COPiL CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **18**, Article 3: pp. 57–90, 2026 | ISSN: 2050-5949

Evaluating Labeling in Minimalist Syntax *

Dallas L. Linn
University of Cambridge

Abstract This paper concerns the topic of labeling in minimalist syntax, the process of deriving a feature or set of features which identifies a syntactic object. Labeling has emerged in large part due to the historical course of generative syntax, so I first examine this historical development, especially of the general minimalist framework, establishing definitions and highlighting important principles. A key theme emerges that in derivations applying set-theoretic Merge, labeling must encode asymmetry as hierarchy in order to enable LF and PF interpretation. I then survey three distinct groups of approaches: projection by selection, the Labeling Algorithm, and alternatives to Merge. For each group I consider the proposed architecture of the labeling process, potential applications to other syntactic phenomena, and whether asymmetry is consistently encoded. The survey provides critical insight into the minimum viability of different approaches, highlights consequential applications of labeling, and suggests key considerations for future inquiry.

1 Introduction

Labeling is a major topic in contemporary minimalist syntax, and yet discussion of its complexities is often somewhat opaque. Working to unpick this complexity and clarify what remains opaque offers insight into how fundamental aspects of the derivation can be approached from different angles. Conducting a comprehensive evaluation requires understanding how proposals are influenced by historical assumptions, intended applications, and theoretical aims. It also requires a fully contrastive approach which equally compares both mainstream theories and more unorthodox proposals.

In this paper, I present an initial survey to contribute to the lack of such a comprehensive evaluation in the literature. I will evaluate three major categories of approaches to labeling in the context of the historical development of the issue. These categories and the associated theories have been selected for their introduction of key ideas or notable theoretical innovations. In conducting this survey, I aim to contribute to the critical appraisal of both these individual theories and of labeling and its role in minimalist syntax more broadly.

To begin, consider a seemingly simple question: what is *labeling*? An action of labeling entails the existence of *labels*. In a natural sense, labels identify items. In a

* I would like to thank Theresa Biberauer, my supervisor, for providing indispensable guidance and feedback throughout the course of this project, and Ksenia Zanon, my director of studies, for her continued support.

©2026 Linn
This is an open-access article distributed by Section of Theoretical & Applied Linguistics, Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge under the terms of a Creative Commons Non-Commercial License (creativecommons.org/licenses/by-nc/3.0).

syntactic sense, labels have been used from the early days of generative theories: Phrase Structure Rules (PSRs) used labels to restrict how constituents could relate to one another (cf. Collins 2002, Adger 2003: ch. 2).

A derivation may begin as follows:

(1) a. S, apply S → NP VP

b. NP VP

The labels NP and VP identify further sets of PSRs, from which one must be applied to continue the derivation, e.g.:

(2) a. NP VP, apply NP → D NP

b. D NP VP

Labels also provide information to identify constituents during the application of transformational rules:

(3) Passivisation rule: $NP_1 - Aux - V - NP_2 \rightarrow NP_2 - Aux + be + en - V - by + NP_1$
(Chomsky 1957: ex. 24)

Once PSRs offer no further decomposition, the remaining labels identify the categories of lexical items (LIs) which may be inserted:

(4) N → car, hat, cat,...

Labels as proposed in PSRs to fulfill the natural sense of identifying constituents. Notably, this identification is a primitive feature of the labels themselves; all labels in PSRs are non-terminal elements, separate from the sets of rules and LIs to which they correspond. The labels derive strictly from the set of derivational rules. Further, the function of LIs in the derivation is derived from their insertion in a category already introduced.

This last insight seems unnatural, as critiqued by Lyons (1968): with some exceptions (notably S), each XP always decomposes to include the corresponding X and a matched LI. The more natural assumption, since it is broadly accepted in generative syntax, is that syntactic objects (SOs) are *endocentric*: the function of the whole derives from the function of one of the parts. To capture endocentricity, the category of SOs must therefore be derived from LIs, rather than from the context of their usage (arguably the method in PSRs).[1]

Endocentricity was formalized through the proposal of X'-theory (Chomsky 1970, Jackendoff 1977). Under X'-theory, each category which enters the derivation is applied to a generalized cross-categorial template which creates phrasal, intermediate, and head levels:[2]

[1] cf. Travis (2005).

[2] This is commonly referred to as projection, but I later establish a definition of the term which excludes X' labels.

(5)


This template ensures endocentricity for every SO in the derivation. Labels are still non-terminal in nature, with heads separate from their corresponding LIs. Chomsky and Jackendoff propose that LIs may not correspond to a single category but rather be underspecified for category features;[3] therefore it cannot be said that the function of the phrases in X'-theory fully derives from LIs. However, the selected category of the head (whether derived directly from the LI or in some other manner) is applied to the X' template, and as a result determines the label identifying the overall SO.

Endocentricity has largely been maintained in contemporary theories, with the notable exception of part of the Labeling Algorithm (Chomsky 2013; see subsection 4.2).[4] As a result of this shift, the role of labels is quite different from their original function in PSRs. Labels no longer restrict a set of constituents or rules which must apply, but do retain the function of identifying SOs, derived from the identity of a constituent (the *head* of the SO).

Following from the above and using a feature-based approach, I will take the following to define labels:

(6) *Label*: a feature or set of features that classifies a constituent.

In minimalist syntax, this role of identification is also linked to the establishment of hierarchy, especially in SOs generated by symmetric Merge. Labels are no longer provided by default in the same manner that the template in X'-theory provides; instead, they must be systematically created by the derivation as part of the structure-building process. As a result, I will define labeling as follows:

(7) *Labeling*: the process of deriving a *label* for a complex SO.

where:

(8) *Complex SO*: an SO formed by Merge or another structure-building operation.[5]

The key to this definition is that *labeling* is a *derivational process*. Some theories of labeling (in a general sense) do not propose such a process. Whether proposed approaches fit the technical definition in (7) will be noted throughout this survey.

[3] Notably reminiscent of the later proposal for separate roots and functional categorizing heads in Distributed Morphology (Marantz 1997).

[4] And also arguably Adger (2025); see subsection 5.2.

[5] This definition is generalized to include approaches which propose alternatives to Merge (considered in section 5).

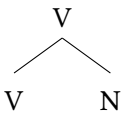
The remainder of this paper is structured as follows: in section 2, I present an overview of key issues in labeling, considering the principles, interactions, arguments, and early proposals which motivate the approaches evaluated. A survey of approaches follows in sections 3–5, grouped by type: projection by selection, the Labeling Algorithm, and alternatives to Merge. Finally, section 6 concludes the paper.

2 The development of labeling

2.1 Labels beyond labeling

Why is *labeling* necessary? Drawing from the definition established in (6), *labels* provide information about the *identity* of complex SOs. This information is used in both derivational processes and in the interpretation of SOs at the interfaces. In these interactions, identity is closely related to structural *hierarchy* (also referred as *asymmetry*). Many approaches to labeling completely conflate identity and hierarchy:

(9)



Regardless of the approach to labeling used to generate this structure, the label V provides both identification and asymmetric hierarchy: the SO is of the type V, and the constituent V is more prominent than N. Without a label, the set {V, N}, created by Merge, would lack asymmetric prominence.

Such a complete fusion of identity and hierarchy is not always characteristic of approaches to labeling. As a result, it is important to consider which function should be considered primitive (if any) and necessary as a feature of well-formed derivations. I suggest that the primitive notion is that of hierarchy, based on two arguments:

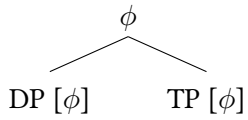
(10) Identity can always be derived from asymmetric hierarchy, but asymmetric hierarchy cannot always be derived from identity.

(11) Asymmetric hierarchy is a necessary condition of interpretation.

To begin, consider again (9). If this structure is taken to reflect a primitive asymmetric relationship between V and N, where V is more prominent than N, then the identity of the SO can be taken from V's prominence. The alternative is also true in this instance: if the structure is taken to represent that V is the identity of the SO, an asymmetric relationship can be derived from identity, since V is a constituent of the SO.

However, deriving an asymmetric hierarchy from identity is not always possible:

(12)



This example draws from Chomsky (2013), which proposes that shared features can label SOs. In this case, shared phi-features between DP and TP provide the label. Because the identity in the label is drawn equally from both DP and TP, it is impossible to derive an asymmetric relationship. This case will be discussed in more depth in section 4; for now, it demonstrates that asymmetric hierarchy cannot always be derived from identity.[6]

Continuing to (11), the statement is true of interpretation at both the LF and PF interfaces. At LF, hierarchy is necessary for compositional semantic interpretation of the derivations; this includes the identification of arguments and determination of the scope.[7] At PF, hierarchy is necessary for the linearization of the derivation under frameworks such as Kayne (1994) and Fukui & Takano (1998). Identity may also contribute to interpretation, but interpretation cannot occur without asymmetric hierarchy. Therefore, hierarchy must be a primitive characteristic of well-formed derivations, and interpretation can be precisely defined:

(13) *Interpretation*: the use of asymmetric hierarchical structure to map SOs to semantic and phonological representations.

In the context of labeling in minimalist syntax, the role of *labeling* is therefore to encode hierarchy. Hierarchy in turn can provide identification, but the primitive status of hierarchy leads to its consideration as the central function of *labeling*.

2.2 Principles of minimalist syntax

The proposals evaluated in this survey broadly fall under the purview of minimalist syntax. They vary widely, to the extent of significant departures from the mainstream Bare Phrase Structure (BPS) variant of minimalist syntax. What unites them is a shared set of theoretical principles following from the *Minimalist Program* (Chomsky 1995). These principles motivate the approaches taken by the proposals and, therefore, merit introduction and definition.

One key principle is that of minimalism as it relates to syntactic theory, most clearly stated as the Strong Minimalist Thesis (SMT):

(14) *SMT*: language is an optimal solution to legibility conditions.
(Chomsky 2001: p. 96)

[6] The discussion of asymmetry being derived from identity parallels *endocentricity*; however, the approach of Adger (2025) (see subsection 5.2) arguably retains hierarchy without endocentricity, so I consider the properties separate.

[7] The exact nature of LF is not a settled matter, but this generalization seems to reasonably hold as an assumption; cf. Sailer (2016) for discussion of relevant issues.

The narrow syntax must generate structures which can undergo *interpretation*; its complexity is assumed to directly link to the complexity of the context forming universal grammar. The SMT therefore compels that the proposed form of the narrow syntax only consist of that absolutely necessary to generate interpretable structures.[8]

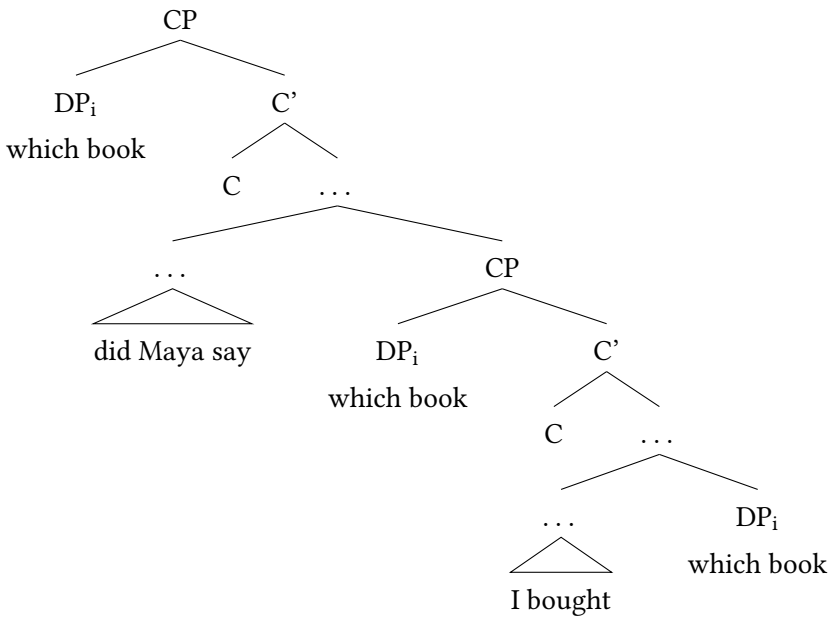
Appeals to the SMT closely interact with the principle of *economy*. The economic considerations are derived with reference to *third factors*, influences on the development of I-language that are not specific to the language (Chomsky 2005); in this case, the finite computational resources in the human brain.

While these are established principles in the field, they cannot be relied upon in isolation to support one proposal over another. To briefly illustrate, consider long-distance movement:

(15) [Which book]$_i$ did Maya say [I bought t_i]

Long-distance movement has been shown to have effects on intermediate constituents between the start and end positions of the movement (cf. Van Urk 2020). The commonly accepted view in minimalist syntax is that this phenomenon involves successive-cyclicity (e.g. Chomsky 1977, 2013), where movement must occur in multiple, shorter steps:

(16)

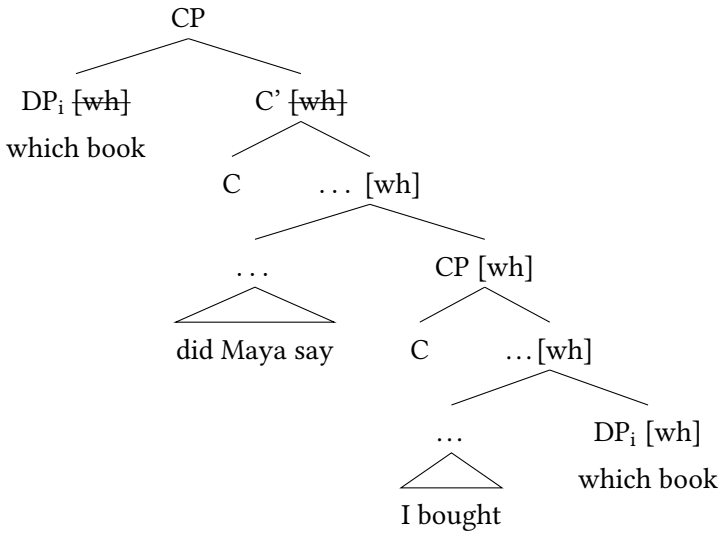


One alternative view is that the movement occurs in a single step, with a percolated selectional feature (e.g. Neeleman & Van De Koot 2002, 2010):[9]

[8] Chomsky (2008) has argued that the narrow syntax is optimized for LF interpretation, and that PF externalization is a secondary consideration. I consider my claim that hierarchy is a condition of both LF and PF to be compatible with such LF-bias views.

[9] Another alternative view is that of multi-dominance (cf. Citko & Gračanin-Yuksek 2021), illustrated in subsection 5.2.

(17)



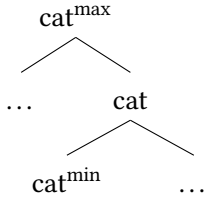
Both alternatives can seemingly be justified at a surface level with reference to theoretical principles: successive-cyclicity reduces the distance of movement, and feature percolation reduces the number of steps of movement. As a result, it is necessary to evaluate these proposals in a broader context, considering the effects on other phenomena (such as the intricacies of the effects of long-distance movement), rather than relating derivational architectures to principles in isolation.[10] This is a key consideration for the evaluation of theories of labeling.

A final relevant principle is that of *Inclusiveness*:

(18) *Inclusiveness*: no new objects are added in the course of computation apart from rearrangements of lexical properties. (Chomsky 1995: p. 209)

Inclusiveness restricts the character of proposed approaches to labeling. This restriction can be understood by comparing minimalist labels with those of X'-theory. Despite both being endocentric (with aforementioned exceptions in some minimalist proposals), minimalist theories of phrase structure cannot create a label in the same way that X' theories create X' and XP levels from a template. Instead, each label must either be an identical copy of a LI distinguished by relative prominence (19), or potentially a copy of a subset of the content of a LI.

(19)



When labels are formed by copying a (subset of a) LI, the LI may be said to project. I take the following to formally define *projection*:

[10] See e.g. Van Urk (2020) for such an evaluation of long-distance movement.

(20) *Projection*: the copying of a feature or features to the *label* of a *complex SO*.

Generally, *Inclusiveness* can be viewed as restricting operations in the derivation to the establishment of relationships between existing LIs. When *complex SOs* are formed, they are ultimately a reflection of a relationship between existing items, rather than something entirely new themselves.

2.3 Early BPS: the first labeling process

In chapter 4 of the *Minimalist Program* (Chomsky 1995, "Categories and Transformations," henceforth: C&T) , the initial BPS proposal introduces the concept of a *labeling* process. The proposal is notable for its inconsistent approach to *labeling*. Contradictory ideas from C&T provided a basis for debate, and later directly led to two competing theoretical strands.

In C&T, labeling is a derivational process (and thus formal *labeling*) consisting of Merge creating a *label* γ for each output of its application:[11]

(21) Merge(α, β) = $\{\gamma\ \{\alpha, \beta\}\}$ (C&T: p. 243)

The inclusion of *labels* is justified by the assumption that they are necessary for interpretation, with reference to the function of identification (C&T: p. 243).[12] C&T acknowledges that the introduction of a label takes Merge away from its simplest form of creating just the set $\{\alpha, \beta\}$ (*Simplest Merge*, an important point of comparison in the development of labeling). The biggest challenge to Simplest Merge is the lack of an automatic encoding of hierarchy: the *symmetry problem* (Cecchetto & Donati 2015: p. 27). Either hierarchy must be introduced in another manner for each Merge, or any symmetry allowed to enter the derivation must be repaired before *interpretation*. C&T, in assuming that a derived label is necessary, takes the former path.

C&T lists three possibilities for the derivation of γ from α and β: an intersection, union, or choice of the feature sets of α and/or β (C&T: p. 244). The first two possibilities are summarily dismissed, leaving a choice between the two constituents to determine which projects.[13]

The inconsistency in C&T arises in addressing how a *label* is selected for each Merge; Move (IM) and Merge (EM) are treated separately. Move is taken to only apply when checking an uninterpretable feature in the label of the SO being extended. This reflects an *Attract*-based approach to movement, rather than *Greed*-based, where the constituent being moved motivates the operation (cf. Bošković & Messick 2017).[14] Greed-based movement is discussed in C&T, but it is assumed that in all

[11] The possibility of representationally-derived labels introduced at a later stage of the derivation is mentioned, though C&T assumes that a system of derivational *labeling* is a better fit to observed empirical data.

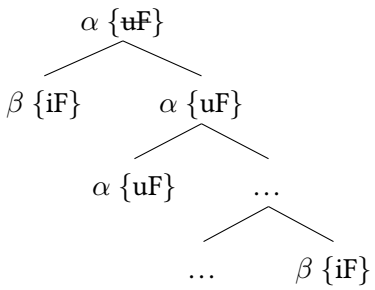
[12] Therefore not reflecting the requirements of *interpretation* as in (13).

[13] The intersection option would later be included in the Labeling Algorithm of Chomsky (2013), and the union option pursued by Zeijlstra (2020); see subsection 4.2 and subsection 5.1, respectively.

[14] The proposals in this survey differ in which of these approaches to movement they allow/require; I will note this distinction as relevant.

cases of Greed, there will be an uninterpretable feature checked in the SO being extended in addition to the SO being Moved (C&T: p. 267).[15] As a result, the SO being extended will always select for Move, and therefore *project* and label the SO:

(22)



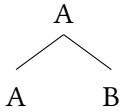
This method of labeling is commonly referred to as *projection by selection* (where *projection* meets (20)). While it provides a systematic approach to labeling for Move, it is not applied to (external) Merge. The proposal relies on the feature-checking inherent to Move, stated to be required to overcome a theorized additional computational cost compared with Merge. Merge is considered to be cost-free, and no clear selectional process is proposed; as a result, the question of which constituent provides a label for Merge is left open.

This dichotomy directly leads into subsequent approaches to labeling. A number of proposals generalize *projection by selection* to all cases of Merge and Move (discussed in section 3). Other approaches generalize Simplest Merge and remove consideration of selection, approaching the *symmetry problem* by allowing symmetry in the case that it is repaired (the Labeling Algorithm (LA), section 4). More recent alternatives entirely reformulate the structure-building operation, eliminating the symmetry problem before it arises (section 5).

2.4 Arguments against Labels

The possibility of applying Simplest Merge was considered at an early stage by Collins (2002). Collins argues that the *projection* of a *label* is unnecessary and uneconomical for the formulation of Merge, proposing replacing (23) with (24) as the output of Merge:

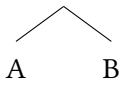
(23) a.



b. {A, {A, B}}

[15] Note that C&T therefore allows for mutual selection; the complexities of labeling mutual selection, and Greed-based movement generally, are discussed in subsection 3.2.

(24) a.



b. {A, B}

Collins focuses his discussion on derivational processes typically assumed to interact with *labels*, especially in relation to selection. Rather than selection occurring through features in *labels*, as in C&T, Collins proposes that selection features should be assumed to proceed with direct reference to heads. The set of selectional features for a constituent is termed the *locus*, and the locus of one LI must be exhausted before it can shift. No other LI can intervene between a locus and a constituent it selects for. Collins gives the following example, where "the" selects for a noun:

(25) a. {the, destruction}

b. *{the, destroy}

c. *{the, {destroy, {the, city}}}

d. *{the, {destroy, cities}} (Collins 2002: p. 52)

Seely (2006) defends and extends the proposal, focusing instead on the theoretical motivation for eliminating *labels*. Seely applies derivational c-command (Epstein 1999: p. 329) to formalize the proposal. Derivational c-command (26) is taken to underlie asymmetric relationships between constituents, and is formulated with reference to formal *terms* (C&T: p. 247):

(26) X c-commands all and only the terms of the category Y with which X was paired (by Merge or by Move) in the course of the derivation.

(27) For any structure K:

a. K is a term of K.

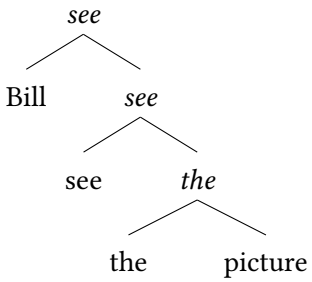
b. If L is a term of K, then the members of the members of L are terms of K.

Seely demonstrates that, when defining syntactic relations in this manner, *labels* do not enter into any relations. Consider the following:

(28) {*see*, {Bill, {*see*, {see, {*the*, {the, picture}}}}}}

(Seely 2006: p. 187, ex. 15)

(29)



Under the derivational definition of c-command, "Bill" c-commands the following constituents:

(30) a. {*see*, {see, {*the*, {the, picture}}}}
b. see
c. {*the*, {the, picture}}
d. the
e. picture

Notably, this does not include any of the *labels*, which are denoted with italics.

Seely further comments that given that *labels* are taken to be members of constituents, it is perhaps stipulative to suggest that the *label* can be accessed preferentially instead of other members. On the basis of these arguments, it seems logical that to suggest that *labels* are not necessary for the purposes of identifying or forming relationships between SOs. However, neither author addresses the need for hierarchy as a condition of *interpretation*, and their proposals therefore cannot be assumed to form SOs capable of meeting these requirements.

Seely also offers a critique of Merge with *labeling* as introducing unnecessary steps into the derivation, given the need to both copy the *label* and make two sets:

(31) With a projecting label: Merge(A, B) →
a. {A, B} →
b. A, {A, B} →
c. {A, {A, B}}

(32) Simplest Merge: Merge(A, B) →
a. {A, B}

This critique has merit outside of the context of eliminating *labels*. Even if formal *labels* should be retained, the process of *projecting* a *label* does appear to have some degree of additional complexity which is not clearly motivated outside of the need for hierarchy. The guiding question of this survey then emerges: what is the best way to ensure that asymmetry hierarchy is encoded in SOs?

3 Projection by selection

Early work in labeling generalized the idea of *projection by selection* (henceforth: PBS) to all instances of Merge (including movement). Most newer work considers PBS to be surpassed by more recent labeling proposals, due to a lack of motivation for the connection between *projection* and selection, and a number of challenges arising from the varied nature of selection. In this section, I argue that the key takeaway from PBS approaches is that selection can create the asymmetric relationships necessary for *interpretation*.

In PBS, the formulation of Merge from C&T is retained (with exceptions, to be noted), with γ corresponding to a *projected* constituent:

(21, repeated) Merge(α, β) = $\{\gamma\ \{\alpha, \beta\}\}$ (C&T: p. 243)

A key issue noted by Zeijlstra (2020) is that the connection between selection and *projection* appears quite stipulative: there is no obvious reason why selection should determine what projects. PBS arose from the assumption that *projection* is necessary to meet *interpretation* requirements (as in C&T; see subsection 2.3). These approaches also generally assume that each instance of Merge must be motivated by selection; this relationship can therefore guide the process of *projection* due to its inherent asymmetry. However, the provision of such inherent asymmetry quite naturally meets the conditions of *interpretation* itself. *Projection* is arguably the stipulative aspect: it may be one way to encode the inherently asymmetry of selection, but it is reasonable to suggest that there could be an alternative method to encode this asymmetry.

In light of the arguments in subsection 2.4, I will assume that PBS is not the ideal manner to encode asymmetry stemming from selection. However, the existing PBS literature provides key insight into the varied dynamics of selection. This diversity has been considered a challenge for PBS, given the difficulty of comprehensively accounting for these dynamics (see e.g. Zeijlstra 2020), but nonetheless provides a possible basis for asymmetry in the derivation.

3.1 Minimalist Inquiries

An early proposal to generalize PBS features in Chomsky's "Minimalist Inquiries" (Chomsky 2000, henceforth: MI). MI retains the projection method proposed for Move by C&T. The selectional requirement is extended to Merge, but considered to take the form of an interpretable semantic feature denoting a theta-role (MI: p. 134), instead of an uninterpretable feature which must by checked by Agreement with another constituent (the method for Move). In either case, the selectional feature is considered to be the only feature which enters the operation; its accessibility is than said to motivate the choice of that constituent to project.[16]

MI is notable for its use of semantic features for selection in the derivation, in response to the dichotomy between syntactic *c-selection* and semantic *s-selection* in local selection.[17] Zeijlstra (2020) illustrates the lack of a one-to-one mapping between c-selection and s-selection. A single theta-role may be filled by constituents of multiple syntactic categories:

[16] This does attempt to motivate PBS; however, an outstanding issue remains in that the primitive status of *projection* is stipulative itself, from an assumption that it is necessary to provide identification of SOs (cf. subsection 2.3).

[17] cf. Grimshaw (1979), Pesetsky (1982).

(33) to know

a. [$_{VP}$ know [$_{DP}$ Mary]]

b. [$_{VP}$ know [$_{PP}$ about Mary]]

c. [$_{VP}$ know [$_{CP}$ that Mary has left]] (Zeijlstra 2020: p. 34, ex. 6)

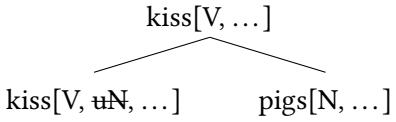
This does not mean that c-selection does not occur; Svenonius (1994b: ch. 2) has argued that all head-complement configurations are characterized by relationships which must reflect c-selection, rather than just s-selection. The use of s-selection in MI is curious, as it is not clear how semantic features can interact with the derivation for *projection*, given the assumption that the narrow syntax is otherwise independent of semantic processes.

MI provides additional interesting commentary which lays groundwork for the proposal of the LA. Chomsky states that labels are redundant, being "generally" predictable for any pair α, β (MI: p. 133), and even "indicated only for convenience" (MI: p. 135), otherwise being determined by a function Label(x). Labels in MI are redundant due to their derivation from inherent asymmetry induced by selection, which does still clearly mark the proposal as PBS. Nevertheless, these statements highlight the presence of ideas which would lead to the LA from an early stage in the development of labeling.

3.2 Core Syntax

Adger's textbook *Core Syntax* (Adger 2003, henceforth: CS) provides an interesting further development on PBS. CS is most notable for its formalization of c-selection through categorial features. Categorial features are postulated to occur as uninterpretable features when selectional, which can only be checked when Merged in a sister relationship with a constituent containing the interpretable counterpart:[18]

(34)



(CS: p. 91, ex. 135)

In every case, the constituent which c-selects then *projects*.

The use of c-selection rather than s-selection as the basis for PBS creates a strong stance on the independence of the narrow syntax from semantics.[19] Theta-roles associated with arguments are taken by Adger to be interpreted at the LF interface (CS: p. 83), which can account for the lack of a one-to-one mapping between theta-roles and properties of c-selection.[20] The use of c-selection brings the mechanism of Merge more in line with Move compared with MI, since selection in both cases uses formal features. However, not every case of Merge is motivated by c-selection. CS also adapts the ideas of the Extended Projection (Grimshaw 1991) as a "Hierarchy of

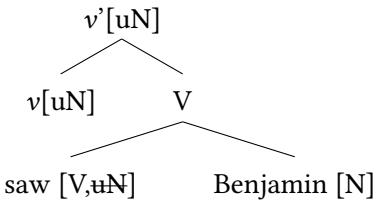
[18] This develops Svenonius's (1994a) proposal that c-selection is feature checking.

[19] cf. discussion by Chomsky (2004).

[20] Alternatively, Baker (2003) has critiqued this view of complete independence.

Projections", which must be followed and allows e.g. little v to *project* when Merged with V:

(35) (CS: p. 144)

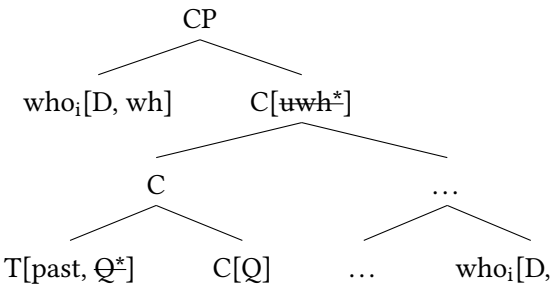


Such operations can be viewed as deriving their asymmetry from the requirements of Extended Projection.

Returning to consideration of the selectional cases, while Merge and Move both engage formal features, the features involved are not the same, reflecting a broader difference between selection in local (Merge) and long-distance (Move) instances. C-selection is notably restricted to local relations (Baltin 1989); selection in non-local relations, such as movement, generally occurs through the checking of non-categorial features, such as a *wh*-feature.

CS models non-local selection using both Attract- and Greed-based agreement. Wh-movement demonstrates Attract-based agreement, where C checks its [uWh] feature by Agreeing with a D-head. The D-head is then moved to a sister configuration with C due to a posited strength distinction on features, where strong features must be checked under sisterhood:

(36)



(adapted from CS: p. 351, ex. 54)

This illustrates an issue that non-categorial feature checking can also occur without subsequent movement, and therefore *projection*. The strength property proposed to account for this in CS is rather stipulative, demonstrating the challenge in adapting existing feature checking to model selection in non-local configurations.[21]

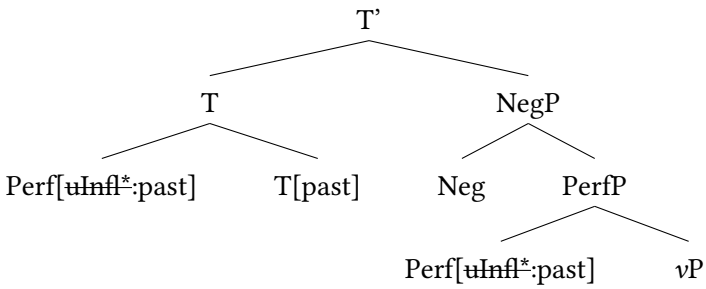
Greed-based Agreement is applied in the case of head movement,[22] illustrated by the movement of perfective auxiliaries to T. T values an uninterpretable inflectional feature[23] on Perf:

[21] Zeijlstra (2020) provides a comprehensive proposal addressing this challenge; see subsection 5.1.

[22] Various approaches to head movement have been proposed (cf. Dékány 2018), many of which would prevent PBS from occuring. I put the issue aside in order to illustrate this point.

[23] The unvalued/valued feature dichotomy was proposed by Chomsky (2001) as a replacement for the (un)interpretable dichotomy; cf. Zeijlstra (2014) for discussion and critique.

(37)



(adapted from CS: p. 180, ex. 120)

The constituent which needs to be checked (and therefore selects for the operation) is the lower constituent, with the uninterpretable and unvalued feature. Despite this, CS clearly proposes that in every case of Move, the already valued constituent is the one that *projects*. PBS does not seem to naturally describe this tendency, demonstrating a challenge for PBS approaches in accounting for selection under Greed.

3.3 The Probing Algorithm

A final notable proposal for PBS comes from Cecchetto and Donati (Cecchetto & Donati 2010, 2015, henceforth: CD).

CD propose the *Probing Algorithm*:

(38) *Probing Algorithm*: the label of a syntactic object $\{\alpha, \beta\}$ is the feature(s) that act(s) as a probe for the merging operation creating $\{\alpha, \beta\}$.

Probes are unvalued features which seek corresponding *goals*, matching features which can value the probe. The featural relationship applied is similar to CS , but is Attract- rather than Greed-based. In this manner, CD generalize PBS more consistently across local and long-distance selection.

CD do not only model probing through the explicit checking of features (though this is employed). Instead, CD also propose that *words* (roughly taken to mean non-functional LIs) are intrinsically probes. This initially seems to reflect the argument by Svenonius (1994b) that head-complement relationships always involve c-selection: the head (a *word*) will always select for the complement, so it can be assumed that they will act as probes, regardless of what the exact nature of the selectional features is. To explore this in further depth, consider cases in which both constituents select for Merge (note that CD follow Chomsky (2004) in viewing Move as a subcase of Merge). Such cases are referred to as *mutual selection* by Zeijlstra (2020). CD propose two categories of outcomes for mutual selection: either a label is unambiguously selected, or an ambiguity arises, with potential consequences.

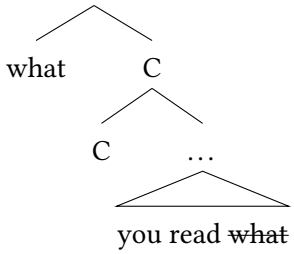
The unambiguous case is presented as being more straightforward, receiving only a brief introduction (CD: p. 40). CD propose this case specifically in relation to the problem of First Merge, where two *words* are Merged:

(39) Merge(saw, John) (CD: p. 40, ex. 20)

In this configuration, CD acknowledge that "saw", a verb, must select for "John", its nominal complement. The issue is that "John" is also a *word*, and therefore an intrinsic probe. CD propose that because "saw" is both a *word* and also must select for an argument,[24] it is a *double probe*, and in all cases will provide the label. The unambiguous case illustrates that CD do not see the intrinsic probe status of *words* as a reflection of c-selection; as a result, it is only mutual selection in the sense created by CD's proposal, but in a more natural sense appears to be straightforward c-selection.

The ambiguous cases are proposed to arise when a probing object triggers internal Merge of a *word*. The classic example is of structures such as (40):

(40) (CD: p. 41, ex. 24)



In this example, C is said to have probed for "what"; since "what" is a *word*, either it or C could provide a label (and C is not a double Probe, since it is not a *word*). CD propose that in such cases, either option for a label is valid. If "what" *projects*, the structure can be a free relative; if C projects, the structure can be an embedded clause:

(41) [$_{?P}$ what [C [you read ~~what~~]]]

a. I read [$_{whatP}$ what you read]

b. I wonder [$_{CP}$ what you read]

CD extensively discuss the prospect of such ambiguous cases, but provide no clear proposal for how the ambiguity proceeds and is resolved, only that it is possible for either option to be selected. It therefore seems plausible to suggest that the two outcomes of the ambiguity may actually reflect two different instances of selection: if C provides a label, Attract-based selection has occurred; if "what" provides a label, then it has selected for the Merge (through Greed-based selection).

Ultimately, both apparent cases of mutual selection as explored in CD seem to potentially reflect fairly standard, asymmetric selection. In general, it appears as though an argument could be made that any such case would actually reflect some asymmetry, even if some degree of mutual selection occurs (as is the case for the proposal by Zeijlstra (2020); see subsection 5.1). Mutual selection does not then provide a major challenge to the notion that selection can provide the

[24] It is unclear whether CD mean to engage syntactic selectional features in this case.

fundamental asymmetry necessary for *interpretation*. The strength of PBS is still limited by the stipulative nature of *projection*, but exploration of the diversity of selection represented by these approaches supports the conclusion that, while there is complexity to account for, asymmetric selection occurs syntactically in all cases of Merge, and could provide a strong account for derivational *labeling* if paired with a less stipulative proposal to capture that asymmetry.

4 The Labeling Algorithm

The Labeling Algorithm (LA) developed as an alternative to projection by selection approaches. Proposals for the LA draw from ideas such as Collins & Seely's arguments against *labels* (subsection 2.4) and Chomsky's suggestion that labels are determined rather than created in MI (section 3). I have established these influences in previous discussion, and in this section, I will focus on direct proposals for the LA and related applications.

4.1 On Phases

The canonical LA appears in Chomsky's "Problems of Projection" (Chomsky 2013, henceforth: POP) and "Problems of Projection: Extensions" (Chomsky 2015, henceforth: POPE)[25]. The initial proposal, however, appears in "On Phases" (Chomsky 2008, henceforth: OP). While less discussed, this earlier version illustrates the development of the LA and presents a notably different view of its role.

OP introduces the first proposal for the LA (henceforth: LA-08) by stating that "the label of an SO must be identifiable with minimal search" (OP: p. 145). This is the key shift which occurs with the development of the LA: labels are not explicitly derived (and therefore not formal *labels* (6)); rather, they are identified, or perhaps identifiable. In this sense, they are formed by relationships which can be determined through minimal search, generally assumed to correspond to relative prominence.

OP is ambiguous as to the issue of whether labels are actively identified by LA-08, or simply identifiable. If labels are identifiable, then they can exist outside of being determined by LA-08, and LA-08 is simply a formalization of what form these relationships must take. If labels are actively identified by LA-08, then it is implied that the information associated with a label would not be available until the application of the LA. OP does not make a clear statement about how the LA-08 should be viewed in this sense. However, Chomsky does discuss labels as triggering further applications of Merge after an SO is formed, conflating the view of OP with that of Collins' *locus* (OP: p. 141; see subsection 2.4 for discussion of the *locus*); this implies that labels are, in some manner, formed each time an SO is created.

The distinction between whether labels are identified or identifiable by LA-08 highlights a general consideration for LA-type approaches. These proposals initially developed from the view that labels are redundant reflections of underlying asymmetric relationships (see discussion of MI in subsection 3.1). If this is the case,

[25] POP and POPE will be distinguished where relevant, but where the distinction is ambiguous the collective proposal will be referred to as POP(E).

then this information from the relationships should always be available, even if not formally encoded in a label. When this assumption is made, it entails by extension that some form of asymmetry will be a fundamental characteristic of each SO to meet the conditions of *interpretation*. OP appears to make this assumption; the later proposal in POP(E) notably does not (see subsection 4.2). This unique aspect of LA-08 reflects the gradual shift of Chomsky's proposals from PBS to the LA.

This is also reflected in the formal proposal for LA-08:

(42) *LA-08*:

a. In {H, α}, H a [lexical item], H is the label.

b. If α is internally merged to β, forming {α, β} then the label of β is the label of {α, β}. (OP: p. 145)

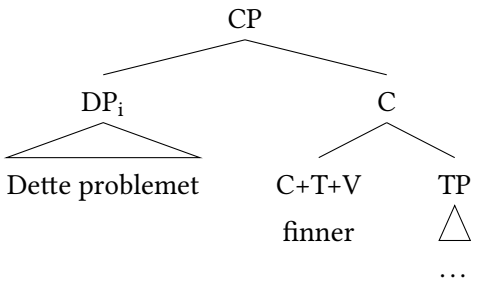
Most notable is the clear split approach between two options, which Chomsky even refers to as separate "algorithms", plural. This split parallels the earlier split in C&T, with the two options reflecting the expectations for labeling in cases of EM(42a) and IM(42b).

The mechanism in (42b) makes the same prediction as C&T: the label is taken from the SO being extended. No explicit reference to selection is made, but given the link made between labels and Collins' *locus*, it can reasonably be concluded that the choice of β as a label occurs because the label of β acted as the selector for the IM. Consider e.g. movement to the left periphery, as exemplified by topicalization in Norwegian (Bokmål):

(43) *Dette problemet*$_i$ *finner de aldri en løsning på* $\_i$
this problem.DEF find they never a solution on

"They will never find a solution to this problem." (Faarlund 2019: p. 220)

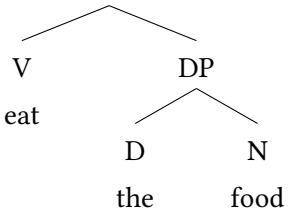
(44)



Assuming that topicalization is feature-driven, the C head would be seen as selecting for the Merge by searching for a constituent to agree with. As a result, it is the label in the sense of having entered into Merge to motivate the IM, and thus provides the label for the new SO. This configuration is striking in its replication of the proposal in C&T, and parallels Attract-based PBS labeling as well.

(42a) seems more distinct, and illustrates the link to prominence in minimal search. If LA-08 must find a LI to provide the label, then the LI of the head will be found rather than an LI in a phrase, given its relative prominence:

(45)



While this mechanism does appeal to prominence, it also can be viewed as reflecting a continued assumption of an asymmetric head-complement relationship, even if not rendered using featural c-command. This shows a clear parallel to the expectations of labeling in EM, but (42a) is notably not restricted to only applying in cases of EM (whereas (42b) is restricted to only IM). It follows that if a head is moved, both algorithms may apply, and would derive different labels. OP allows for this prediction by referring to CD's cases of free relatives (see subsection 3.3).

OP further references other configurations where LA-08 would fail to determine a label; these directly correspond to cases explored in depth in POP(E), so will be discussed in subsection 3.2. I note, however, that these cases undermine the potential view that LA-08 is simply a reflection of how labels should be identifiable, rather favoring the view that LA-08 explicitly identifies labels, and that the asymmetry necessary for *interpretation* is not always available. The consequences of this stance provide both interesting predictions and major challenges, and are illustrated in the forthcoming discussion of POP(E).

4.2 Problems of Projection and Extensions

The LA is developed in more depth in POP(E). Labeling is said to be "the process of providing…information" for the identification of SOs at the interfaces (POP: p. 43). This represents a more consistent shift in the role of the LA to actively identifying labels. However, the focus on identification overlooks the requirement of asymmetric hierarchy in *interpretation* (subsection 2.1). As I will argue, this leaves the proposal insufficient in accounting for the *symmetry problem* of set-theoretic Merge (subsection 2.3).

No succinct formal definition of this version of the LA (henceforth: LA-13) is given; I adapt wording from POP to take the following:

(46) *LA-13*: The labeling algorithm LA licenses the interpretation of SOs at the interface by applying minimal search to find the most prominent head or feature.

POP generally focuses on demonstrating possible configurations of LA-13 rather than exploring theoretical properties of LA-13; POPE provides a clearer picture, though

conflates existing terminology in a confusing manner. LA-13 is stated to determine what projects to license the interpretation of SOs by enabling their identification. At the same time, LA-13 is said to not create any new item, only determining a property of SOs. These two statements clearly conflict with the formal notion of *projection* (20); it should be taken that no labels of LA-13 *project* in this sense, rather only providing identification, and leaving the requirement for asymmetric hierarchy unaddressed in certain cases (as illustrated below). LA-13 is also situated within a cyclic view of the derivation (Chomsky 2000): the operation is proposed to apply during *Transfer* of phases to the interfaces, and cannot search within SOs which have been spelled-out (the Phase Impenetrability Condition (Chomsky 2001: p. 13)).

I will demonstrate the character of LA-13 and especially the problematic instances of symmetry by considering the three possible configurations encountered by LA-13, and the corresponding proposed outcomes.

To begin with the most straightforward configuration:

(47) SO={H, XP}, H a head: H labels the SO.

No issues are encountered: LA-13 can consistently identify H, the head, as the label, to provide identification information. H is identified due to its relative structural prominence, which can also be taken to meet the need for asymmetric hierarchy in *interpretation*. The asymmetry is not created as a reflection of inherently asymmetric relationships from selection, but arguably could be taken to reflect the assumption of asymmetric c-selection in head-complement configurations. It is therefore not a complete departure from the approach of PBS, though does still derive in a different manner.

Next, consider cases where both constituents are heads:

(48) SO={H, H}, both heads: assumed to only occur as root-categorizer constructions, where roots are not visible to LA-13 and the categorizer provides the label.

POP adopts the proposal of Distributed Morphology (Marantz 1997) that roots must be combined with a functional element which provides a category. This is taken to be the only instance of head-head configurations which arise, since all LIs must be categorized, and subsequently will be *complex SOs*.[26] Another potential instance of {H, H} structures, head movement, is taken by Chomsky (2000: p. 38) to occur at PF, meaning that no head-head configuration ever arises syntactically for these cases, and there is therefore no need to derive a label for these constructions.

The third configuration consists of two Merged phrases. In this case, POP(E) states that neither head is more prominent, so some repair or alternate labeling strategy must be engaged:

[26] This appears to then restrict {H, XP} configurations to only occuring with functional heads, giving them a special status reminiscent the Borer-Chomsky Conjecture of parametric variation (Borer 1984).

(49) SO={XP, YP}, neither a head: LA(SO) requires either:

a. IM of XP or YP, leaving a single visible head; or

b. Agreement between XP and YP, with shared feature(s) labeling the SO.

Beginning with (49a), the solution of triggering further movement was first mentioned in OP, influenced by the proposal of *dynamic antisymmetry* (Moro 2000), which suggests that XP-YP configurations are inherently unstable, and must trigger repair movement.[27] Consider the following example, where β is the constituent being labeled:

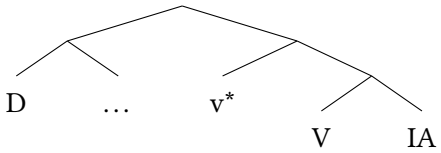
(50) T [$_\beta$ (EA) [v* [V IA]]] (POP: p. 44, ex. 17)

In the case that the EA raises, v* is clearly the most prominent lexical head (though application of LA-13 is still not straightforward, as will be discussed). However, consider a case in which [V IA] moves. POP states in the main text that "the part of the structure visible to [LA-13] is EA-v*", but admits in a note that "technically, what is visible to [LA-13] is {EA, {v*}}":

(51) T [$_\beta$ EA [v* ~~[V IA]~~]]

Chomsky posits that this forces search into singleton sets to be considered minimal (POP: note 34), but this does not easily follow from minimal search using structural distance.[28] An easier solution would be the assumption that the EA is a phase itself (reasonably the case if it were a DP). Since LA-13 is subject to the Phase Impenetrability Condition, it cannot access the EA to search for a head to provide the label. However, it would then appear possible for v* to be selected as the label from the start:

(52)



This example illustrates a lack of motivation for the inability of LA-13 to label all {XP, YP} configurations. Instead, it must be stipulated that, for the movement repair, LA-13 is simply prevented from applying until movement has occurred.[29] This introduces a further key issue: there is no accepted approach to distinguishing copies from repetitions (Collins & Groat 2018). This is problematic not just for the

[27] OP and POP argue that the LA can derive dynamic antisymmetry, a view later developed in Moro & Roberts (2024).

[28] This undermines a core motivation for LA-13: minimal search is considered a third factor cognitive function, and so does not need to be posited as content of UG (POP: p. 39). If LA-13 cannot follow from the application of minimal search, it is then rendered quite stipulative.

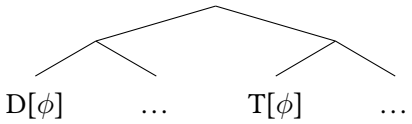
[29] Though this movement is also considered to have been triggered by LA-13, leading to questions of timing and ordering, especially in relation to the assertion that LA-13 applies at Transfer.

conclusion above, but also for the entire repair approach predicated on moved (and thus copied) constituents being unable to provide a label.

It additionally entails that the movement repair approach does not naturally induce asymmetric hierarchy. In the example above, a symmetric structure will remain after movement has occurred; this requires hierarchy to be derived from the identity determined by LA-13. This is possible in the case of repair movement, since a single constituent is selected as the label. However, as discussed in subsection 2.1, this is not possible in all cases, a notable issue for the feature-sharing strategy.

Shifting to evaluate (49b), consider the classic example for feature sharing, Spec-T configurations:[30]

(53)



Firstly, it is important to note the unique property which led to the proposal of a separate repair strategy for this and similar {XP, YP} structures: neither constituent should be moved from this position, and so for labeling to occur, another strategy must be proposed. This is also referred to as the "halting problem" (cf. Rizzi 2015), in relation to the question of why successive-cyclic movement stops at certain positions – relevant given that the movement repair strategy is assumed to derive successive-cyclic movement in the first place. As a result, the proposal for this solution is motivated less by the nature of LA-13 itself, and more by the challenge that these structures posed; the observation that there are often shared features which Agree provides a convenient mechanism, but not a well-motivated one.

The positioning of the features used to label (53) further illustrate the surface-level nature of this proposal. It was assumed in the discussion of (50–52) that DP is a phase, since this was necessary to enable the proper outcome of LA-13 in the example given in POP. As a result, LA-13 should be unable to examine within DP to find a potential feature to use as a label. This clearly is at odds with the assumption that a feature on the D head is used in conjunction with T to label this SO. In POP(E), Chomsky only considers CP and v*P to be phases, which would allow this to occur; however, a clear theoretical inconsistency emerges as a major limitation of the proposal. Further, if DP is not a phase, its head is more prominent than T; it is thus not clear why a shared feature of two heads on different levels would be more prominent than this single head.

Finally, as mentioned above, while this strategy can provide identification, it is not possible to derive hierarchy from the identity chosen.[31] As a related consequence, the endocentric character of SOs, otherwise retained since its introduction in X'-theory (cf. section 1), is now assumed to not arise for a subset of SOs.

[30] I assume phi-features are hosted in D for simplicity of this illustration; for discussion of relevant issues, see Inokuma (2013).

[31] The one possibility which comes to mind would be to appeal to the original Probe-Goal relationship formed by Agree; however, this is clearly a departure from the dynamics of labeling as proposed in POP(E).

Consideration of the three possible cases for LA-13 has illustrated the nature of LA-13 and its consequences. The departure from previous approaches to labeling by allowing symmetric structures to be created provides some interesting predictions which motivate known syntactic phenomena. However, the dynamics of LA-13 are quite internally inconsistent, and the allowed symmetry which is at times creates interesting effects is also not always repaired, leaving this theory of labeling unable to consistently generate derivations which meet the need for asymmetric hierarchy in *interpretation*.

4.3 Applications of the LA

The LA has been adopted and applied to a number of different phenomena. These applications have produced interesting results, but have responded to the inconsistency in POP(E) with diverging assumptions about the exact character of the LA. As a result, it is not clear that these applications can be considered to reflect a general success of the LA as an approach to labeling. Further, it is uncertain whether these departures from POP(E) are trending toward a single unified account of the LA, or are too different to be considered a single approach to labeling. I will briefly introduce a set of applications to illustrate this point.

Blümel (2024, 2025) has applied the strength distinction posited for T heads in POPE to the nominal categorizer n. This is proposed to motivate an NP/DP typology, with DP languages requiring a DET-P to enable shared feature labeling, to rectify the inability of "weak" n (n_{wk}) to provide a label:

(54) a. *$[_a\ n_{wk}\ \sqrt{root}]$

b. $[_a$ DET-P $[n_{wk}\ \sqrt{root}]]$ (adapted from Blümel 2024: p. 16, ex. 34)

Blümel does not depart from the mechanism of POPE ; however, the lack of label in (54a) is not rectified, since the shared features only label the SO which contains it. This highlights an issue from POPE which remains unresolved in recent a recent application.

Bošković (2016, 2018, 2020) has argued that the timing of labeling with phases can derive successive-cyclic movement and island constraints, such as the subject condition:

(55) *I wonder who$_i$ [friends of t_i] left. (Bošković 2016: ex. 15)

Bošković's analysis relies on labeling only applying at Transfer of phases, and an *antilocality* restriction requiring movement to cross a distinct labeled projection. In (55), "who" would have to move to the phase edge of DP before DP is Transferred; it is then prevented from moving to SpecCP due to *antilocality*, since the feature-sharing label of the Spec-T constituent cannot be labeled until CP is Transferred:

(56) … $[_{CP}$ C $[_?$ $[_?$ who $[_{DP}$ subject]] $[_{TP}$ T… $[_{vP}$

(adapted from Bošković 2016: ex. 17)

The architecture of labeling as applied by Bošković is sharply contrasted by that employed by Roberts (2019). Roberts proposes a theory of word order variation arising from the requirement to label *complex SOs* consisting of a categorized LI and a functional head:

(57) $[_{FP} \ldots F_{[Cat]} [_{xCatP} XP\ x_{[Cat:x]}]]$ (Roberts 2019: p. 159, ex. 111)

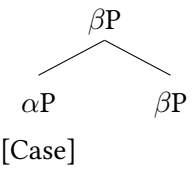
In order to label FP, Roberts (2019) proposes that F's Cat-feature must be valued; this can co-vary with the strength of Cat, where only strong and valued Cat can provide a label. If Cat is strong and unvalued, it must agree with x_{Cat}, creating head movement. If Cat is weak, then the x_{Cat}P must "roll-up" and move to c-command F:

(58) $[_{FP} [_{xCatP} XP\ x_{[Cat:x]}] [F_{[Cat]} ([_{xCatP} XP\ x_{[Cat:x]}])]]$
(Roberts 2019: p. 160, ex. 112)

If weak cat is valued, then labeling proceeds via feature-sharing; if it is unvalued, head movement applies after roll-up.[32]

These potential configurations are stated to derive a typology of word order which does not violate the final-over-final condition (cf. Sheehan, Biberauer, Roberts & Holmberg 2017). In order for LA to trigger these derivations, Roberts explicitly states that he breaks with POP(E) and posits that "labeling must take place as soon after Merge as possible" (Roberts 2019: p. 157), to meet an assumption that a label is required for any further operation involving a *complex SO*. This assumption was further developed by Moro & Roberts (2024), who argue that this approach to labeling can (partially) derive dynamic antisymmetry (Moro 2000).

Another notable application is by Saito (2016, 2017). Saito proposes that suffixal case in languages like Japanese makes phrases opaque to search, and therefore unable to provide a label. In (59), Case prevents αP from providing a label, so the head of βP will label the SO:

(59)


(adapted from Saito 2016: ex. 20)

Saito (2016) argues that this proposal explains phenomena including DP-scrambling and argument ellipsis patterns, and Saito (2017) extends the account to argument doubling. While these explanations are compelling, the nature of the proposal itself notably departs from POP(E). There is no clear motivation for why Case should render a constituent opaque to search; it is reminiscent of the proposal in POP(E) that moved constituents cannot provide a label in that both are quite stipulative, and therefore appears to move the theory of the LA in an undesired direction.

[32] This roll-up movement appears to violate the antilocality central to Bošković (2016: et seq.), a further inconsistency.

The examples discussed are only a sample of proposed applications for the LA. However, even this sample illustrates the clear diversity of assumptions in this literature, many of which diverge from POP(E). Determining the exact consequences of this diversity is beyond the remit of this dissertation, but the necessity of such consideration sows doubt as to the success of the LA.

5 Alternatives to Merge

In the final section of this survey, I consider two recent proposals which replace the process of *labeling* entirely, by proposing new structure-building operations which inherently encode asymmetric hierarchy. These proposals build on the key finding from section 3: selection provides fundamental asymmetry in the derivation, but *projection* by copying is a rather stipulative approach to encoding this asymmetry as hierarchy for *interpretation*.

5.1 Merge as a union

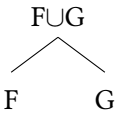
Zeijlstra (Zeijlstra 2020; henceforth: Z) proposes that rather than producing a set containing independent constituent feature sets (60a), Merge should create a union of the formal features of the constituents (60b):[33]

(60) a. Merge(F, G) = {F, G}

b. Merge(F, G) = {F∪G}

Z then posits that this Merged set can essentially function in the same manner as a traditional projected *label*, yielding representations of the sort:

(61)



Z proposes a number of concordant characteristics of the narrow syntax, primarily reflecting the importance of features. Following Zeijlstra (2014), formal and semantic features are entirely separated. The uninterpretable-interpretable dichotomy is taken to simply reflect a dependence relationship, rather than anything about formal *interpretation*. The features themselves percolate up the tree through subsequent applications of Merge, until they are checked. This reflects the proposal of Neeleman & Van De Koot (2002, 2010) discussed in relation to long-distance movement in subsection 2.2, and further creates a consistent Greed-based approach to agreement. Feature checking follows from the formalization in CS (subsection 3.2); when an iF-uF pair is Merged, the features are checked and do not appear in the output union

[33] This proposal instantiates the union option for labels which in C&T is acknowledged as a logical possibility but summarily dismissed.

set. Merge is motivated by the resolution of at least one featural dependency, so every *complex SO* is created with an asymmetric selectional relationship.[34]

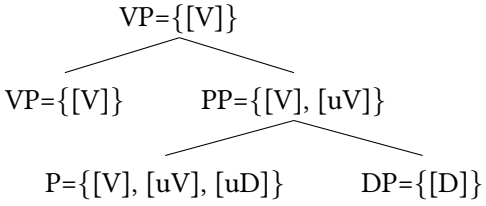
Before considering the encoding of asymmetry in further depth, I will first demonstrate the theory. Z proposes and illustrates a number of predictions which follow from the proposal. I will present one application: the argument structure of PPs. Z proposes that the label of PP must have the feature set {[V], [uV]} in order to act as adjuncts to VPs.[35] Given that Ps select for DPs, P is given the feature set {[V], [uV], [uD]} (Table 1):

Feature	Function
[V]	establishes hierarchy through percolation
[uV]	selects for a VP
[uD]	selects for a DP

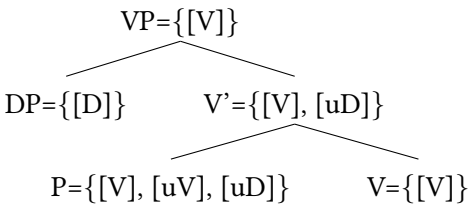
Table 1 Feature set of P.

The two selectional elements are unordered in the set; as a result, either should be able to be checked first. This yields alternatively:

(62) a. (Z: p. 44, ex. 22–23)



b.



(62a) demonstrates the typical assumed ordering of selection, with P selecting a DP, and then selecting a VP. In this application, Z provides an account of adjunction (notably difficult for theories of selection due to its optionality). Z proposes that adjunction arises when a constituent selects for an interpretable feature which it also carries; here, PP carries both [V] and [uV]. As a result, when PP and VP Merge, the percolated interpretable feature is still V, despite it also having been selected.

[34] Z explicitly sees the theory as an evolution of *projection by selection* approaches. There is a clear relationship, but I have separated Z's proposal from the approaches discussed in section 3 due to its unique approach to Merge.

[35] PPs can of course also be used in a similar manner with nominals, which requires the feature specification {[N], [uN], [uD]}; this is accounted for by a later proposal that V and N are subcategories of a single lexical category Pred (Z: p. 53).

(62b) at first seems abnormal, with P first Merging with a V, then later a D. Z proposes that this availability of this configuration predicts particle verb constructions, where complex verb heads are formed with adpositions (see e.g. Zeller 2001):

(63) [$_{V'}$ pick up [$_{DP}$ the food]]

The PP selectional ordering and its consequences illustrates the style of applications and predictions proposed by Z. While the theory does not appear to have been applied beyond Z itself, these predictions could provide the basis to properly evaluate the proposal in a broad context (as I advocated in subsection 2.2).

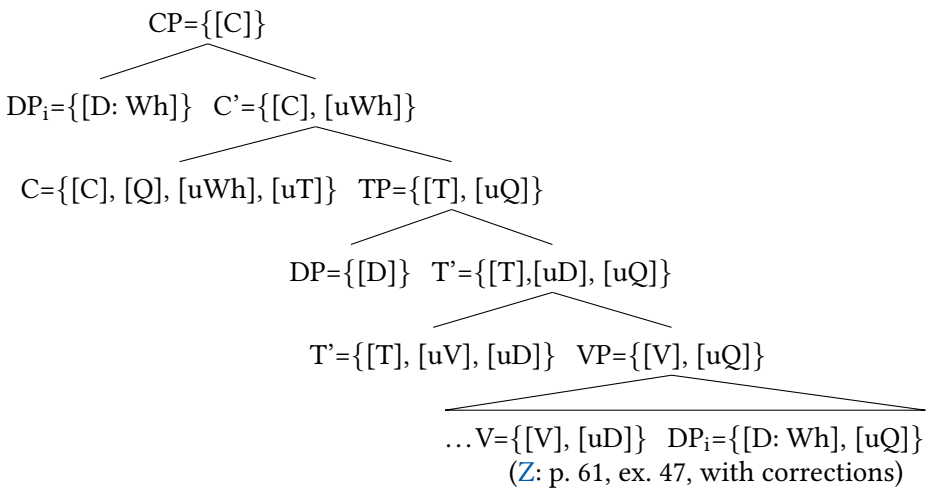
Returning to the question of encoding asymmetric hierarchy, since the proposal does not fully copy selecting constituents through *projection*, asymmetry must be encoded in an alternative manner. I argue that the percolation of interpretable features must be taken as the encoding of asymmetry. Firstly, feature checking occurs during the application of Merge, but is not retained in the output of Merge. This reflects the more general observation that input constituents are not retained in the output (the case with traditional PBS):

(64) a. PBS: Merge(F{F, uG}, G{G}) = {F {F}, {F {F, uG}, G{G}}

b. Union Merge: Merge(F{F, uG}, G{G}) = {F {F}}

Secondly, Z allows for cases of mutual selection, where both constituents select for a Merge. Consider the proposal for wh-movement:

(65)



(Z: p. 61, ex. 47, with corrections)

The DP selects for the movement with the [uQ] feature, while C selects for the movement with the [uWh] feature. These features are actually checked in two sequential applications of Merge: [uQ] is checked when C is Merged with TP, and [uWh] when the DP is Internally Merged to C'. However, the Merger of C to TP shows mutual selection in a single Merge (Table 2):

C	TP	Merge(C, TP):	C'
[C]	-	percolates	[C]
[uT]	[T]	checked	-
[Q]	[uQ]	checked	-
[uWh]	-	percolates	[uWh]

Table 2 Feature relationships and outcome of Merge(C, TP) in (65).

Further, (65) demonstrates that an uninterpretable feature can percolate much further than its head can reasonably be assumed to take scope. As a result, asymmetric hierarchy must be encoded by the interpretable features which percolate.[36] If it is also assumed that each set must contain at least one interpretable feature, it follows that Z encodes the asymmetry of selection as hierarchy for each *complex SO* formed, fully meeting the requirements of *interpretation.*

5.2 Mereological Syntax

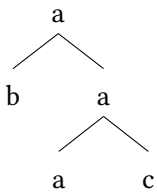
Adger's *Mereological Syntax* (Adger 2025; henceforth: MS) replaces set-theoretic Merge with the mereological function *Subjoin.* Subjoin creates part-whole relationships between the objects it acts on:

(66) Subjoin (x, y) $\rightarrow$ y : x $<$ y (MS: p. 42, ex. 2)

This operation can be viewed as acting on y by making x a part of it. Relationships are therefore determined by parthood, which is inherently asymmetric.[37] This eliminates the need for any process of *labeling* to encode asymmetry.

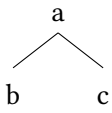
MS proposes that Subjoin can act on any object twice; each iteration places a dependent part into a different "dimension", used to derive two types of part-whole relations, corresponding to the traditional notions of complements and specifiers. Compare the following:

(67) a. Traditional BPS:



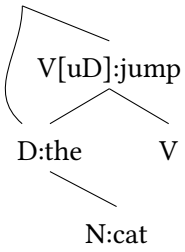
[36] This also entails that Z retains endocentricity for every SO.
[37] This approach does not retain endocentricity, since these constituent parts do not determine the function of the whole.

b. MS:



MS implements selectional features (following from CS), but only for the establishment of specifier-type parthood, in dimension 2:

(68) T[uD]:ed (MS: p. 62, ex. 41)



Subjoin in dimension 1 corresponds to the Extended Projection Complement relation (Grimshaw 1991), with the relevant objects (being functional categories) hierarchically ordered and inserted to satisfy the requirements of a given Extended Projection, rather than being selected. The Extended Projection does still provide a clear basis for asymmetry which can be encoded as hierarchy by Subjoin; still, this forms an interesting contrast with Z, which generalizes selection, and suggests the possibility of alternative sources of inherent asymmetry beyond just selection.

(68) also displays another key characteristic of MS: *multiparthood*. Multiparthood is illustrated with the DP "the cat", which is first selected by v, then also by T. Instead of a copy theory of movement, each constituent can form multiple parthood relationships. This clearly parallels *multidominance* in Merge-based syntax (cf. Citko & Gračanin-Yuksek 2021), providing an alternative which avoids the issue of distinguishing copies and repetitions (Collins & Groat 2018).[38] MS explores traditional conceptions of movement through this approach of multiparthood quite extensively, deriving a notion of locality related to the parthood dimensions which is proposed to predict a number of attested island constraints. Similarly to Z, this provides the basis for a new and untested theory to be evaluated in a broadly comparative perspective, rather than on theoretical architecture alone.

A final interesting point is that MS makes a clear effort to integrate earlier ideas in generative syntax into a proposal under the minimalist framework. For instance, Adger's critique of the general set-based system common under minimalism is situated within a historical survey of labels (MS: ch. 2), which forms the basis of the argument for a shift to again focusing on relationships between objects rather than sets. Z similarly looks beyond the mainstream ideas in minimalism, though rather than looking historically, draws acknowledged inspiration from categorial grammar. As a result, both proposals address issues with existing approaches in minimalist

[38] Z also provides an alternative, but through feature percolation.

syntax by integrating ideas from non-minimalist proposals. They clearly depart from mainstream BPS, nevertheless applying the common minimalist principles discussed in subsection 2.2. Their reformulated structure-building operations inherently encode asymmetric relationships as hierarchy, a natural reflection of both the availability of asymmetry and the requirements of *interpretation*. While still recent proposals adopted only by a single author each, their departures lead to wide-ranging, testable predictions, and they do not yet have the issue of inconsistency of application (see subsection 4.3). Moving forwards, their clear connection to labeling as an issue motivates consideration of these theories as direct alternatives to the LA (and traditional PBS), ideally within a broadly comparative perspective.

6 Conclusion

At the beginning of this survey, I posed the question: what is the best way to ensure that asymmetry is encoded in SOs generated by the narrow syntax? In section 2, I argued that the process of *interpretation* requires asymmetric hierarchy, and that identity, a frequently conflated characteristic, can be assumed to follow from asymmetry. I also examined the common minimalist framework and principles which underlie all of the proposals evaluated. From this discussion, it followed that approaches to labeling cannot be compared on adherence to principles in isolation. While this survey is limited in its scope, I established the above question to guide an initial evaluation of approaches to labeling.

To briefly recap the findings from each group of approaches: projection by selection (section 3) faces clear theoretical issues, but provides insight into how selection can provide an asymmetric relationship from which to encode hierarchy. The Labeling Algorithm (section 4) faces a serious issue in its lack of consistently ensuring that asymmetric hierarchy is encoded. It makes a number of interesting predictions which have been explored in a variety of applications, but the existing literature contains diverging assumptions and likely cannot be taken as reflecting the success of a unified approach. Finally, two recent proposals change or supplant Merge with operations which inherently encode asymmetry in *complex SOs* (section 5), and therefore consistently meet the requirements of *interpretation*.

On the basis of this survey, the recent proposals for alternatives to canonical Merge seem most likely to be the best approaches to ensuring the encoding of asymmetry, on the basis of their inherent approach to the issue. The proposals remain relatively untested, however, and merit additional application and consideration. Additionally, it is reasonable to suggest that other proposals could be made which inherently encode asymmetry in different manners, and that it is more this general character which motivates the success of the proposals evaluated. This further validates the necessity of broadly comparative evaluation of theories which meet the minimum requirements of approaches to labeling. On that note, it remains to be seen whether the LA can be considered to meet such minimum requirements as *interpretation*, motivating a clear need for focused evaluation of the LA and its applications to resolve outstanding inconsistencies.

References

Adger, D. 2003. *Core syntax: A minimalist approach*. Oxford University Press. doi:10.1093/oso/9780199243709.001.0001.

Adger, D. 2025. *Mereological syntax: Phrase structure, cyclicity, and islands*. MIT Press. doi:10.7551/mitpress/15777.001.0001.

Baker, M. C. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge University Press. doi:10.1017/CBO9780511615047.

Baltin, M. R. 1989. Heads and projections. In M. R. Baltin & A. S. Kroch (eds.), *Alternative conceptions of phrase structure*, 1–16. University of Chicago Press.

Blümel, A. 2024. Labeling therory and the nominal phrase. In J. Lu, E. Petersen, A. Zaitsu & B. Harizanov (eds.), *Proceedings of WCCFL 40*, 11–21.

Blümel, A. 2025. Revisiting diachronic change in the nominal domain from Latin to modern Romance. *Linguistic Variation* doi:10.1075/lv.24056.blu.

Borer, H. 1984. *Parametric syntax: Case studies in Semitic and Romance languages*. De Gruyter Mouton. doi:10.1515/9783110808506.

Bošković, Ž. 2016. On the timing of labeling: Deducing Comp-trace effects, the Subject Condition, the Adjunct Condition, and tucking in from labeling. *The Linguistic Review* 33(1). 17–66. doi:10.1515/tlr-2015-0013.

Bošković, Ž. 2018. On movement out of moved elements, labels, and phases. *Linguistic Inquiry* 49(2). 247–282. doi:10.1162/LING_a_00273.

Bošković, Ž. 2020. On the Coordinate Structure Constraint, across-the-board-movement, phases, and labeling. In J. V. Craenenbroeck, C. Pots & T. Temmerman (eds.), *Recent developments in phase theory*, 133–182. De Gruyter. doi:10.1515/9781501510199-006.

Bošković, Ž. & T. Messick. 2017. Derivational economy in syntax and semantics. *Oxford Research Encyclopedia of Linguistics* doi:10.1093/acrefore/9780199384655.013.130.

Cecchetto, C. & C. Donati. 2010. On labeling: Principle C and head movement. *Syntax* 13(3). 241–278. doi:10.1111/j.1467-9612.2010.00140.x.

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

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

Chomsky, N. 1970. Remarks on nominalization. In R. Jacobs & P. Rosenbaum (eds.), *Reading in english transformational grammar*, 184–221. Ginn.

Chomsky, N. 1977. On wh-movement. In P. W. Cullicover, T. Wasow & A. Akmajian (eds.), *Formal syntax*, 71–132. Academic Press.

Chomsky, N. 1995. *The minimalist program*. MIT Press.

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

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

Chomsky, N. 2004. Beyond explanatory adequacy. In A. Belletti (ed.), *Structures and beyond*, 104–131. Oxford University Press. doi:10.1093/oso/9780195171976.003.0004.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36(1). 1–22. doi:10.1162/0024389052993655.

Chomsky, N. 2008. On phases. In R. Freidin, C. P. Otero & M. L. Zubizarreta (eds.), *Foundational issues in linguistic theory: : Essays in honor of Jean-Roger Vergnaud*, 133–166. MIT Press. doi:10.7551/mitpress/7713.003.0009.

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

Chomsky, N. 2015. Problems of projection: Extensions. In E. Di Domenico, C. Hamann & S. Matteini (eds.), *Linguistik Aktuell/Linguistics Today*, vol. 223, 3–16. John Benjamins Publishing Company. doi:10.1075/la.223.01cho.

Citko, B. & M. Gračanin-Yuksek. 2021. *Merge: Binarity in (multidominant) syntax*. MIT Press. doi:10.7551/mitpress/12800.001.0001.

Collins, C. 2002. Eliminating labels. In S. D. Epstein & T. D. Seely (eds.), *Derivation and explanation in the minimalist program*, 42–64. Wiley. doi:10.1002/9780470755662.ch3.

Collins, C. & E. Groat. 2018. Distinguishing copies and repetitions. M.S.

Collins, C. & T. D. Seely. in press. Labeling without Labels. In K. K. Grohmann & E. Leivada (eds.), *The Cambridge handbook of the minimalist program*, Cambridge: Cambridge University Press.

Dékány, É. 2018. Approaches to head movement: A critical assessment. *Glossa: a journal of general linguistics* 3(1). doi:10.5334/gjgl.316.

Epstein, S. D. 1999. *Working minimalism*. MIT Press.

Faarlund, J. T. 2019. *The syntax of mainland scandinavian*. Oxford University Press. doi:10.1093/oso/9780198817918.001.0001.

Fukui, N. & Y. Takano. 1998. Symmetry in syntax: Merge and demerge. *Journal of East Asian Linguistics* 7(1). 27–86. doi:10.1023/A:1008240710949.

Grimshaw, J. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10(2). 279–326.

Grimshaw, J. 1991. Extended projections. M.S.

Inokuma, S. 2013. Distribution of phi-features within DPs and the Activity Condition. *English Linguistics* 30(1). 292–312. doi:10.9793/elsj.30.1_292.

Jackendoff, R. 1977. *$\bar{X}$ syntax: A study of phrase structure*. MIT Press.

Kayne, R. S. 1994. *The antisymmetry of syntax*. MIT Press.

Lyons, J. 1968. *Introduction to theoretical linguistics*. Cambridge University Press. doi:10.1017/CBO9781139165570.

Marantz, A. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2). 201–225.

Moro, A. 2000. *Dynamic antisymmetry*. MIT Press.

Moro, A. & I. Roberts. 2024. The duality of syntax: Unstable structures, labelling and linearisation. *Natural Language & Linguistic Theory* 42(2). 609–631. doi:10.1007/s11049-023-09588-z.

Neeleman, A. & H. Van De Koot. 2002. The configurational matrix. *Linguistic Inquiry* 33(4). 529–574. doi:10.1162/002438902762731763.

Neeleman, A. & H. Van De Koot. 2010. A local encoding of syntactic dependencies and its consequences for the theory of movement. *Syntax* 13(4). 331–372. doi:10.1111/j.1467-9612.2010.00143.x.

Pesetsky, D. 1982. *Paths and categories*: Massachusetts Institute of Technology dissertation.

Rizzi, L. 2015. Cartography, criteria, and labeling. In U. Shlonsky (ed.), *Beyond functional sequence: The cartography of syntactic structures, Volume 10*, 314–338. Oxford University Press. doi:10.1093/acprof:oso/9780190210588.003.0017.

Roberts, I. 2019. *Parameter hierarchies and Universal Grammar*. Oxford University Press. doi:10.1093/oso/9780198804635.001.0001.

Sailer, M. 2016. The syntax–semantics interface. In M. Aloni & P. Dekker (eds.), *The cambridge handbook of formal semantics*, 629–663. Cambridge University Press. doi:10.1017/CBO9781139236157.022.

Saito, M. 2016. (A) Case for labeling: Labeling in languages without -feature agreement. *The Linguistic Review* 33(1). 129–175. doi:10.1515/tlr-2015-0017.

Saito, M. 2017. Labeling and argument doubling in Japanese. *Tsing Hua Journal of Chinese Studies* 47(2). doi:10.6503/THJCS.2017.47(2).07.

Seely, T. D. 2006. Merge, derivational C-command, and subcategorization in a label-free Syntax. In C. Boeckx (ed.), *Minimalist essays*, 182–217. John Benjamins Publishing Company.

Sheehan, M., T. Biberauer, I. Roberts & A. Holmberg. 2017. *The Final-Over-Final Condition: A syntactic universal*. MIT Press. doi:10.7551/mitpress/8687.001.0001.

Svenonius, P. 1994a. C-selection as feature-checking. *Studia Linguistica* 48(2). 133–155. doi:10.1111/j.1467-9582.1994.tb00853.x.

Svenonius, P. 1994b. *Dependent nexus: Subordinate predication structures in English and the Scandinavian languages*: University of California, Santa Cruz dissertation.

Travis, L. D. 2005. Lexical, functional, crossover, and multifunctional categories. In H. Cohen & C. Lefebvre (eds.), *Handbook of categorization in cognitive science*, 319–346. Elsevier. doi:10.1016/B978-008044612-7/50069-X.

Van Urk, C. 2020. Successive cyclicity and the syntax of long-distance dependencies. *Annual Review of Linguistics* 6(1). 111–130. doi:10.1146/annurev-linguistics-011718-012318.

Zeijlstra, H. 2014. On the uninterpretability of interpretable features. In P. Kosta, S. L. Franks, T. Radeva-Bork & L. Schürcks (eds.), *Language faculty and beyond*, vol. 11, 109–128. John Benjamins Publishing Company. doi:10.1075/lfab.11.05zei.

Zeijlstra, H. 2020. Labeling, selection, and feature checking. In P. W. Smith, J. Mursell & K. Hartmann (eds.), *Agree to Agree: Agreement in the minimalist programme*, 31–70. Language Science Press. doi:10.5281/ZENODO.3541745.

Zeller, J. 2001. *Particle verbs and local domains*. John Benjamins Publishing Company. doi:10.1075/la.41.

Evaluating Labeling in Minimalist Syntax

Dallas Linn
University of Cambridge
dallaslinn@outlook.com