triggered movement attracts the closest dominant head within the extended projection, and must be followed by Coalescence. There is indeed no evidence from these cases that the moved and target heads (e.g. the fronted participle + auxiliary sequence in Bulgarian) can move together in later steps of the derivation, as would be expected if they were bundled after movement. This has some similarities to the analysis of Harizanov and Gribanova (2018), who argue that only long head movement involves the syntactic displacement of heads, whereas "short" head movement is postsyntactic morphological amalgamation. Although their approach also accounts for the differences in locality and bundling between long and "short" head movement, it does not explain cases in which apparently local head movement affects the licensing of phrasal movement to the target projection.

4   Case studies

This section illustrates key aspects of the proposals on Coalescence and movement in several cases of variation in the realization of functional projections in the inflectional

and complementizer domains. It considers two instances in which languages permit a given head to have both dominant and recessive variants, the Kashmiri V2/V3 alternation and the realization of sentential negation in English. In addition, within-language variation in the realization of the Catalan past perfect is attributed to differences in the distribution of dominant and recessive features.

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

(55)   [$_{CP}$ XP   V-C°   [$_{IP}$ ... ~~V°~~ ... ~~XP~~ ...

However, Rizzi (1997) and later research in the Cartographic Program have produced a range of evidence that the traditional "CP" contains a series of projections associated with clause type and information structure features, as illustrated in the "core" structure of Rizzi (1997). 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 simulateously instantiated in phrase structure.

(56)   [$_{ForceP}$ ... [$_{TopicP}$ ... [$_{FocusP}$ ... [$_{FinitenessP}$ ... [$_{IP}$
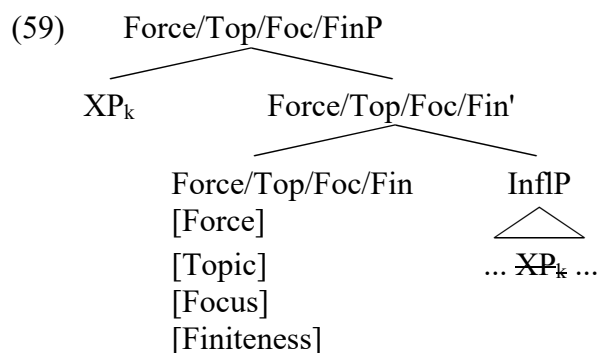
In addition, there are a variety of 'relaxed' V2 languages that show verb movement to the C-domain, but allow more than one phrase to precede the verb (either optionally or obligatorily). In these languages, the ordering of pre-verbal constituents is restricted based on discourse properties, consistent with the order in (56) (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*.

(57) Ingush: Topic + Focus + V
      [Jurta   jistie] [joaqqa   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.

(58) 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.

(59)
```
        Force/Top/Foc/FinP
         /              \
      XPₖ        Force/Top/Foc/Fin'
                  /              \
          Force/Top/Foc/Fin      InflP
          [Force]                 /\
          [Topic]             ... X̶P̶ₖ ...
          [Focus]
          [Finiteness]
```
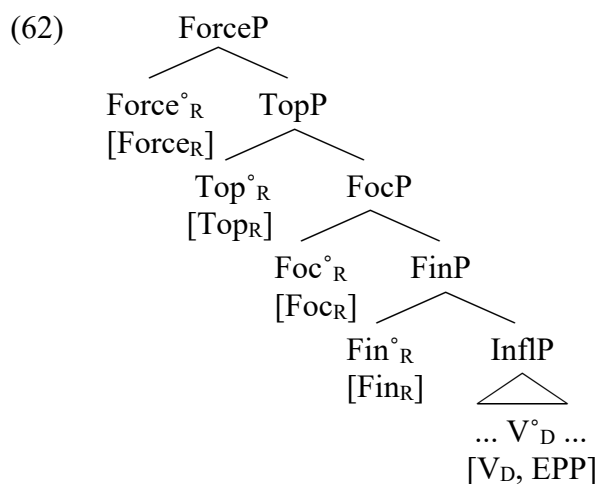
Relaxed V2 patterns are generated when left-peripheral features are distributed across multiple heads, whose hierarchical order remains consistent with those proposed in strictly Cartographic analyses.
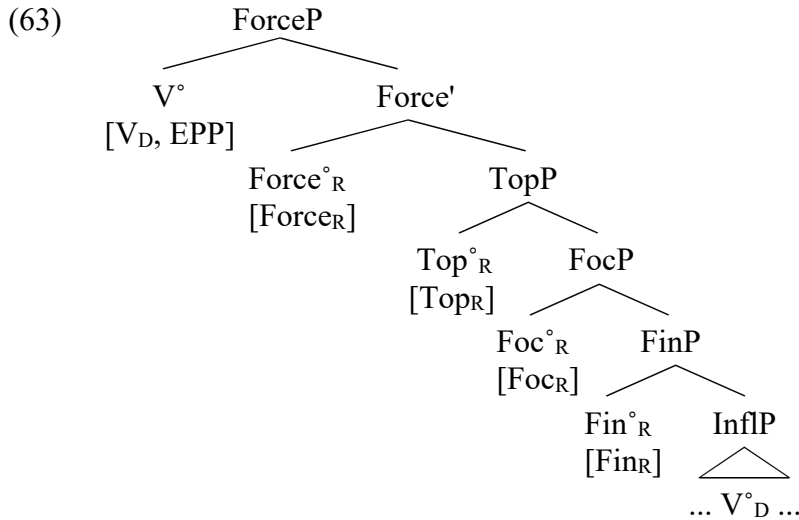
(60) Ingush *Topic Focus V ...*
[Force/TopP XP_top   [Foc/FinP (XP_foc)   V-Foc/Fin°   [IP ... V̶°̶

(61) Old Italian *Frame-setter Topic Focus V ...*
[ForceP XP_frame [TopP (XP_foc) [Foc/FinP (XP_top)   V-Foc/Fin°   [IP ... V̶°̶
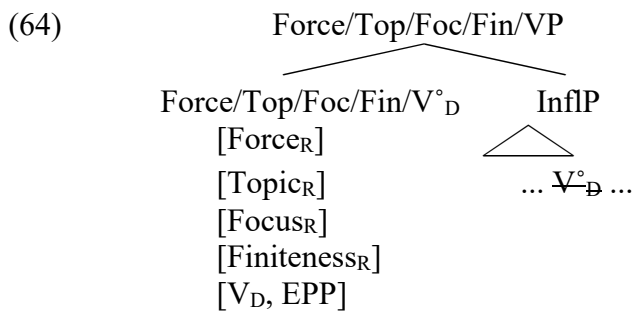
I now present a derivational analysis of these patterns in terms of Coalescence. To generate the idealized strict V2 pattern, suppose that all left-peripheral category features are recessive, and a recessive head corresponding to each feature is Merged.

(62)
```
              ForceP
             /      \
       Force°_R      TopP
       [Force_R]     /   \
               Top°_R     FocP
               [Top_R]    /   \
                    Foc°_R      FinP
                    [Foc_R]     /   \
                         Fin°_R      InflP
                         [Fin_R]      /\
                                  ... V°_D ...
                                  [V_D, EPP]
```

41

If the derivation were to end here, the structure would not be interpretable at PF, as it contains recessive heads. One strategy to avoid this (the only one available in main clauses) is Last Resort movement of $V°_D$ to Spec, ForceP.

(63)

```
                    ForceP
            ┌──────────┴──────────┐
          V°                    Force'
      [V_D, EPP]         ┌────────┴────────┐
                      Force°_R            TopP
                      [Force_R]      ┌─────┴─────┐
                                  Top°_R        FocP
                                  [Top_R]   ┌────┴────┐
                                          Foc°_R     FinP
                                          [Foc_R] ┌───┴───┐
                                               Fin°_R    InflP
                                               [Fin_R]   ◁
                                                      ... V°_D ...
```

This produces the environment for Coalescence, which applies iteratively until all recessive features are bundled into a dominant head.

(64)

```
                Force/Top/Foc/Fin/VP
           ┌──────────────┴──────────────┐
      Force/Top/Foc/Fin/V°_D           InflP
            [Force_R]                     ◁
            [Topic_R]                ... V̶°̶_̶D̶ ...
            [Focus_R]
            [Finiteness_R]
            [V_D, EPP]
```
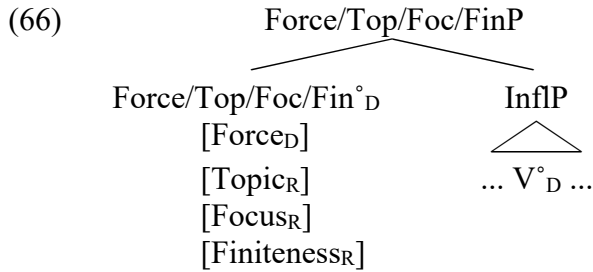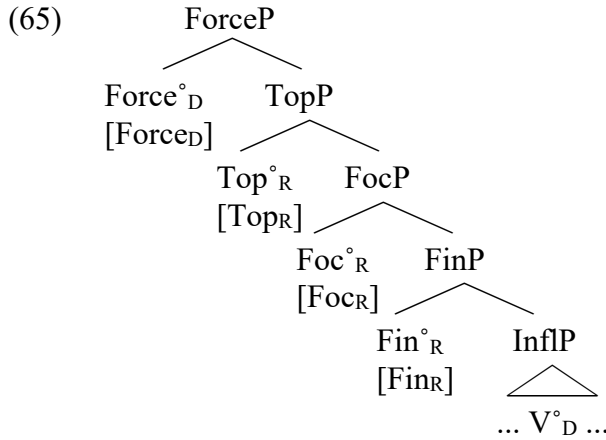
The association of the [EPP] property with $[V_D]$, rather than a left-peripheral feature, accounts for the fact that subsequent phrasal movement to Spec, ForceP has several possible pragmatic functions. In Standard German the first position can be occupied by either a given information topic, contrastive 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 these can be

checked via Agree with a constituent lower in the clause.[xviii] However, phrasal movement occurs only if the probe is found on a dominant head with [EPP], a situation that arises only as the result of 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 the [EPP] feature on V° makes an additional prediction about word order in nominal extended projections. My translation of Baker's (2003) argument that only verbs license specifiers into the claim that all V°$_D$ heads have [EPP] also implies that N° heads lack [EPP], even if they are dominant. Consequently, even if N° movement to the highest DP projection is possible, as is likely the case for languages with noun-initial order within DP, we do not expect to find "noun second" patterns in which nouns are preceded by exactly one syntactic phrase within DP, because N° never carries [EPP]. To the best of 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 (65). This allows Coalescence to apply, obviating verb movement out of InflP (66).

(65)

```
                    ForceP
                   /      \
          Force°_D         TopP
          [Force_D]       /    \
                    Top°_R       FocP
                    [Top_R]     /    \
                          Foc°_R      FinP
                          [Foc_R]    /    \
                               Fin°_R      InflP
                               [Fin_R]     /\
                                          /  \
                                       ... V°_D ...
```

(66)

```
                 Force/Top/Foc/FinP
                /                  \
    Force/Top/Foc/Fin°_D            InflP
         [Force_D]                   /\
         [Topic_R]                  /  \
         [Focus_R]              ... V°_D ...
         [Finiteness_R]
```

Even if [Topic_R], [Focus_R], and [Finiteness_R] are associated with the same probes as before, agreement cannot trigger phrasal movement, because [Force_D] is not associated with the [EPP] property.

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° head that lacks [EPP].[xix] V2 clauses are generated by movement of dominant V_D°, with subsequent movement to first position permitted by [EPP] on V_D°.[xx]

(67)   a. Er sagte [dass er morgen   kommt]
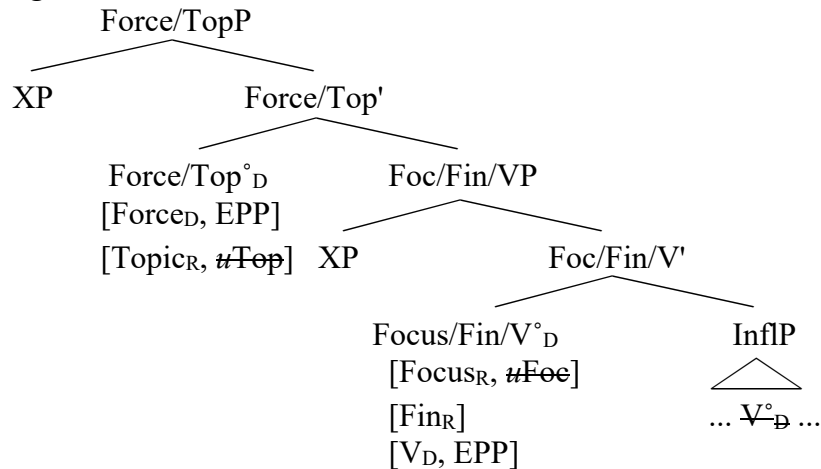   He said   that   he tomorrow comes

b. Er sagte [er kommt morgen]
   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

(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 independent 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.

(68)   *Ingush V3*


```
                    Force/TopP
                   /          \
               XP              Force/Top'
                              /          \
                    Force/Top°D           Foc/Fin/VP
                    [ForceD, EPP]         /        \
                    [TopicR, u̶T̶o̶p̶]  XP          Foc/Fin/V'
                                              /            \
                                    Focus/Fin/V°D          InflP
                                    [FocusR, u̶F̶o̶c̶]         /\
                                    [FinR]               ... V̶°̶D̶ ...
                                    [VD, EPP]
```

Note that derivation requires verb movement to take place 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. I suggest the following explanation: First, suppose that the Dominance Condition is enforced on projections that constitute phasal spell-out domains (returned to Section 5.1). Further assuming that languages can vary as to whether a given projection is a phase (Abels

2003), Ingush may differ from German in that the projection with [Topic] is a phase head only in Ingush. 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).

(69)   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. In the descriptions of 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.

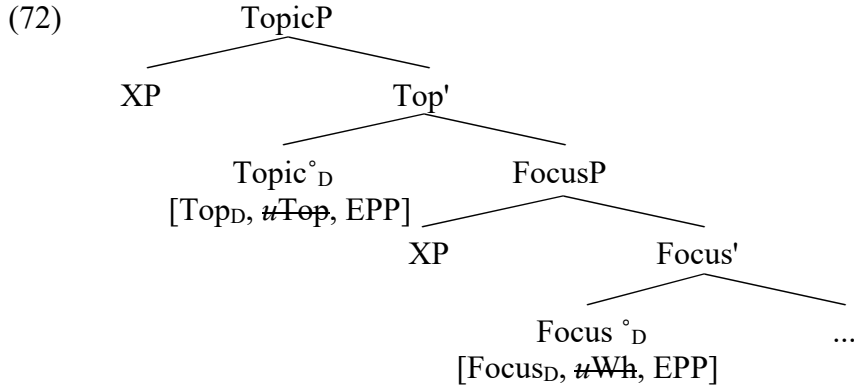(70) a. tse   kyaa   dyutnay   Rameshan
       you   what   gave   Ramesh
       'As for you, what is it that Ramesh gave?'

      b. ?kyaa   dyutnay   Rameshan tse
        what   gave   Ramesh   you   (Bhatt 1999)

Munshi & 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 [Topic] and [Focus] in separate projections. Because both heads contain [EPP], this creates the V3 word order in interrogative clauses.

(71) Numeration for V3
[Focus$_D$, uWh, EPP]
[Topic$_D$, uTop, EPP]

(72)

```
                    TopicP
            _____|_____
          XP               Top'
                    _____|_____
              Topic°_D            FocusP
          [Top_D, u̶T̶o̶p̶, EPP]  ____|____
                            XP          Focus'
                                    ____|____
                                Focus °_D      ...
                            [Focus_D, u̶W̶h̶, EPP]
```

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

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

(73) Numeration for V2
[Focus$_R$, uContrast]
[Topic$_D$, uTop, EPP]

(74)

```
              TopicP                               Topic/FocusP
        _____|_____                        _____|_____
    Topic°_D          FocusP      →        Topic/Focus°_D        ...
[Topic_D, u̶T̶o̶p̶, EPP]  ___|___           [Topic_D, u̶T̶o̶p̶, EPP]
               Focus°_R      ...           [Focus_R, u̶C̶o̶n̶t̶r̶a̶s̶t̶]
          [Focus_R, u̶C̶o̶n̶t̶r̶a̶s̶t̶]
```
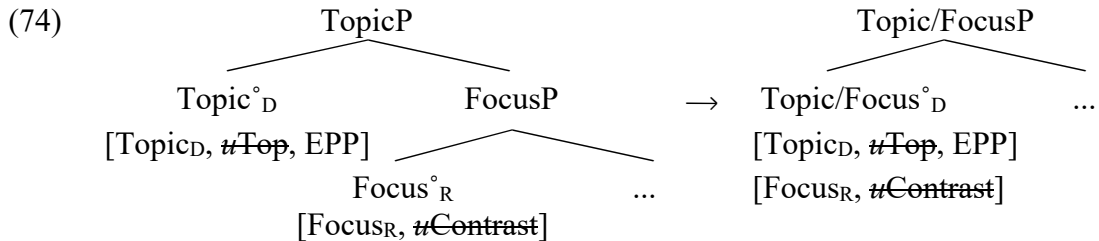
In this case, I suggest that the contrast between varieties that permit either a topic or a focus in first position (Munshi & Bhatt 2009) and those that permit only a focus in first position (Bhatt 1999, Manetta 2011) results from a difference in the priority with which probing features in the bundled head can trigger phrasal movement in conjunction with [EPP]. In the former case, either [uTop] or [uContrast] can trigger phrasal movement, whereas in the latter variety [uContrast] must take precedence over [uTop], even if [EPP] is originally associated with [Topic$_D$].[xxi]

Although there are other languages that permit *XP$_{Topic}$ XP$_{Wh}$ V* verb-third order, but not **XP$_{Topic}$ XP$_{Focus}$ V*, such as Badiotto Rhaetoromance (Poletto 2002) and Yiddish (Diesing 2005), the opposite pattern remains unattested, suggesting that no language can contain both [Focus$_D$, uContrast] and [Focus$_R$, uWh]. Such patterns can be explained by the proposal of Cinque (1999) that all functional features can be associated with either a marked or default value. Suppose that [Focus] always has [uContrast] as its default value and [uWh] as its marked value. The apparent restriction is that it is not possible for a category feature with a marked value to be recessive, while the same feature with the unmarked value is dominant. It is possible however, for both instances of a category head to be recessive or for both to be dominant (as is the case for [Focus] in Ingush).

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.

(75) 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 (76a). This gives the effect of contraction feeding raising. On the hand, only the full form is possible if the negative remains in a post-subject position (76b). 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.

(76) 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 bundling by M-Merger optionally applies to Neg˚ and Aux˚ as soon as the items are Merged, and that M-Merged Neg˚ corresponds to *n't*. If Aux˚ and Neg˚ are M-Merged prior to movement, 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]
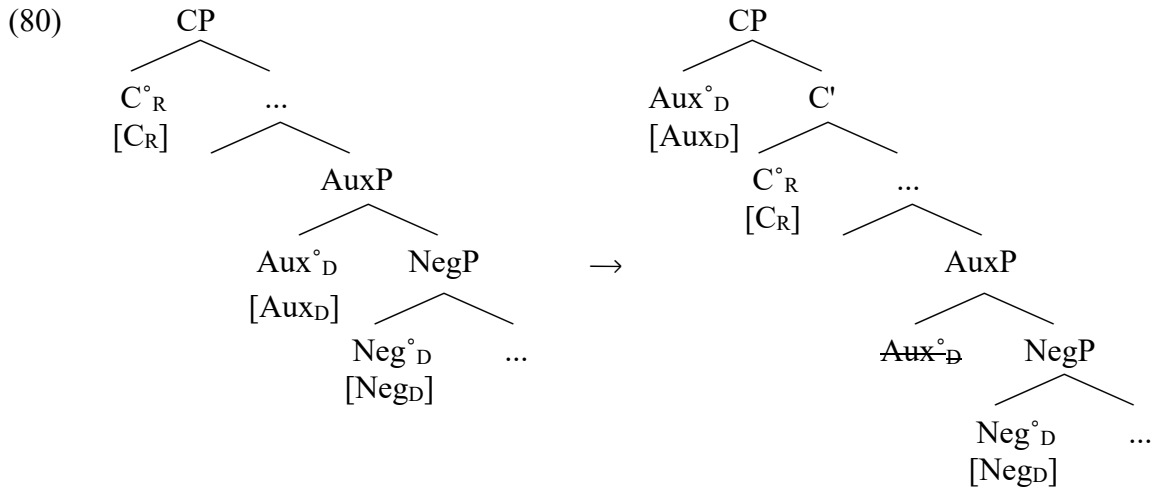
This analysis can be reframed using Coalescence as follows. I first illustrate the basic analysis with a structure in which Aux-to-T movement has already applied, and Aux˚ immediately c-commands Neg˚. The different distribution of the full and contracted

forms is accounted for if the full form enters the derivation with a dominant category

feature [Neg$_D$], while the contracted form is first Merged with a recessive feature [Neg$_R$].

Recessive Neg°$_R$ undergoes Coalescence with the dominant auxiliary (77). Thus, when

recessive C° attracts the closest dominant head, it attracts bundled Aux/Neg° (78).

(77)   AuxP   Aux/NegP

Aux°$_D$ NegP   →   Aux/Neg°$_D$ ...
[Aux$_D$]   *Coalescence*   [Aux$_D$]
          Neg°$_R$ ...   [Neg$_R$]
          [Neg$_R$]

(78)   CP   CP

C°$_R$ ...   Aux/Neg°$_D$ C'
[C$_R$]   [Aux$_D$]
              Aux/NegP   [Neg$_R$]   C°$_R$ ...
                                                     [C$_R$]
        Aux/Neg°$_D$ ...   →   Aux/NegP
        [Aux$_D$]
        [Neg$_R$]   ~~Aux/Neg$_D$~~   ...
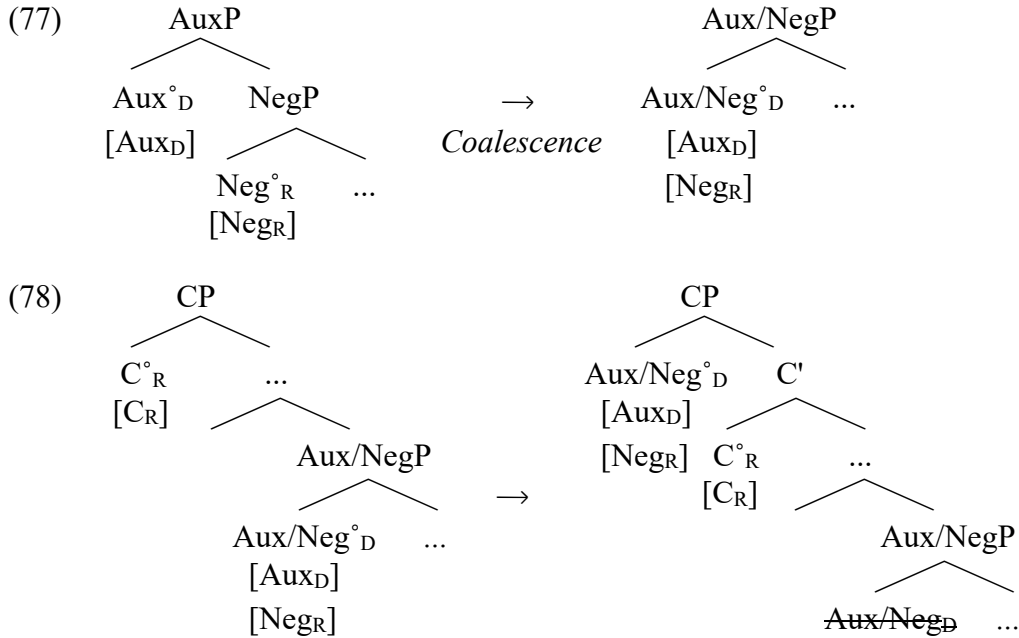
On the other hand, dominant Neg°$_D$ that corresponds to the full form does not undergo

Coalescence. In inversion contexts, interrogative C° thus attracts Aux$_D$°, leaving Neg$_D$° in

its first-Merge position (80).

(79)   AuxP

Aux°$_D$ NegP
[Aux$_D$]   *Coalescence does not apply*
          Neg°$_D$ ...
          [Neg$_D$]

(80)

```
           CP                                      CP
          /  \                                    /  \
        C°R    ...                             Aux°D    C'
        [CR]     \                             [AuxD]   /  \
                AuxP              →                    C°R    ...
               /    \                                  [CR]     \
           Aux°D    NegP                                       AuxP
           [AuxD]   /  \                                      /    \
                 Neg°D  ...                                Aux°D    NegP
                 [NegD]                                   ‾‾‾‾‾    /  \
                                                              Neg°D   ...
                                                              [NegD]
```
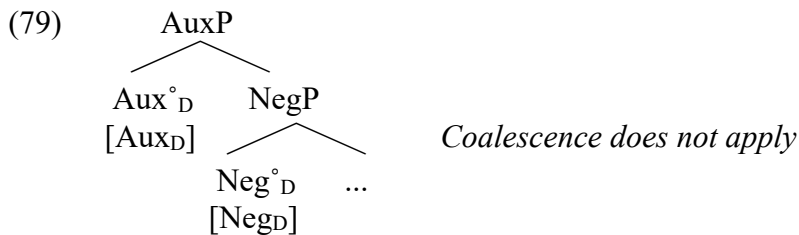
Some minor additions to the theory are needed in light of the standard analysis of English clause structure in which auxiliary projections are first Merged below NegP, and the highest auxiliary moves to T. This pre-movement clause structure is shown in (81).

(81)

```
        TP
       /  \
     T°    NegP
          /   \
        Neg°   AuxP
              /   \
           Aux°    …
```
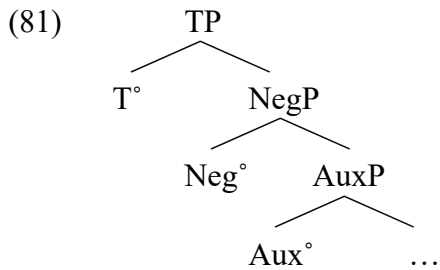
Within this structure, Aux-to-T movement followed by Coalescence is generated if both T° and Neg° are recessive, and dominant Aux°$_D$ moves to TP. However, in derivations that contain a dominant Neg°$_D$ (as argued for *not*), we incorrectly predict that Neg°$_D$, the closest dominant head, will move to Spec, TP, leaving the auxiliary below negation.[xxiii]

I propose a tentative solution appealing to Category Percolation. Suppose that the dominant and recessive Neg° heads differ in their featural content, such that the recessive head contains a subset of the features of T, but the dominant head does not. In other

words, Neg°$_D$ is not part of the verbal extended projection, while Neg°$_R$ is (a similar parameter is proposed in Roberts' 2018 account of cross-linguistic variation in the realization of clausal negation). To illustrate with one possible feature set, recessive negation contains category subfeatures {Aux, V}, a subset of those of T: {T, Aux, V} (82). All three heads can thus undergo Coalescence after movement of Aux$_D$. The dominant negation head contains a feature that is not found within the verbal extended projection. This prevents both Last Resort movement of Neg$_D$ to TP and Coalescence with T$_R$, as it would violate the No Tampering Condition. The only option is to move Aux$_D$. I label this antagonistic feature as {-V} in (83), but leave open whether category features have privative or binary values.

(82)
```
              TP
          ┌────┴────┐
   T_R {T, Aux, V}  NegP
                ┌────┴────┐
         Neg_R {Aux, V}  AuxP
                      ┌────┴────┐
               Aux_D {Aux, V}   …
```

(83)
```
              TP
          ┌────┴────┐
   T_R {T, Aux, V}  NegP
                ┌────┴────┐
         Neg_D {Aux, -V}  AuxP
                      ┌────┴────┐
               Aux_D {Aux, V}   …
```
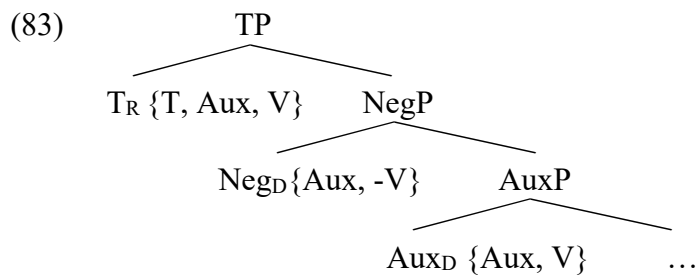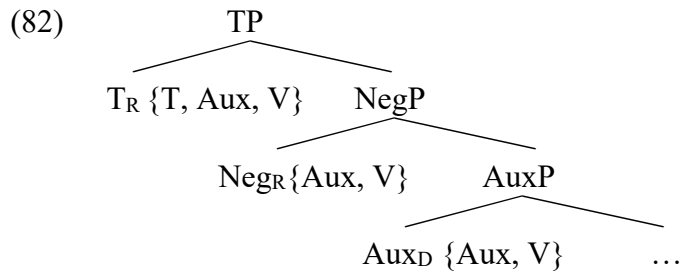
Finally, it is worth noting that negative contraction differs in key ways from auxiliary reduction (Zwicky 1970; Anderson 2008), which affects English auxiliary forms *is, has, would, had, have, am, are,* and *will*. First, reduced auxiliaries do not form syntactic units

with their hosts in the manner of contracted negation. For instance, reduced auxiliaries do not pattern as if they raise with elements that they are affixed to.

(84) 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 not 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*.[xxiv] On the basis of these differences, Zwicky & Pullum (1983) categorize reduced auxiliaries as *simple clitics*, phonologically reduced variants that occur in 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 & 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 syntactic derivation, and its restriction to a tensed auxiliary host results from the head-adjacent configuration of $Neg_R{}°$ and $Aux/T_D{}°$. 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 recession 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 contracted, 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 a recessive head is

eliminated by external Merge or movement: the realization of the past perfect in Catalan.

As described by Oltra-Massuet (2013), some dialects of Catalan express the past perfect

either in a synthetic form where the subject and tense and aspect morphemes are realized

as suffixes to a lexical verb (85), or an analytic form where subject agreement is realized

on an auxiliary verb *anar* 'to go' followed by a lexical verb participle that resembles the

infinitive form (86).

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

(86)   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, "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." It is further clarified that both

within and across speakers, variation depends on the lexical items and conjugations used;

speakers do not probabilistically use both forms for any given verb and conjugation pair.

    Oltra-Massuet proposes that the synthetic past perfect structure is derived by head-

movement of the verb through Aspect° and Tense° heads. Abstracting away from the

distinction between V and little *v,* not relevant to the present analysis, she proposes the following: In synthetic forms like *purificares*, V° moves to a T° head that carries specifications for past tense, perfective aspect, and telicity.
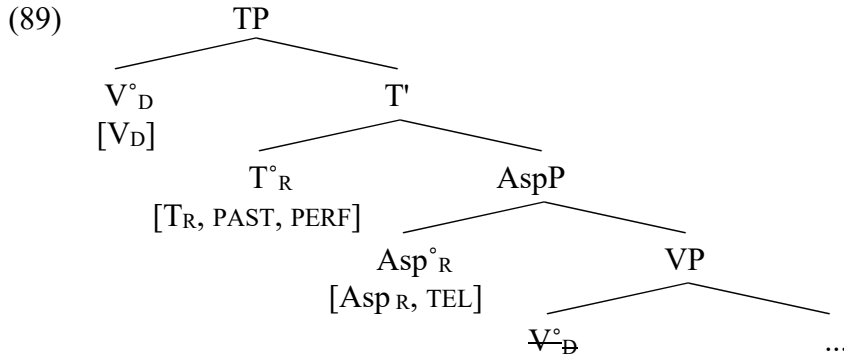
(87)

```
              TP
            /    \
          T°      ...
        /    \
      V°      T°
           [PAST, PERF, TELIC]
```
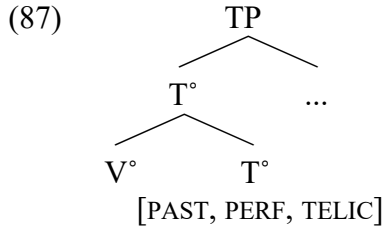
In the analytic form, [+PAST, +PERF] are associated with T°, but [+TELIC] is on a separate Aspect° head. Verb movement stops at Asp°, and V-Asp° is pronounced as the participle.

(88)

```
              TP
            /    \
          T°      AspP
     [PAST, PERF]  /  \
                Asp°   VP
               /   \    ...
             V°     Asp°
                   [TELIC]
```
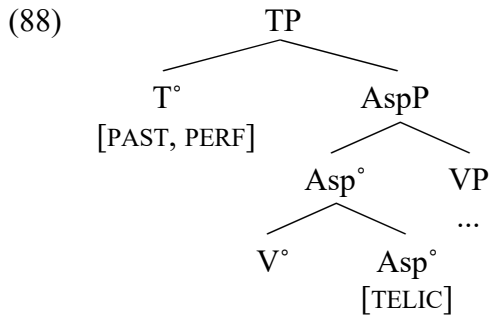
Because T° is obligatorily a suffix, this structure triggers *go*-support, the insertion of an *anar* auxiliary verb that supports the inflection of both tense and subjecthood.

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 Coalescense bundles V°$_D$ with T°$_R$ and Asp°$_R$.[xxv]
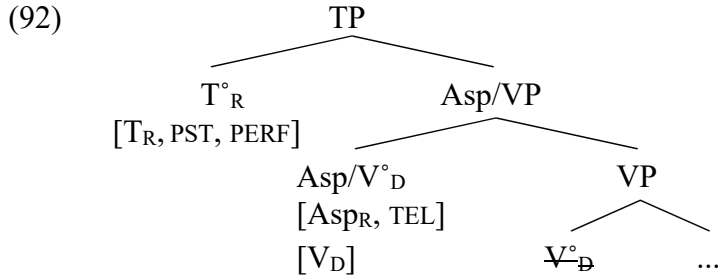
(89)

```
                    TP
          ┌──────────┴──────────┐
        V°_D                    T'
        [V_D]          ┌─────────┴─────────┐
                     T°_R                  AspP
                [T_R, PAST, PERF]    ┌───────┴───────┐
                                  Asp°_R            VP
                                [Asp_R, TEL]   ┌─────┴─────┐
                                             V̶°̶_̶D̶          ...
```

(90)

```
                T/Asp/VP
          ┌────────┴────────┐
      T/Asp/V°_D            VP
    [T_R, PAST, PERF]  ┌─────┴─────┐
    [Asp_R, TEL]     V̶°̶_̶D̶          ...
    [V_D]
```
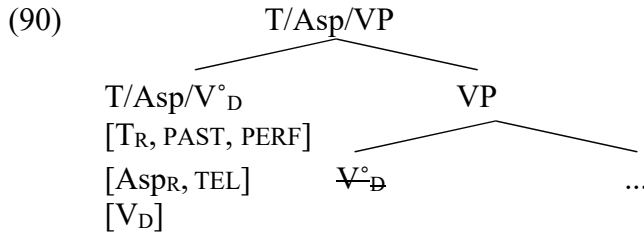
Suppose that in these dialects of Catalan, it is also possible for verb movement to proceed only as far as Spec, AspP. At this point, Coalescence bundles V°_D and Asp°_R.

(91)

```
            AspP                                        Asp/VP
      ┌──────┴──────┐                             ┌───────┴───────┐
     V°_D         AspP'                         V/Asp°_D          VP
     [V_D]    ┌─────┴─────┐         →          [Asp_R, TEL]  ┌─────┴─────┐
            Asp°_R        VP     Coalescence   [V_D]        V̶°̶_̶D̶         ...
          [Asp_R, TEL] ┌───┴───┐
                      V̶°̶_̶D̶      ...
                      [V_D]
```
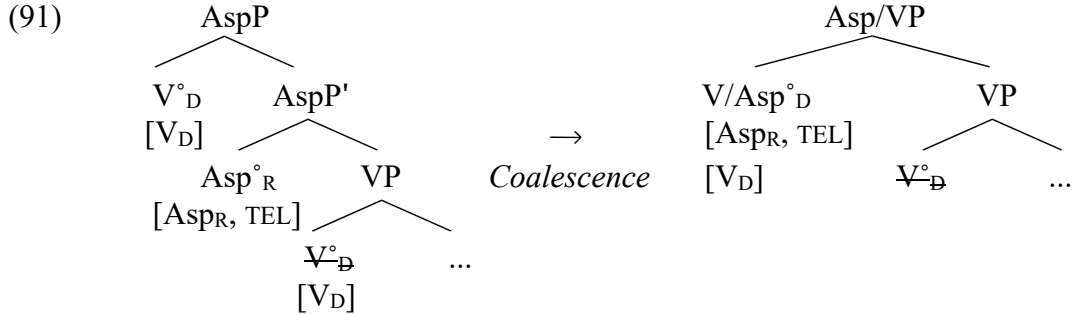
Note that even though V°_D has the [EPP] property, subject movement to Asp/VP will not take place if [Asp_R] lacks a subject agreement probe. Indeed, there is no agreement morphology on the participle to suggest such a probe associated with [Asp_R], and *auxiliary-subject-infinitive participle* orders are ungrammatical.

In the next step, the T°$_R$ head is Merged (92), followed by a dominant auxiliary head corresponding to *anar* (93). I will remain agnostic as to the categorial status of the auxiliary, labeling it simply as [Aux]. Coalescence then bundles Aux°$_D$ and T°$_R$ (94).

(92)
```
                        TP
             ┌──────────┴──────────┐
           T°R                   Asp/VP
      [TR, PST, PERF]       ┌──────┴──────┐
                         Asp/V°D          VP
                        [AspR, TEL]    ┌───┴───┐
                          [VD]        V°D̶      ...
```

(93)
```
                  AuxP
          ┌────────┴────────┐
        Aux°D               TP
       [AuxD]         ┌──────┴──────┐
                    T°R           Asp/VP
               [TR, PST, PERF]  ┌────┴────┐
                             Asp/V°D      VP
                            [AspR, TEL] ┌──┴──┐
                               [VD]    V°D̶    ...
```

(94)
```
                     Aux/TP
           ┌───────────┴───────────┐
        Aux/T°D                  Asp/VP
        [AuxD]              ┌──────┴──────┐
   [TR, PST, PERF]  Asp/V°D              VP
                   [AspR, TEL]        ┌───┴───┐
                     [VD]            V°D̶      ...
```
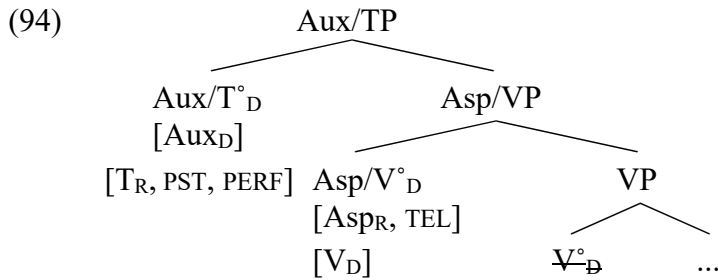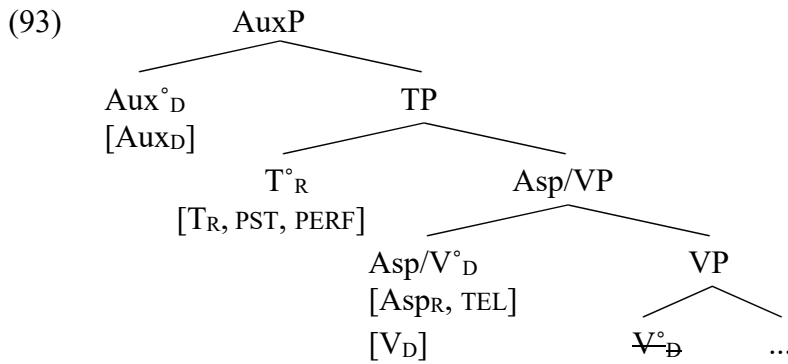
The analysis raises the question of how the derivation "knows" which option is used after T°$_R$ is Merged, given that the choice is uniquely determined by the verbal predicate. As a tentative solution, I propose that individual predicates can potentially introduce more than one corresponding object 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 V/Asp°$_D$.[xxvi]

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 should be understood as a well-formedness requirement on syntactic structures that is evaluated when they are transferred to PF. This makes clear predictions about when the Dominance Condition should apply, in the context of phase theory (Chomsky 2000). Under the assumption 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.[xxvii] We thus expect TP (or the highest head within a series of inflectional heads) to obligatorily have a dominant head. This

prediction is borne out in the English and Catalan case studies – 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 (95). In this structure, 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 [~~*u*Topic~~] probe to trigger phrasal movement of the topic *kaffe*.

(95)   [CP Kaffe drikker   [TP Peter   ofte   [VP om   morgonen]]
        Coffee drinks   Peter   often   in   morning
        'Coffee, Peter often drinks in the morning'

(96)

```
                    CP
           _____|_____
          DP                C'
        kaffe        _____|_____
                 C/T/V°D                TP
               [CR, uTopic]         ____|____
               [TR, uD]            DP         T'
               [VD, EPP]         Peter     ___|___
                 drikker               T/V°D       …
                                          _____|_____
                                        AdvP               VP
                                        ofte           _____|_____
                                                      DP          V'
                                                             _____|_____
                                                           V°D           DP
```
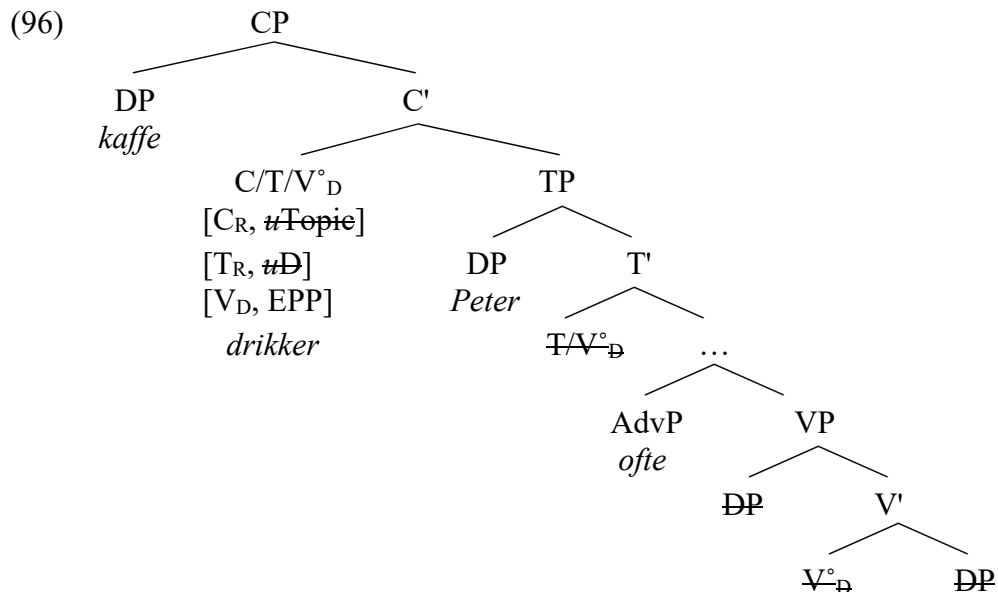
However, consider an alternative derivation in which head movement of $V°_D$ takes place after $T°_R$ and $C°_R$ have been Merged, and Coalescence creates a single head with

features of T°$_R$ and C°$_R$. Assuming that [EPP] can trigger phrasal movement in concert

with either the [~~uD~~] or [~~uTopic~~] probes, we predict the generation of structures like (97),

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.

(97)
```
                          CP
                 ┌─────────┴─────────┐
               DP                    C'
             kaffe          ┌─────────┴─────────┐
                      C/T/V°D                   …P
                      [CR, uTopic]        ┌──────┴──────┐
                      [TR, uD]          AdvP            VP
                      [VD, EPP]         ofte      ┌──────┴──────┐
                        drikker                  DP             V'
                                               Peter      ┌─────┴─────┐
                                                         V°D         DP
```
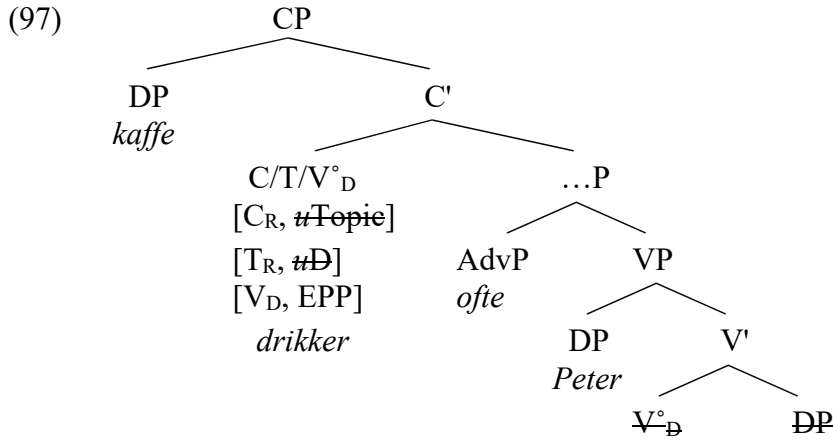
This derivation can be ruled out as follows. Suppose that spell-out applies to the

complement of the head that contains [C] when it ceases to project. There is then no

prohibition against the creation of a bundled C/T/V°$_D$ 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 results from a separate restriction that requires [~~uD~~] to

take precedence over [~~uTopic~~] in associating with [EPP] (recall that the Kashmiri V3

analysis similarly requires relative priority of probes in their association with [EPP]).

This ensures that topicalization is only possible if C and T are realized as separate

projections, each the target of verb movement, as is the case in Danish. In some ways,

this recasts 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 order. It allows for a simple mapping from syntactic structure to its morphological realization: the assocation 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 (M. Baker 1985).

As an illustration of a standard analysis of head movement, cyclic head movement of V° through the Aspect° and Tense° inflectional heads produces the complex head (98), with the projecting head at each level on the right (Williams 1981).

(98)
```
              T°
            /    \
        Asp°      T°
       /    \
     V°      Asp°
```
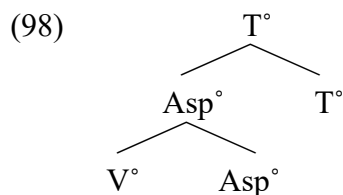
If vocabulary items are inserted in this structure, with no modification, both aspect and tense are realized morphologically as verbal suffixes. Prefixes can be generated by positing a postsyntactic operation like Local Dislocation (Embick & Noyer 2001), which alters the linear ordering of sister nodes prior to vocabulary insertion.

(99)
```
        Asp°
       /    \              Local Dislocation      [Asp - V]
     V°      Asp°                 →
```

If Local Dislocation applies between sister nodes of complex heads from the bottom up, a V/Asp/T° head can ordered as [[V-Asp]-T], [[Asp-V]-T], [T-[V-Asp]], [T-[Asp-V]]. However, it is not possible to generate orders where the Mirror Principle is violated among affixes on one side of the root (Harley 2013), in particular *[Asp-T-V], *[V-T-Asp].

The Mirror Principle is not exceptionless; languages can permit affix orders that contradict scopal or derivational relations. For example, derivational suffixes in Bantu languages largely occur in a fixed 'Causative-Applicative-Reciprocal-Passive' order (Hyman 2003; Good 2003). In Chichewa, causativized applicatives and applicativized causatives occur with the order V-CAUS-APPL, creating a mismatch in scope and affix order for causativized applicatives.

(100)   a.   lil-its-il   b.   takas-its-il
             cry-CAUS-APPL   stir.with-CAUS-APPL
             'cause to cry with'   'cause to stir with'
             [[cause cry] with]   [[stir with] cause]   (Hyman 2003)

The Mirror Principle can also be violated in favor of more strictly phonological restrictions, such as relative sonority (Arnott 1970; Paster 2005, 2006) or prosodic properties of stems (McCarthy and Prince 1993; Ussishkin 2007). Nonetheless, even if affix ordering is determined by competing functional preferences (cf. Manova and Aronoff 2010, Rice 2011 for an overview), to the extent that it predicts a fundamental preference on affix ordering, head-internal branching in syntactic representations provides explanatory power.

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 containing multiple
category features. Alternatively put, the order of affixes reflects the order in which their
corresponding heads entered the derivation via external Merge. This is possible if c-
selection features are not deleted from the syntactic representation upon checking, and
accessible to affix linearization operations. This conclusion is consistent with the analysis
of delayed EPP patterns in Section 4, which require checked agreement probes to remain
accessible to later stages of the derivation. I illustrate briefly with the simplifed clause
structure [$_{TP}$ [$_{AspP}$ [$_{VP}$ …]]]. External Merge of each functional head is triggered by c-
selectional uninterpretable features [uV] on Asp˚ and [uAsp] on T˚. Subsequent head
movement of V˚ and Coalescence then create the bundled head in (102).

(101)

                            TP

          T˚$_R$ AspP
     [T$_R$, ~~uAsp~~]
               Asp˚$_R$ VP
             [Asp$_R$, ~~uV~~]
                   V˚$_D$ …
                   [V$_D$]

(102)

            T/Asp/VP

   T/Asp/V˚$_D$ …
   [T$_R$, ~~uAsp~~]
   [Asp$_R$, ~~uV~~]
   [V$_D$]
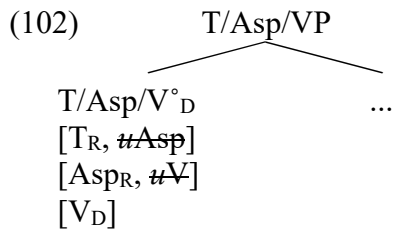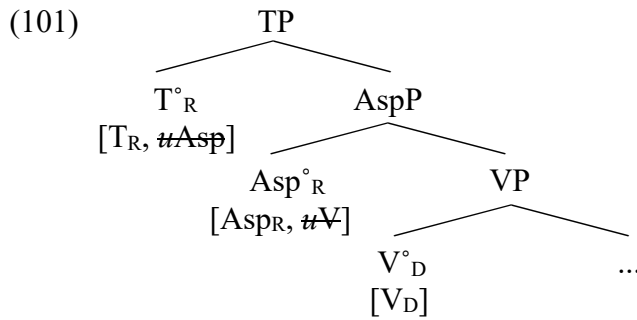
The Mirror Principle can thus be stated as a preference for linearization to reflect c-selection relations among features of a head, such that the exponent of a feature must be equidistant or closer to the word edge than the exponent of a feature that it c-selects.

The key claim for the current purposes is that the absence of branching structure does not preclude an account of the Mirror Principle generalization on affix order. There are reasonable concerns with the approach, however. ~~While~~ there is an independent conceptual necessity for linearization procedures to be sensitive to branching structure. However, it is less clear why c-selection relations should be accessible to affix linearization, though it is plausibly driven by a preference to make scope relations transparent. In addition, my proposal does not adhere as strictly to the desideratum of informational encapsulation in modular systems, in that c-selection features are visible to two components 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 structure from syntactic representations.

Finally, it is worth noting that the case studies in this paper have focused on movement and bundling patterns observed within the complementizer and inflectional domains of the clause. Important questions remain in determining whether the approach can be extended to syntactic approaches to derivational morphology, in which acategorial roots must combine with category-defining functional heads (Marantz 2007; Embick and Marantz 2008). In particular, consider structures that contain more than one category-defining head like *gloriousness* [*n* [*a* [√GLORY]]] (Embick & Marantz 2008). Coalescence might not be expected to apply to such structures, as they contain apparently

incompatible features [*n*], [*a*]. Nonetheless, *gloriousness* patterns as a single noun head with respect to syntactic operations. Detailed consideration of this issue will have to await future work.[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 for counterarguments). While this adds a certain level of 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. In *do*-support contexts (sentential negation, questions, ellipsis), an auxiliary *do* appears in T with tense suffixes.

(103) [$_{TP}$ Lucille   [often [$_{VP}$ sees   George   ]
(104) [$_{TP}$ 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 the structures in (103) and (104) both contain a

dominant auxiliary (labeled $Aux_{do}°$) *as* the head of TP (cf. Emonds 1970; Pollock 1989; Wilder and Ćavar 1994a for similar instantiations of a null Aux or null *do* proposal). I posit that like other auxiliaries, $Aux_{do}°$ first head-moves and undergoes Coalescence with $T°_R$ (105).

(105)

```
           T/Aux_do P
          /        \
        DP          T/Aux_do'
                   /       \
          T/Aux_do°_D       VP
          [Aux_D, EPP]     /  \
          [T_R, u̶P̶h̶i̶]   D̶P̶     V'
                              /  \
                           V°_D   ...
```

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}:

(106)   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 the proposal of Embick & Noyer (2001), *do* is the realization of a head that is (syntactically) Merged onto T if T lacks a VP complement and no auxiliary moves to T. However, given the guiding assumption in Minimalism that Merge is driven by features, this triggering mechanism is unorthodox and difficult to implement. 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 all 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 of English

where *do*-support occurs with auxiliaries in ellipsis, e.g. *George will have escaped to*

Mexico, and Buster will have done too. The placement of *do* below auxiliaries in these

cases is not predicted in proposals in which *do* is inserted in T.[xxix]

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. In this paper, 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 projections (as opposed to terminal nodes, as

commonly assumed in Distributed Morphology), also known as *spans*.[xxx] Word order

variation within extended projections arises through two means: languages vary in 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 theory is presented as a means of elimiting head movement altogether 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, one could propose that semantic features can restrict possible placement of @. Second, it is difficult to account for patterns in which the placement of specifiers is dependent on the position of heads, as observed in delayed gratification patterns. This is because @ determines where heads are linearized relative to specifiers in span, which is expected to be an independent parameter of whether individual projections can have specifiers. Delayed gratification patterns could be accounted for in Spanning with an ad hoc restriction that some projections can only have @ if they can host a specifier. In contrast, they are generated in syntax by my 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 Coalescence 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. In

particular, 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 primary 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. Ph.D dissertation, University of Connecticut.

Aboh, Enoch Oladé. 2006. Complementation in Saramaccan and Gungbe: the case of C-type modal particles. *Natural Language & Linguistic Theory* 24: 1–55. doi:10.1007/s.

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. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16: 373–415.

Baker, Mark. 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. 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. 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. 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. Ph.D 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. Robert Borsley and Ian G. Roberts, 53–74. Cambridge: Cambridge University Press.

Bošković, Željko. 1995. Participle movement and second position cliticization in Serbo-Croatian. *Lingua* 96: 245–266.

Brody, Michael. 2000. Word order, restructuring and Mirror Theory. In *Derivation of VO and OV*, ed. Peter Svenonius, 27–43. Amsterdam/Philadelphia: John Benjamins.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20*, ed. 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. 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. 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. Bernd Heine and Heiko Narrog, 1–21. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199544004.013.0003.

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: a journal of general linguistics* 3: 65. 1–43.

Delsing, Lars-Olof. 1993. The internal structure of noun phrases in the Scandinavian languages: a comparative study. Ph.D dissertation, Lunds University.

Diesing, Molly. 2005. The upper functional domain in Yiddish. In *Focus on Germanic*

typology, ed. 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. doi:10.1515/TL.2007.001.

Douglas, Jamie. 2017. Unifying the that-trace and anti-that-trace effects. *Glossa: a journal of general linguistics* 2: 60. 1–28.

Embick, David, and Roumyana Izvorski. 1995. On long head movement in Bulgarian. In *ESCOL '94*, ed. Janet M. Fuller, Ho Han, and David Parkinson, 104–115. Ithaca, NY: CLC Publications.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39: 1–53. doi:10.1162/ling.2008.39.1.1.

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 E. 1970. Root and structure-preserving transformations. Massachusetts Institute of Technology.

Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language & 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. Artemis Alexiadou, Jorge Hankamer, Thomas McFadden, Justin Nuger, 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 & Linguistic Theory* 29: 169–209. doi:10.1007/s11049-010-9109-x.

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.

Giorgi, Alessandra, and Fabio Pianesi. 1997. *Tense and aspect: from semantics to morphosyntax*. New York: 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. 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. 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.

Haegeman, Liliane. 1995. *The syntax of negation*. Cambridge: Cambridge University Press.

Haider, Hubert. 1988. Matching projections. In *Constituent structure: papers from the 1987 GLOW Conference*, ed. Anna Cardinaletti, Guglielmo Cinque, and Giuliana Giusti, 101–121. Dordrecht: Foris.

Harizanov, Boris. 2014. Clitic doubling at the syntax-morphophonology interface. *Natural Language & Linguistic Theory* 32: 1033–1088.

Harizanov, Boris, and Vera Gribanova. 2018. Whither head movement ? *Natural Language & Linguistic Theory*. Springer Nature B.V. doi:10.1007/s11049-018-9420-5.

Harley, Heidi. 2013. Getting morphemes in order: Merger, Affixation and Head Movement. In *Diagnosing syntax*, ed. 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. doi:10.1162/LING_a_00050.

Höhn, Georg F.K. 2016. Unagreement is an illusion: Apparent person mismatches and nominal structure. *Natural Language and Linguistic Theory* 34. Springer Science+Business Media Dordrecht: 543–592. doi:10.1007/s11049-015-9311-y.

Hsu, Brian. 2017. Verb second and its deviations: An argument for Feature Scattering in the left periphery. *Glossa: a journal of general linguistics* 2: 35. 1–3. doi:10.1016/0025-326X(78)90261-8.

Hsu, Brian, and Saurov Syed. *to appear*. Variation in the co-occurrence of Indexical elements : evidence for split indexical projections in DPs. To appear in *Proceedings of the 36th West Coast Conference on Formal Linguistics*.

Hyman, Larry M. 2003. Suffix ordering in Bantu: a morphocentric approach. In *Yearbook of Morphology 2002*, ed. Geert Booij and Jaap van Marle, 245–281. Dordrecht: Springer. doi:10.1017/CBO9781107415324.004.

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 & Linguistic Theory* 27: 305–344.

Kayne, Richard. 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. Guglielmo Cinque and Richard S. Kayne, 3–69. New York: Oxford University Press. doi:10.1093/oxfordhb/9780195136517.013.0001.

Kayne, Richard S. 2005b. *Movement and silence*. Oxford: Oxford University Press.

Keine, Stefan. 2019. Selective opacity. *Linguistic Inquiry* 50: 13–62. doi:10.1162/ling.

Koopman, Hilda. 1996. The Spec Head configuration. In *Syntax at Sunset, UCLA Working papers in syntax and semantics Volume 1*, ed. Edward Garret 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. Gautam Sengupta, Shruti Sircar, Gayatri Raman, and Rahul Balusu, 23–47. Dordrecht: Springer. doi:10.1007/978-981-10-4295-9.

Krapova, Iliyana, and Tsena Karastaneva. 2002. On the structure of the CP field in Bulgarian. *Balkanistica* 15: 293–321.

Lechner, Winfried. 2006. An interpretive effect of head movement. In *Phases of Interpretation*, ed. Mara Frascarelli, 45–71. Berlin: Mouton de Gruyter.

Ledgeway, Adam. 2010. Subject licensing in CP. In *Mapping the Left Periphery*, ed. Paola Benincà and Nicola Munaro, 257–296. Oxford: Oxford University Press.

Lee, Hyunjung, and Gereon Müller. 2018. Asymmetries in long-distance dependencies: a view from Gradient Harmonic Grammar. Paper presented at the *Workshop on Long-Distance Dependencies, Humboldt-Universität zu Berlin, October 4-5.*

Lema, José, and María Luisa Rivero. 1989. Long head movement: ECP vs. HMC. In *Proceedings of NELS 20*, 333–347. Amherst, MA: GLSA.

Li, Yafei. 1990. X^0-binding and verb incorporation. *Linguistic Inquiry* 21: 399–426.

Lundquist, Björn. 2008. Nominalizations and participles in Swedish. Ph.D dissertation, University of Tromsø.

Mahajan, Anoop Kumar. 2001. *Word order and remnant VP movement*. In *Word order and scrambling (Explaining Linguistics 4)*, ed. 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. doi:10.1007/s11525-010-9153-6.

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. S. H. Choe, 191–222. Seoul: Dong-In Publishing Co.

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37: 69–109. doi:10.1162/002438906775321184.

Matyiku, Sabina Maria. 2017. Semantic effects of head movement : Evidence from negative auxiliary inversion. Ph.D dissertation, Yale University.

McCarthy, John J, and Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology*: 79–153.

Migdalski, Krzysztof Marek. 2006. The syntax of compound tenses in Slavic. Tilburg University.

Mohr, Sabine. 2009. V2 as a single-edge phenomenon. In *Selected Papers from the 2006 Cyprus syntaxfest*, ed. Kleanthes K. Grohmann and Phoevos Panagiotidis, 141–159. Cambridge: Cambridge Scholars Publishing.

Munshi, Sadaf, and Rajesh Bhatt. 2009. Two locations for negation: Evidence from Kashmiri. *Linguistic Variation Yearbook* 9: 205-205–239. doi:10.1075/livy.9.06mun.

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. 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. 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.

Paster, Mary Elizabeth. 2006. Phonological conditions on affixation. Ph.D dissertation, University of California, Berkeley.

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. 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. 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. Massachusetts Institute of Technology. doi:10.7551/mitpress/9780262162548.001.0001.

Rice, Keren. 2011. Principles of affix ordering: An overview. *Word Structure* 4: 169–200. doi:10.3366/word.2011.0009.

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. 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.

Smolensky, Paul, and Matthew Goldrick. 2016. Gradient symbolic representations in grammar: the case of French Liaison. *ROA 1286*.

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. doi:10.7557/12.213.

Steriade, Donca. 1995. Underspecification and markedness. In *The Handbook of Phonological Theory*, ed. 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*.

Svenonius, Peter. 2016. Spans and words. In *Morphological metatheory*, ed. Heidi Harley and Daniel Siddiqi, 199–220. Amsterdam: John Benjamins.

Thráinsson, Höskuldur. 1996. On the (non)-universality of functional projections. In *Minimal ideas*, ed. Werner Abraham, Samuel David Epstein, Hoskuldur Thrainsson, and C. Jan-Wouter Zwart, 253–81. Amsterdam: John Benjamins.

Travis, Lisa. 1984. Parameters and effects of word order variation. Ph.D dissertation, Massachusetts Institute of Technology.

Ussishkin, Adam. 2007. Morpheme position. In *The Cambridge handbook of phonology*, ed. Paul de Lacy, 457–472. Cambridge: Cambridge University Press.

Vangsnes, Øystein Alexander. 1999. 'Identification' and the role of morphology in the Scandinavian noun phrase: 99–176.

Vikner, Sten. 1994. Finite verb movement in Scandinavian embedded clauses. In *Verb

movement, ed. Norbert Hornstein and David Lightfoot, 117–147. Cambridge: Cambridge University Press.

Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. New York and Oxford: Oxford University Press.

Walkden, George. 2017. Language contact and V3 in Germanic varieties new and old. *Journal of Comparative Germanic Linguistics* 20: 49–81. doi:10.1007/s10828-017-9084-2.

Walker, Rachel. 2011. *Vowel patterns in language*. New York: Cambridge University Press.

Wilder, Chris, and Damir Ćavar. 1994a. Long head movement? verb movement and cliticization in Croatian. *Lingua* 93: 1–58.

Wilder, Chris, and Damir Ćavar. 1994b. Word order variation, verb movement, and economy principles. *Studia Linguistica* 48: 46–86. doi:10.1111/j.1467-9582.1994.tb00849.x.

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. doi:10.1371/journal.pone.0005772.

---

[i] This asymmetry is not found in the extraction of arguments from embedded clauses, which appears insensitive to the presence of a complementizer. Giorgi & Pianesi (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 & Pianesi 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. A tentative explanation for cases where bundling "suppresses" a probe is discussed in Section 4.1, in context of a similar pattern in Kashmiri involving a left-peripheral [Topic] probe.

[iii] In addition to pre-syntactic bundling, "Scattering B," Giorgi & Pianesi hypothesize the existence of "Scattering A," an 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 & Pianesi (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 assumption: I claim that while only dominant heads can license specifiers, not all dominant heads have this ability.

[v] The latter issue does not arise within Kayne's (1994) revised definition of c-command that distinguishes between segments and categories.

[vi] In Den Dikken's proposal, the probe in (21a) is unable to agree with the object because vP is a phase. Head movement of v+V to F extends the phase upwards, such 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 interpretation, that agreement between F and the object is always possible, but F must inherit its ability to attract a specifier from head-moved v+F.

[vii] This structural characterization of movement resembles the one posited in reprojection approaches to 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 I show in Section 3.2, reprojection is obviated if Category Percolation (Keine 2019) is adopted.

[viii] The observation here resembles the claim in other works 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 carry the same features), but in how they are mapped to phrase structure.

[ix] This structural definition, which precludes Coalescence from applying to two heads with an intervening specifier, often resembles that of of linear adjacency. The two structural definitions would differ in a context in which a head-initial projection dominated a head-final one with no specifier, [ZP Z [YP [XP …]

Y], satisfying structural but not linear adjacency. While I know of no instances of bundling that takes place in this configuration, this could be explained if true head-final structures are excluded from syntactic representations (R. Kayne 1994).

[x] The output of Coalescence is identical to that of Matching Projections in the theory of 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 phonetically realized head.

(i)   X/ZP
           ╱   ╲
   Spec,X/Z   X/Z'
               ╱   ╲
           X/Z   ...
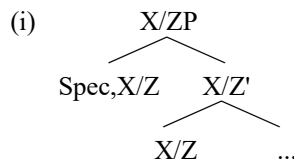
[xi] I thank an anonymous reviewer for pointing out the relevance of such examples.

[xii] I cannot yet provide a deeper explanation for why bundling seems to require the dominant head to c-command the recessive head, but not vice versa. Potentially, it may follow from restrictions on the kinds of elements that can trigger syntactic operations, or on the search space of operations. However, a fuller explanation must await future work.

[xiii] In the formulation of Chomsky (2008), 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.

[xiv] Tentatively, we may explore the idea that the movement of dominant heads is *always* anti-local, crossing at least two projections prior to Coalescence.[xiv] It may be the case, given the many empirical arguments made in the Cartographic Program for a highly articulated hierarchy of features, that the apparently local movement patterns used to justify the Head Movement Constraint in fact require movement across intervening heads. However, this discussion necessarily remains speculative, given the current variety 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.

[xv] It could turn out that either of these conditions is 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 comprehensive evaluation of these conditions and where they (don't) apply must await another occasion.

[xvi] The approach resembles the system of Chomsky (2001) in the sense that the [EPP] feature is distinct from [uF] 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.

[xvii] This requires a slight modification in a theory where roots obtain their 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 v, rather than individual roots.

[xviii] For arguments that a low C-domain feature like Finiteness is associated with subject probes, see Poletto (2000) Aboh (2006), Ledgeway (2010).

[xix] This does not rule out the possibility for languages to have complementizers with [EPP]. For example, this is necessary to account for languages in which topics or foci move to a position immediately preceding a complementizer, such as Bulgarian (Krapova and Karastaneva 2002).

(ii) *Bulgarian* (Krapova and Karastaneva 2002)
Ivan   misli   [kniga-ta   če   Marija ja   e   kupila]
Ivan   think-3SG   book-the   that   Mary   it.CL   has   bought
'Ivan thinks that the book, Mary has bought.'

[xx] In a similar vein, Leu (2015) argues 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 an "-*ass*" head that attracts a lower complementizer head corresponding to "*d*-", or filled by movement of a remnant vP (Müller 2004).

[xxi] Similar asymmetries in the priority given to probes are discussed by Lee and Müller (2018), who assume a constraint-based grammar in which symbols have gradient activity levels (Smolensky and Goldrick 2016). In this approach, individual probes vary in their activity levels (i.e. [uTop$_{0.5}$] vs. [uContrast$_{0.8}$]); The penalty incurred by a failure to trigger movement is proportional to the activity of the probe. On this view,

variation between the two dialects of Kashmiri arises from a difference in the relative activity levels assigned to [uTop] and [uContrast].

[xxii] For Matushansky (2006), sentential negation is Merged as a specifier of AuxP that is simultaneously 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] This issue remains in an analysis where *n't* is a Neg head, while *not* is a specifier of NegP (Haegeman 1995). Under the assumption that only dominant heads license specifiers, one still needs a way to exclude movement of (null) Neg°$_D$, instead of Aux°$_D$, to TP.

[xxiv] The reduced forms of *will*, *have*, *are*, and *am* are more restricted, appearing only after pronouns and *wh*-words.

[xxv] I do not exclude the possibility that V° moves first through AspP before raising to TP. However, this is not critical for the purposes of this discussion, since the two possible derivations will result in the same bundled head structure.

[xxvi] A reviewer suggests an alternative account in which all clauses contain an Aux, but individual verbs select whether they occur with Aux$_D$ *anar* or an Aux$_R$ null auxiliary that will undergo Coalescence. This has the benefit of maintaining a uniform set of functional heads in all clauses, and a more restrictive way of determining the content of the numeration. The proposal also comes with potentially significant implications. For instance, if auxiliary selection arises through Agree between Aux and V, it suggests that the dominance vs. recessiveness is not always an inherent property of heads when they are first Merged, but can be contextually determined by agreement. I will have to leave detailed consideration of this hypothesis and its predictions to future work.

[xxvii] 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 (Rizzi 1997). I refer the reader to Douglas (2017) for a recent approach.