

Labelling in Syntax*

RICHARD STOCKWELL
UNIVERSITY OF CAMBRIDGE

1 INTRODUCTION

This paper considers syntactic labelling within the framework of the Minimalist Program (MP) (Chomsky 1995b). The MP pursues the Strong Minimalist Thesis (SMT) that “Language is an optimal solution to legibility conditions” (Chomsky 2000: 96). Labels were integral to the theories from which the MP arose. Both the Phrase Structure Rules (Chomsky 1975) of transformational generative grammar and X-bar theory (Chomsky 1970, Jackendoff 1977) in the Government and Binding era (Chomsky 1981) were formulated in terms of labels. However, in the MP, in view of SMT, some theorists (e.g. Collins 2002, Seely 2006, Narita 2011, 2014) claim that labels are non-optimal and should therefore be eliminated. I defend labels against such claims, showing instead that they are part of language’s optimal solution to legibility conditions. Nonetheless, prominent current approaches to labelling are unworkable, leading me to propose an alternative.

In outline, I begin by defending the need for labels in Section 2. I argue that headedness information is required by both the Phonological Form (PF) and Logical Form (LF) interfaces, and so would be most efficiently established in narrow syntax. Since a label-free syntax (Narita 2011, 2014) cannot provide the interfaces with headedness information, I pursue labelling as the means for doing so. However, existing approaches to labelling are highly problematic. I consider two approaches in detail in Section 3. First, Chomsky’s 2005, 2008, 2013, 2014 minimal search Labelling Algorithm fails to overcome {XP, YP} ambiguities, either through movement or feature sharing. Second, I consider selectional asymmetries in Merge (e.g. Pesetsky & Torrego 2006). I argue that it is conceivable for labels to play a role in selection from a crash-proof perspective (Frampton & Gutmann 2002), which is conceptually preferable to its free-merge alternative (Boeckx 2010). However, this role must fall short of stipulating that the selector always projects the label (e.g. Collins 2002), because this does not occur in free relatives. Existing analyses of free relatives (Donati 2006, Cecchetto & Donati 2010) fail to derive their characteristics, suggesting that the interpretive difference between free relatives and indirect questions reduces to labelling optionality. In addition, I argue in Section 4 that neither a Labelling Algorithm nor selectional asymmetries can account for adjunction, where {XP, YP} ambiguities are rife and there is no selection. Instead, I offer

* This paper is largely modelled on my undergraduate dissertation, Stockwell (2014a). My thanks go to David Willis for supervising, Theresa Biberauer for additional comments, and András Bárány for formatting.

an approach to labelling in Section 5 that emphasises the role of parallel derivations during language acquisition in overcoming {XP, YP} ambiguities. Section 6 concludes.

2 THE NEED FOR LABELS

This section defends labels as satisfying SMT. Section 2.1 argues that headedness information is required by both interfaces. Rather than being computed separately at each interface, it is therefore efficient for headedness information to be established just once in narrow syntax. Section 2.2 finds Narita’s 2011, 2014 proposal that this can be achieved in a label-free syntax to be seriously flawed, leaving me to consider labelling in syntax for the rest of this paper.

2.1 Computational efficiency

By SMT, interface requirements determine what is present in narrow syntax, and the interfaces require labels. Chomsky (1995a: 396) asserts that Merge must generate labels due to output conditions: “thus verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component.” Chomsky (2013: 43) maintains this position: “For a syntactic object SO to be interpreted, some information is necessary about it: what kind of object is it? Labeling is the process of providing that information.”

Regarding LF, the distinction between participants and events, encoded by nominal and verbal heads, is uncontroversial. By contrast, the role of headedness at PF is contested. Indeed, part of Collins’ (2002: 59f.) argument for eliminating labels from narrow syntax is that they are not relevant at PF. It would then suffice to establish headedness at LF. However, since the head directionality parameter (Chomsky 1981), theories of linearization have made reference to labels (see Narita 2014: 142–3 for discussion and references). In addition, research in phonology shows that PF does require headedness information about phrasal XPs, and more specifically about X categories.

In prosodic phonology, X-bar theoretic maximal XP projections (Jackendoff 1977) have a role in shaping phonological phrases (Selkirk 2011: 453). Nespor & Vogel (1986: 168) emphasise the non-isomorphism of prosodic and syntactic constituents, but make reference to the syntactic notions of phrase and head, particularly in the formation of the phonological phrase, φ . Maximal projections are likewise referenced in Selkirk’s (1986, 1995a) cross-linguistic theory of edge alignment between phonological phrases and syntactic XPs: $\text{Align}(\text{XP}, \varphi)$.¹

Moreover, there is evidence that PF requires more detailed information about XPs than just their head/phrasal status. Mappings in prosodic phonology concern only lexical elements as distinct from functional ones, as recognised in Selkirk’s (1995a) Lexical Category Condition. Some phonological rules refer to specific syntactic categories (Nespor & Vogel 1986: 32): Unrounded First Vowel Deletion in Greek (Kaisse 1977) occurs only within NP, while Verb Final Vowel Deletion in Italian

¹ cf. Truckenbrodt’s (1999: 228) WRAP-XP.

(Vogel, Drigo, Moser & Zannier 1983) applies specifically to verbs. Focal stress is triggered by specific syntactic heads in many theories of focus: first Jackendoff (1972: ch. 6), and subsequently proposals for F-marking (Selkirk 1984, 1995b) and a Focus head in an articulated CP (Rizzi 1997).

Therefore, PF requires finely detailed phrasal and headedness information. Since the same information is uncontroversially required by LF, this precipitates a conceptual argument for labelling in narrow syntax in terms of the third factor principle (Chomsky 2005) of efficient computation. Since headedness information is relevant at both interfaces, it is more efficient to establish it once, in narrow syntax, rather than twice, independently at each interface (*contra* Seely 2006: 189ff.).

2.2 Endocentricity without labels?

Narita (2011, 2014) argues that narrow syntax can supply the interfaces with headedness information without recourse to labels. Narita (2011: 15) motivates his proposal as taking up Chomsky’s (2007: 23) concern that “reference to labels ... is a departure from SMT”. Still, Narita (2011: 8) remains committed to endocentricity as a “basic fact” about phrase structure in human language, following Lyons (1968) and the insights of X-bar theory (Chomsky 1970, Jackendoff 1977). However, he proposes that endocentricity is established as part of Transfer to the interfaces rather than by labels. A part of Transfer is Minimal Head Detection (MHD) (Chomsky, p.c. to Narita and MIT lectures, Fall 2010), where SO is a syntactic object and LI a lexical item (Narita 2011: 122):

- (1) MHD: The head of an SO Σ is the most prominent LI within Σ .²

MHD motivates, and aligns with, an H- α schema for Merge, where H, a head, is an LI, and α a phrase (Narita 2011: 82):

- (2) Merge(H, α) \rightarrow {H, α }
Merge must take at least one LI as its input.

The H- α schema provides MHD with unambiguously headed phrases. Narita (2014: 77) reformulates the H- α schema as a constraint:

- (3) Endocentric Structuring Constraint (ESC):
Merge can only generate SOs whose head LI H is immediately detectable via minimal search at Interpret/Spell-Out.

Thus the merger of two phrases is rendered ungrammatical; i.e. *{XP, YP} (Narita 2014: 76):

- (4) Ban on Exocentricity:
Exocentric (non-endocentric) SOs of the form {XP, YP} are ruled out by [Full Interpretation].

² Narita (2014: 71) minimally rewords MHD: “For each SO Σ , define the most prominent LI within Σ as the head of Σ .”

To maintain (1)–(4), Narita (2014: 79) argues that all apparent instances of {XP, YP} must involve a phase head, which is returned to the derivation as an atomic element by Transfer of its complement.³

- (5) If Transfer eliminates an SO α , the phrasal SO {X, α } within the workspace of [narrow syntax] is reduced to X.

However, Narita’s thesis is challenged conceptually by the recombination problem (Boeckx & Grohmann 2007), and empirically by the interpretation of adjuncts.

The recombination problem concerns how transferred complements are reunited with their phase heads at the interfaces, for interpretation at LF and linearization at PF. Narita (2011: 178, 2014: 78) stipulates that despite the stripping away at Transfer, the interfaces interpret the phase interior as related to the phase head. However, this places a massive working memory burden on the interfaces (Boeckx & Grohmann 2007: 209): each interface must independently keep track of the relations between remerged phase heads and their spelled out complements. Narita (2011) does not consider the recombination problem. Subsequently, Narita (2012: 106)⁴ recognises that “cyclic Transfer will presumably need to integrate (‘recombine’) separately Transferred bits of structures in order to achieve the full-fledged compositional interpretation at SEM and PHON.” He claims that the problem will not arise if Transfer “let[s] the phase-head X stand as the most prominent element in place of {X, YP}” (Narita 2012: 105). That is, YP remains combined with X throughout the rest of the derivation, while the entire phase {X, YP} is stipulated to behave as an LI for further computation. This resort to stipulation to address the recombination problem severely weakens the conceptual elegance of Narita’s H- α schema.

Empirically, Narita’s proposal leads to a distinctly odd conclusion regarding the interpretation of adjuncts. By MHD, an LI head will always be more prominent than a phrase for minimal search. Therefore, when an adjunct LI merges with a phrase, the adjunct is interpreted as the head of the phrase. For example, (6) would be headed by the adjective *valuable*:

- (6) {valuable, {vase}}

Narita (2011: 214f., 2014: 129) claims that “we just have to swallow this conclusion.”⁵ Yet *valuable vase* is surely just as noun-y as *vase*, empirically undermining Narita’s theory.

Overall, Narita’s (2011, 2012, 2014) case for endocentric interpretation in label-free syntax is unconvincing. Narita stipulates his way out of the recombination problem, while he is also forced to say that adjunct LIs are interpreted as heads. Since Section 2.1 established that both interfaces require headedness information,

³ cf. Narita (2011: 83).

⁴ cf. Narita (2014: 37, fn. 20).

⁵ Narita (2011: 215, 2014: 129) defers to Chomsky’s (2004) Pair Merge as an alternative to digesting his conclusion, but see Section 4.1 for criticism.

in the next section I move on to examine two proposals for encoding it in syntax through labelling.

3 LABELLING FAILURES

Having established that headedness is most efficiently established in narrow syntax, and that Narita’s label-free theory fails to do so, I will instead pursue labelling in syntax as the means for providing the interfaces with headedness information. In this section I consider two approaches to labelling: Chomsky’s minimal search algorithm (Section 3.1) and selectional asymmetries in Merge (Section 3.2). However, neither proposal is persuasive, and when added to the problem of adjuncts (Section 4), these labelling failures lead to the alternative proposal in Section 5. But first I will scrutinise Chomsky’s (2005, 2008, 2013, 2014) minimal search Labelling Algorithm, and the {XP, YP} ambiguities it fails to overcome.

3.1 Chomsky’s Labelling Algorithm: {XP, YP} ambiguities

Chomsky (2005: 14) introduces the notion of a “natural algorithm” for determining the label of an SO. Appropriating minimal search, a third factor principle of efficient computation (Chomsky 2005: 6),⁶ Chomsky (2008: 145) sets out a “simple algorithm” for labelling:

- (7) In {H, α }, H an LI, H is the label.
 If α is internally merged to β forming $\{\alpha, \beta\}$ then the label of β is the label of $\{\alpha, \beta\}$.

Chomsky (2008: 145) claims that “These principles suffice for virtually every case.” This is certainly true for SOs of the form {H, XP}, where minimal search identifies H, an LI, as the label. However, SOs of the form {XP, YP} are ambiguous for minimal search (Chomsky 2007: 23).⁷ Chomsky (2013: 43) proposes a “fixed labeling algorithm LA”, and develops two ideas for how to resolve an {XP, YP} ambiguity: either the SO is modified by movement, or X and Y share features in common. I consider each in turn.⁸

3.1.1 Modifying {XP, YP}

Chomsky (2013: 43ff.) argues that {XP, YP} SOs can be labelled after they are modified by movement.⁹ As Chomsky acknowledges, this proposal derives from Moro’s (2000) theory of Dynamic Antisymmetry (DA). By DA, the PF requirement for linearization triggers movement, neutralizing a point of symmetry in the geometry of phrase structure. Chomsky recasts the effects of DA as triggered by labelling

⁶ See already Chomsky (2004: 105f.).

⁷ In addition, we will see in Section 3.2.3 that free relatives are a counterexample to the claim in (7) that the probe always projects over the goal in Internal Merge.

⁸ For empirical applications of Chomsky’s (2013) proposal, see, e.g., Epstein, Kitahara & Seely (2014, 2015), Saito (2013), and Schneider-Zioga (2014).

⁹ An idea mooted by Chomsky (2008: 145, fn. 34).

ambiguities at {XP, YP} points of symmetry. Chomsky (2013: 44) considers {DP, v*P}, the merger of the external argument with v*. He argues that the reason for EPP raising of DP to Spec-T is to allow the structure {<DP>, v*P} to be labelled v*, where <DP>¹⁰ has been moved.¹¹

However, Chomsky's analysis relies entirely on the notion of trace invisibility. In DA, "The crucial observation is that traces are not visible to the PF component" (Moro 2000: 2). Chomsky (2013: 44)¹² claims further that trace invisibility is an "intuitive idea". There are two major problems with this approach: the notion of a chain and the notion of traces as literal copies.¹³

First, trace invisibility rests on the notion of a chain. Chomsky (2013: 44) claims that taking the whole chain to be the relevant SO is not a stipulation. He argues that for α to be in the domain of D every occurrence (OCC) (Chomsky 2001: 39f.) of α must be a term of D. On this view, movement creates a discontinuous element which is no longer wholly contained in its original set. However, contrary to Chomsky's claim, chains are a stipulation that violate the Inclusiveness Condition (IC) (Chomsky 1995b: 228), because they are not part of the features of an LI. Without chains, the notion of trace invisibility cannot be formulated. Chain formation would require LA to look in two directions: both into the {XP, YP} SO it is trying to label, and to the phase edge, to see if either XP or YP have been moved there. This would mean LA having eyes in the back of its head, which would compromise computational efficiency.

Second, by the Copy Theory of Movement (CTM) (Chomsky 1993), movement does not leave a trace in its original position, but a full copy of itself. This theory is motivated by Inclusiveness and by interface phenomena. Regarding Inclusiveness, CTM does not resort to traces and indices, which are not features of LIs and so would violate IC. Regarding the interfaces, at PF lower copies can have a phonological manifestation. This is the case in split spellout phenomena:

- (8) What hope of ~~finding survivors~~ could there be ~~what hope of finding survivors~~
(Radford 2004: 194)

Meanwhile at LF, CTM derives the ambiguity in the antecedent of the anaphor in 'picture-of' sentences, without resorting to an LF reconstruction operation:

- (9) John wondered [which picture of himself] [Bill saw ~~which picture of himself~~]
(adapted from Chomsky 1993: 37)

¹⁰ Chomsky (2007: 23; 2013: 44, fn. 36) wonders why it is DP that moves out of {DP, v*P} rather than v*P, but leaves the matter aside. Rizzi (2013: 7) answers this question: if v*P moves then the SO would be labelled DP, which would provide a nominal interpretation to the interfaces rather than the required verbal one.

¹¹ Chomsky (2013: 44f.) develops his analysis further in relation to successive cyclic movement.

¹² Chomsky (2014: 5) continues to assume trace invisibility.

¹³ See also Narita (2014: 215) on the incompatibility of trace invisibility and head movement, if head movement is syntactic.

In sum, since traces are literally copies by CTM, it is difficult to see how they can be ignored by LA.

The conceptual problems with the notion of a chain, combined with conceptual and empirical support for CTM, seriously question trace invisibility and, consequently, the theory of DA. {XP, YP} labelling ambiguities cannot be resolved by movement, because moving XP or YP does not in fact modify the original SO {XP, YP}. Next, I show that an appeal to shared features fares no better.

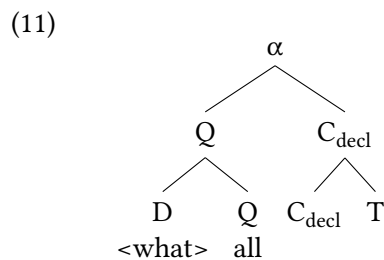
3.1.2 X and Y share features

Chomsky (2013: 43) claims that a second solution to the labelling problem for {XP, YP} SOs is for X and Y to share features. LA can then select these features as the label.¹⁴ Chomsky (2013: 45) invokes this strategy in view of {DP, TP}. This SO is the result of EPP-raising of the external argument out of v*P. The ϕ -features of D and T are identical, following the Agree relation in which T probed the goal DP to value its ϕ -features. Chomsky argues that these identical ϕ -features could serve as the label of the SO: $\langle \phi, \phi \rangle$ (Chomsky 2014: 6).¹⁵ Rizzi (2013, 2015) develops this idea in terms of his theory of criterial freezing (Rizzi 2006, 2010). {XP, YP} is a criterial configuration when both phrases have the same most prominent feature, such as [Q] or [Person], which then serves as the label. However, this analysis faces an empirical challenge from floating quantifiers, in addition to a host of conceptual issues: computational efficiency, interface interpretation, and feature interpretability.

Rizzi (2013: 14, 2015: ch. 7) acknowledges the problem raised by floating quantifier data from West Ulster English (McCloskey 2000: 61):

- (10) *What* did he say *all* (that) he wanted *t*?

Rizzi's analysis predicts a labelling ambiguity in (8) if the quantifier *all* is in Spec-C, because Q and C_{decl} do not share a criterial feature to serve as the label of $\alpha = \{Q, C_{decl}\}$, where $\langle X \rangle$ is a moved element:¹⁶



¹⁴ This marks a revision of Chomsky's previous (1995a: 397) position that the intersection of the two merged elements could not provide the label.

¹⁵ From this, Chomsky (2014: 3) concludes that all Internal Merge is exocentric. This implies a restrictive definition of endocentricity as applying only to heads. However, as Stockwell (2014b) notes, endocentricity remains at the level of features: the label of an SO is still sourced internally to that SO. See also Narita's (2011: 217) term "bifurcated endocentricity".

¹⁶ Rizzi (2013, 2015) assumes Chomsky's modifying by movement proposal outlined in Section 3.1.1 as the trigger for successive cyclic movement, along with trace invisibility.

Rizzi speculates that floated quantifiers must vacate the position where they were stranded, and move to an adverbial position in the low IP space. However, *that*, a reliable C position diagnostic, surfaces optionally after *all*, providing overwhelming evidence that *all* is in Spec-C, not somewhere in the IP space. With *all* in Spec-C, Rizzi's analysis predicts a labelling ambiguity in (8), contrary to the evidence.

Regarding computational efficiency, Chomsky (2013: 46) critiques his own analysis by expressing concern that labelling by feature sharing is a worrisome complication of LA. LIs are usually considered computational atoms, but LA would have to be able to search for features internal to LIs. Given that other operations, such as Agree, apply to features, this in itself need not be a cause for concern. However, LA has to retread the path of Agree, searching the features internal to both SOs to ascertain whether they are identical. Redoing operations is highly questionable with respect to computational efficiency.

In addition, it is unclear how labelling by feature sharing is useful for interface interpretation. As Stockwell (2014b) notes, labelling the traditional tense phrase {DP, TP} with nominal φ -features is interpretively very odd. Considering LA is meant to be motivated by interface requirements, by SMT, the label $\langle\varphi, \varphi\rangle$ does not seem helpful.

A final conceptual issue is feature interpretability, as addressed by Narita (2012: 107ff., 2014: 217ff.). By the principle of Full Interpretation (FI) (Chomsky 1986, 1995b), uninterpretable features must be deleted before reaching the interfaces. If, *pace* Pesetsky & Torrego (2007), uninterpretable features remain uninterpretable after they are valued, then the uninterpretable φ -features on T should not form part of the label of {DP, TP}. If they did, they would reach the interfaces and cause the derivation to crash.

Overall, labelling {XP, YP} structures by feature sharing faces an empirical problem from floated quantifiers, as well as conceptual issues regarding inefficient computation, interface interpretation, and feature interpretability. Since movement of one of the phrases out of the SO is likewise problematic, there is no satisfactory proposal for how {XP, YP} can be labelled by Chomsky's minimal search LA. In the next section, I investigate whether labelling by selectional asymmetries in Merge is any more successful.

3.2 Labelling by selectional asymmetries in Merge

This section examines whether selectional asymmetries could play a role in labelling. In Section 3.2.1 I introduce the framework of crash-proof syntax, which I argue is conceptually preferable to its free-merge alternative. From a crash-proof perspective, labels may have a role in constraining selection in the derivation, a possibility I examine in Section 3.2.2. However, it should not be stipulated that the selector always projects, since the selector does not project in free relatives (Section 3.2.3). Furthermore, as I show in Section 4, such an analysis could not extend to adjuncts, where there is no selection. This leads me to introduce an alternative analysis for labelling in Section 5.

3.2.1 *Crash-proof versus free-merge*

Crash-proof syntax involves a highly constrained derivation which “generates only objects that are well-formed and satisfy conditions imposed by the interface systems” (Frampton & Gutmann 2002: 90). From a crash-proof perspective, it is conceivable that labels could play a selectional role in syntax. This would be inconceivable from a free-merge perspective, whereby selection has no place in narrow syntax. Before investigating the role that labels may play in selection, I will therefore establish that crash-proof syntax is conceptually preferable to the free-merge alternative.

Boeckx (2010) criticises crash-proof syntax for simply re-encoding the filters of the Government and Binding (GB) era as features. Crash-proof syntax reconceptualises GB filters as properties of the interface systems, reducing them to a single generalised filter of interpretability, FI. Uninterpretable features must be checked and deleted before reaching the interfaces, otherwise the derivation will crash. The alternative is a free-merge approach (Boeckx 2010). In the free-merge framework, features are reduced to a generic edge feature permitting Merge (Chomsky 2007, 2008), and the external interface systems retroactively interpret derivations.¹⁷ Thus output filters, such as the Theta Criterion (Chomsky 1981), can be discarded, with theta-roles determined configurationally (Hale & Keyser 1993) rather than derivationally. As for labelling, LA, along with other operations, applies at the phase level as part of Transfer (Chomsky 2013: 43, 2014: 4), with the outcome interpreted at the interfaces. In sum, free-merge opposes “the lingering idea, carried over from earlier work, that each operation has to be motivated by satisfying some demand” (Chomsky 2014: 11). On the free-merge approach, therefore, labels can have no role internal to the derivation, ruling out a role for labels in selection.

I offer three conceptual arguments in favour of the crash-proof approach, and against the free-merge approach. First, the free-merge approach merely relocates complexity at the interfaces rather than eliminating it. Second, crash-proof syntax has better prospects for future investigation of the interfaces. Third, the free-merge approach is computationally inefficient, a criticism which is not countered by an appeal to phases.

First, the interfaces are not a better location for complexity than narrow syntax. In Ott’s (2010: 99) words, free-merge pursues “the methodological desideratum of attributing as little structure to UG as possible, while relegating as much complexity as possible to the interfacing systems”. I do not share this conception of minimalism as shunting complexity from narrow syntax to the interfaces. Rather, a minimalist disposition should aim to eliminate, rather than relocate, complexity.

Second, the free-merge approach does not, and cannot, provide a theory of the interfaces. As Ott (2010: 100) himself notes, the free-merge approach does not offer an account of the “interface conditions, most of which remain to be identified.” Instead, the crash-proof perspective is necessary to reach an understanding of the interfaces (Frampton & Gutmann 2002: 103). This conclusion accords with

¹⁷ See Ott (2010: 91ff.) on the distinction between acceptability and grammaticality in this regard.

Chomsky's (2000: 98) observation that progress in understanding syntax and the interfaces must proceed hand-in-hand.

Third, the free-merge approach entails apocalyptic computational inefficiency, to the extent of infinite ill-formed derivations. A derivation could be condemned to crash due to a problem early on, but this would not be noticed until Transfer at the end of the derivation. Frampton & Gutmann (2002: 94) illustrate this point with reference to the example of the derivation in (12), where External Merge of *it* has consigned *Max* to caselessness:

(12) it to be believed Max to be happy

Thus, “[d]erivations can continue from [(12)] and proceed indefinitely, futilely, never detecting the fatal flaw already present in [(12)], until the question of convergence arises at the interface.” The free-merge approach means giving up on SMT, since allowing crashing derivations to pass to the interfaces is a sub-optimal way of meeting the conditions they impose.

In defence of free-merge, Boeckx (2010: 108) follows Chomsky (2000) in invoking “cyclic computation (phase-based), which addresses computational complexity concerns”. Likewise, Ott (2010: 101) argues that “With cycles reduced to phases, deviance of an Exp[ression] will be detected quickly, upon TRANSFER of the phase.” However, this attempted defence of free-merge in terms of phases does not stand for two reasons: first, there is nothing that forces a phase head to be merged; second, merging a phase head would not even solve the problem.

First, let us assume that merging a phase head could resolve the problem of noticing deviance. If Merge is entirely free, there is nothing that could force a phase head to be merged. This means that a derivation could continue through an infinite number of derivational steps without ever being transferred to the interfaces. Rather than ensuring that a fatal problem is spotted reasonably early, a problem may never be spotted at all. Nor is this issue resolved by positing a Numeration (Chomsky 1995b) that a derivation must exhaust. A Numeration merely shunts the issue to Select, which can then be accused of severe look ahead, presaging the course of the entire derivation.¹⁸ Therefore, even if merging a phase head could reduce the computational burden of noticing deviance, there is nothing in the free-merge approach that requires this to happen.

Second, even merging a phase head would never necessarily resolve the issue. The specifier of the phase head is not transferred to the interfaces in cyclic spell out; only its complement is. It seems that on the free-merge approach Internal Merge can operate as ‘interface bounceback’. The edge feature allows any SO with an unvalued feature that is rejected by the interfaces to be merged as a specifier of the phase head. This would allow any deviant or crashing element to escape from the phase, allowing the derivation to continue through infinitely many phases without

¹⁸ Chomsky (2013: 41) suggests that he is moving away from positing a Numeration in his definition of Merge, which “must access the workspace of already generated objects and *the lexicon*” (emphasis mine).

the deviance being noticed. Far from identifying deviance early on, phases could in fact prolong the agony.

Thus, the phase-based defence of free-merge against computational inefficiency fails. There is nothing forcing a phase head to be merged, and even then this would not prevent deviant syntactic objects avoiding transfer via the phase head’s specifier. Such unconstrained generation of deviant derivations would severely reduce computational efficiency.

Crash-proof syntax is therefore conceptually preferable to the free-merge approach. With this established, it is now appropriate to investigate the role that labels play in ensuring that syntax is crash-proof.

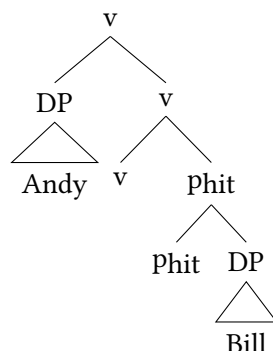
3.2.2 Labels in selection

From a crash-proof perspective, labels may help to ensure that the derivation converges by playing a role in selection.¹⁹ In this section, I argue that acategorial roots (Marantz 1997) need not undermine this view. However, the case of free relatives in Section 3.2.3 will show that it is too far to stipulate that labels are determined by the selectional asymmetry between selector and selectee.

Frampton & Gutmann (2000) conceive of labels as having a role in selection. A head is a pivot whose features must be satisfied before a new cycle can begin. Assuming theta roles can be conceptualised as features, this approach recalls the insight of Relativised X’-theory (Fukui & Speas 1986) that lexical categories project as they discharge their theta-features, allowing them to contribute to further computation.

However, Chomsky (2004: 112) argues that selection can play no role in narrow syntax, concluding that “derivations cannot be failure-proof (“crash-free”).” He arrives at this conclusion due to concerns over acategorial roots (Marantz 1997) as proposed in the framework of Distributed Morphology (Halle & Marantz 1993). On this view, a root, such as *√hit*, only becomes verbal in the environment of the verbaliser, little *v*:

(13)

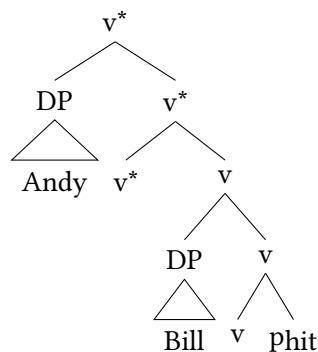


¹⁹ See also Cecchetto & Donati (2015: 29ff.) on the internal role of labels in selection as a core part of the computation, in addition to their external interpretive role at the interfaces.

For the root to select its internal argument in (13), it would have to do so before it receives verbal interpretation by merging with the functional head, *v*. However, if the functional nominaliser *n* was merged instead of *v* then such an internal argument would be entirely illicit. Selection in this derivation would entail very strong look ahead. This concern leads Chomsky (2004) to banish selection from narrow syntax.

However, the crash-proof selectional view can be made compatible with acategorical roots by separating out the verbaliser *v* from the head *v*^{*} (Chomsky 1995b)²⁰ that introduces the external argument:²¹

(14)



This analysis has the advantage of introducing arguments uniformly by functional heads. Moreover, it offers a satisfactory resolution to the First Merge problem, where two heads are merged at the start of the derivation. If roots are taken to have no syntactic features, then they could not provide the label under any theory of labelling, because they have nothing to contribute to further computation (Chomsky 2013: 47, 2014: 5). This maintains the generalisation that Merge is a binary operation between two unique elements, unlike other proposed solutions to the First Merge problem. Some of these involve binary Merge with identical inputs, such as Self Merge (Guimarães 2000) or singleton set formation (Kayne 2009). Others involve singular Merge, as in Unary Merge (Zwart 2004) or first Merge with the empty set (Fortuny 2008: 18f.). The analysis here is preferable, because it does not make a special case of First Merge.

It therefore seems that functional heads are involved in selection. This role is extended by some scholars to the stipulation that the selector always projects. This stipulation (cf. Pesetsky & Torrego 2006, Boeckx 2008: ch. 3, Chomsky 2008: 145,²² Panagiotidis 2014: ch. 5) is conveyed most clearly by Collins (2002: 55):

- (15) “Accessibility Condition:
A lexical item *X* (and the features it contains) is accessible without search to a syntactic operation *OP* if *X* contained the probe/selector for the last operation in the derivation.”

²⁰ Or Kratzer’s (1996) Voice head.

²¹ I assume that movement of {*v* *hit*} to *v*^{*} yields the correct linear order in (14).

²² See note 7.

However, this stipulation does not hold in the case of free relatives, to which we now turn.

3.2.3 Free relatives: where the selector does not project

Free relatives refute the stipulation that the selector projects. Free relatives (16) contrast minimally with indirect questions (17) in terms of projection:

(16) John likes what Mary likes.

(17) John wonders what Mary likes.

In both cases, C is the selector and *what* is the selectee. I assume that there is an uninterpretable and unvalued [wh]-feature on C, which causes C to Probe for a matching Goal in its c-command domain. The [wh]-element *what* is identified, allowing valuation of C's uninterpretable [wh]-feature by Agree. An [EPP] movement diacritic feature, which is parasitic on C's [wh]-feature, triggers movement of *what* to Spec-C. In (16), the selectee *what* projects, in accord with the selectional requirements of the verb *like*, and with the fact that free relatives are complex nominal structures. Thus free relatives are an example of the Project Goal option of Citko (2008). In (17), on the other hand, the selector C projects, in accord with the selectional requirements of the verb *wonder*. Thus the embedded clause is interpreted as an interrogative clause.

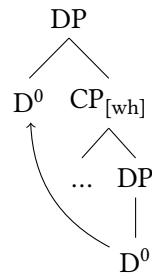
(18) John [_{VP} likes [_{DP} [_D what] [_{CP} [_C \emptyset] Mary likes ~~what~~]]]

(19) John [_{VP} wonders [_{CP} [_D what] [_C \emptyset] Mary likes ~~what~~]]

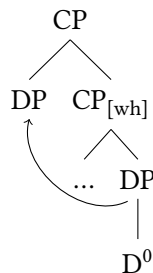
Attempts have been made to analyse this labelling phenomenon in terms of phrase structural ambiguity (Donati 2006) and probing asymmetries in Merge (Cecchetto & Donati 2010). I take these analyses in turn (Sections (19) and (30)), demonstrating that both fail. Furthermore, I show that the C of free relatives and indirect questions is featurally identical (Section (31)). This forces the conclusion that in free relatives labelling shows true optionality. Therefore, just as the minimal search Labelling Algorithm was shown to fail in Section 3.1, selectional asymmetries in Merge are also unable to determine labels unambiguously.

Phrase structural ambiguity Donati (2006) sets out to distinguish free relatives from indirect questions in terms of contrasting phrase structural status. Heads and phrases retain their phrase structure status after movement by the Condition on Uniformity of Chains (CUC) (Chomsky 1995b: 253): heads project, whereas phrases are projections. Regarding free relatives and indirect questions, both involve wh-movement triggered by a [wh]-feature on C. In free relatives (20), the D head then projects a DP. In standard wh-constructions (21), by contrast, phrasal movement is required to preserve the CP categorial status of the clause (cf. Donati 2006: 33):

(20) Free relative:



(21) Interrogative:



This analysis requires that free relatives involve movement only of D, not DP. Donati (2006: 32) provides evidence from Italian and English (22)–(24) that phrasal pied-piping is disallowed in free relatives:

(22) *I shall visit [*what* town] you will visit [t](23) I wonder [*what* town] you will visit [t](24) I shall visit [*what*] you will visit [t]

However, (25) suggests that free relatives can involve movement of DP, not just D. Donati (2006: 32, fn. 10) claims that (25) is not a free relative, citing Bianchi (1999) and Kayne (1994) in analysing it as (26):

(25) I shall visit [*whatever* town] you will visit [t](26) [DP [D *what*_i [D *ever*]] [CP [DP *t*_i town] [...]]]

Still, it would be very difficult to apply the analysis in (26) to data that do not include *whatever* (Kayne 1994: 154):²³

(27) We gave him what little money we had.

Moreover, as Citko (2008: 930) notes, the analysis in (26) entails that free relatives involving more than just a D head must be headed relative clauses. However, evidence from case matching in Polish, extraposition in German, and compatibility with overt complementizers, presented here, show that simple free relatives (28) pattern together with *whatever*-relatives (29), to the exclusion of headed relatives (30) (adapted from Citko 2008: 931):

²³ See Chomsky (2013: 46f.) for further difficult examples for Donati's (2006) analysis.

- (28) Buy what (*that) you like.
(29) Buy whatever thing (*that) you like.
(30) Buy the thing (that) you like.

In sum, the standard assumption that only the Probe can project in Internal Merge rests on the fact that the Probe is the selector. However, it seems unjustified to stipulate that the selector projects, since the selected Goal projects in free relatives. Donati's (2006) analysis does not derive this option, which suggests that the interpretive difference between free relatives and indirect questions reduces to labelling optionality.

Probing asymmetries in Merge Cecchetto & Donati (2010)²⁴ revise the analysis of Donati (2006) on conceptual grounds, attempting a more minimal analysis that does without phrase structure theory and the CUC.²⁵ Instead, they analyse the ambiguous labelling possibilities of free relatives in terms of a Probing Algorithm (PA) (2010: 245):²⁶

- (31) “Probing Algorithm:
The label of a syntactic object $\{\alpha, \beta\}$ is the feature(s) that act(s) as a Probe of the merging operation creating $\{\alpha, \beta\}$.”

Their analysis rests conceptually on a probe-goal asymmetry introduced by the Edge Feature (EF). Chomsky (2008: 139,144; 2007: 11) proposes that every LI has an EF, which permits it to enter into a computation by being merged with some SO. EF is conceptually problematic in two ways and Cecchetto & Donati's (2010) analysis is fundamentally flawed in any case.

First, Chomsky's (2007: 11) claim that “The property of unbounded Merge reduces to the statement that LIs have EF” is unfounded. As Fukui (2011: 85) notes, whereas features usually distinguish between classes of LIs, EF is associated with every LI. Being a potential input to Merge is an intrinsic property of an LI and, further, an SO. Rather than reducing Merge to EF, EF should be reduced to the definition of LI.

Second, EF is unique among lexical features in not being deleted after it has probed (Fukui 2011: 85f.). EF is not deleted in narrow syntax, since the same LI can be internally merged multiple times in the same derivation. Yet, as an uninterpretable feature (uF), EF must be deleted before reaching the interfaces, by FI. Chomsky (2007: 11, fn. 16) therefore concedes that EF is not a standard uF, speculating that it may be deleted as part of Transfer. EF's uniqueness gives further reason to suspect that it is superfluous to endow every LI with EF.

²⁴ See also Donati & Cecchetto (2011).

²⁵ See in particular Cecchetto & Donati (2010: 267, fn. 22).

²⁶ This definition is retained by Cecchetto & Donati (2015: 2); cf. Adger (2003: 91).

Even putting these conceptual issues with EF aside,²⁷ Cecchetto and Donati's analysis is fundamentally flawed. Cecchetto & Donati (2010: 246) extend the concept of EF to a probe-goal asymmetry: "any time an LI is merged, it qualifies as a Probe by virtue of its EF." This is claimed to derive the fact that the head projects when merged with a phrase. The head is an LI with EF, which probes the phrase. The phrase does not have EF, so does not probe the head. Thus, by PA, the head projects.

Applied to free relatives and indirect questions, Cecchetto & Donati (2010) argue that the distinct structures arise from a labelling conflict. When the wh-element is an LI, it is a probe by virtue of its EF. C is also a probe, by virtue of its wh-uF. Therefore the wh-LI and C mutually probe one another. This ambiguity means that either the wh-LI or C can provide the label by PA, resulting in a free relative or a question. There is no ambiguity when phrasal movement is involved. The wh-phrase is not a probe because it does not have EF; only C is a probe by virtue of its wh-uF. C unambiguously provides the label, resulting in a question.

Cecchetto & Donati's (2010) analysis thus retains the prediction from Donati (2006) that free relatives may only involve wh-heads, contrary to the data in (25)–(30).²⁸ Moreover, their analysis of labelling conflict disintegrates upon closer examination. Cecchetto & Donati (2010: 246) "propose that the EF of a word is to be identified with its categorial feature", and further that the "categorial feature can provide the label." The EF is the categorial feature, which is in turn the label. A phrase would have EF at least in, if not as, its label. EF does not delete, because IM is possible. So there is an active EF in the label of a phrase. Thus, both a head and a phrase have EF, in direct contrast to Cecchetto & Donati's (2010) reasoning. By the PA, they mutually probe one another, causing a labelling conflict. As such, there is no asymmetry inherent to head-phrase merger. Therefore C would be doubly a probe, due to its being a phrase CP and having a wh-uF. This predicts that only C can provide the label, regardless of whether it has probed for a head or a phrase. This prevents D from projecting in free relatives, contrary to what is required for selection and interpretation. Thus, Cecchetto & Donati (2010) fail to derive labelling in free relatives. This again suggests that the interpretive difference between free relatives and indirect questions reduces to labelling optionality.

The same C The difference between free relatives and indirect questions cannot be attributed to the features of C, since the structural relationship between the wh-phrase and C is the same.²⁹ Free relatives (32)–(33) and indirect questions (34)–(35)

²⁷ Cecchetto & Donati (2015: 33) indeed abandon edge features as "infelicitous". However, the problems of the following paragraphs remain, because Cecchetto & Donati (2015) revert in effect to the problematic analysis of Donati (2006) in claiming that words are special. Words intrinsically act as Probes on account of their wordhood, always activating PA, and providing (at least) their categorial feature as the label. However, rather than appealing to contrasting phrase structural status, as did Donati (2006), Cecchetto & Donati (2015) support their claim through lexicalism: words have a special status as the outputs of the morphology module, and the inputs to syntax.

²⁸ This issue also remains unsatisfactorily accounted for in Cecchetto & Donati (2015: 51ff.).

²⁹ cf. Ott (2011: 186), who argues that C is featurally distinct in free relatives and indirect questions. Free relative C bears no interpretable formal features, so can be transferred to the interfaces along

pattern together in not showing do-support, in contrast with direct questions (36)–(37):

- (32) *I will buy what did you buy.
 (33) I will buy what you bought.
 (34) *I wonder what book did you buy.³⁰
 (35) I wonder what book you bought.
 (36) What did you buy?
 (37) *What you bought?

These data suggest that C is featurally the same in free relatives and indirect questions, as opposed to the C in direct questions. A possible analysis is that a [Q] feature on C in direct questions triggers head movement from T to C, requiring do-insertion to support the tense and person features in C. By contrast, the embedded C in both free relatives and indirect questions has [wh] but not [Q]. Whatever the specifics of the featural analysis, the same C is involved in free relatives and indirect questions, which is different from the C in direct questions. Since C is featurally the same, the interpretive difference between free relatives and indirect questions reduces to labelling optionality: in free relatives, a DP label is projected; in indirect questions, a CP label is projected.

Overall, selectional asymmetries in Merge cannot be responsible for determining the label, contrary to the stipulation that the selector projects. In any case, such an analysis could never account for adjuncts, as discussed in the next section. This urges exploration of an alternative approach in Section 5.

4 LABELLING FAILURES FOR ADJUNCTS

The failed proposals for a theory of syntactic labelling in Section 3 limited themselves to the domain of argument structure. Adjuncts are crucially different from arguments in that they are not involved in selection. The optionality of adjuncts leads Chametzky (2000: 141f.) to the conclusion that a minimalist theory of adjuncts may be impossible, because minimalist theorising focuses only on what is necessary. Despite Chametzky's reticence, theories of adjunction have been proposed in the MP by Chomsky (1995a, et seq.) and Hornstein (2009), though both are unsatisfactory.³¹ The failure of general labelling theories, as well as those restricted to adjunction, leads to the alternative proposal sketched in Section 5.

with its complement at the end of its phase. This leaves behind only the wh-phrase, yielding the nominal distribution of free relatives. Since indirect question C bears an interpretable formal feature Q, by contrast, it cannot be transferred, but remains in the derivation to give indirect questions their clausal distribution. Such an analysis does not straightforwardly account for the data presented here.

³⁰ (34) is not grammatical as an indirect question in standard English. It is only possible with 'I wonder' marked off by intonation as introducing a direct question.

³¹ Recall also from Section 2.2 the odd prediction of Narita (2011: 214f., 2014: 129) regarding the labelling of adjunct LIs.

4.1 Chomsky's Pair Merge

Chomsky (1995a, et seq.) advocates a separate operation for adjunction. He argues that there are two types of pure Merge: Set Merge and Pair Merge (2000: 133). Set Merge, formerly substitution, takes two objects α and β and forms the set $\{\alpha, \beta\}$, labelled $\{\delta, \{\alpha, \beta\}\}$, where either $\delta = \alpha$ or $\delta = \beta$ (Chomsky 1995a: 396ff.). In Pair Merge, formerly adjunction, on the other hand, the head does not project; rather, the most recent projection of the head is replaced by a two segment category. The resulting SO is labelled with an ordered pair determined by the head: $\{<k, k>, \{\alpha, K\}\}$ (Chomsky 1995a: 402). This difference in labelling is required to show the inherent asymmetry of adjunction. Chomsky (2013: 45f.) continues to argue for a distinct Pair Merge operation.

Chomsky's treatment of adjunction is conceptually unappealing. It proliferates operations threefold by introducing Pair Merge, Simplify, and redundancies in other syntactic operations. First, Pair Merge introduces a second kind of structure-building operation where one would be optimal by SMT. Since both arguments and adjuncts have to be interpreted at the interfaces (Chomsky 2004: 118), ideally both should be captured by the same Merge operation. Second, Pair Merge requires a further Simplify operation SIMPL as part of TRANSFER (Chomsky 2004: 118). SIMPL allows the output of Pair Merge to be linearised, by converting it from an ordered to an unordered set. Thus, Chomsky is forced to posit a further operation to render Pair Merge, which is itself conceptually unsound, interpretable to PF. Third, Pair Merge introduces a redundancy in narrow syntactic operations (Hornstein 2009: 86). Movement, replacement and deletion operations target phrasal levels, optionally including or excluding adjuncts. Such operations must therefore target structures labelled both $H(K)$ by Set Merge, and $\langle k, \dots k_n \rangle$ by Pair Merge, rather than targeting just one type of label. Thus, where there would optimally be one structure-building and labelling operation by SMT, Chomsky's analysis of adjunction is uneconomical and conceptually unappealing because it introduces three operational complications.

4.2 Hornstein: dangling in a separate plane

Hornstein's (2009) account of adjunction is likewise unsatisfactory. It is framed in an analysis of Merge as a two-part operation, comprising Concatenate and Labelling.³² Hornstein (2009: ch. 3) stipulates that only LIs are concatenable. Labelling maps a concatenate to one of its elements, an LI, returning it to the domain of concatenable expressions. Thus the concatenate X^Y is labelled $[_X X^Y]$, mapping it to X – a concatenable LI. As for adjuncts, Hornstein (2009: ch. 4) argues that they

³² This analysis is motivated by a concern for evolutionary plausibility. Hornstein (2009: ch. 1) seeks a solution to Darwin's problem, or the logical problem of language evolution. If the human language faculty is only 50,000–100,000 years old, as is commonly assumed (Hornstein 2009: 4), then it must make use of predominantly general cognitive resources, with probably just one, or maximally two, innovations. Hornstein argues that Concatenate is a general cognitive operation that predates Language. Concatenate was rejected by Chametzky (2000: 127ff.) on the grounds that it generates flat strings rather than hierarchical structures. However, the innovation of endocentric Labelling, combined with Concatenate, means hierarchical structure can be generated.

do not have to be labelled; instead they can dangle off the structure. There are at least three problems with Hornstein's analysis of adjuncts: interface interpretation, the proliferation of dimensions, and a lack of empirical support.

First, it is unclear how unlabelled adjuncts would be interpreted at the interfaces. I established in Section 2.1 that both interfaces require headedness information for interpretation. In Section 2.2, I argued that such information could not be provided by a label-free syntax. If syntactic labelling is required for interface interpretation, it is unclear how adjuncts could be interpreted without labels.

Second, Hornstein's analysis proliferates dimensions. Hornstein (2009: 92, fn. 16) notes that his theory could be viewed as a development of Chomsky's (2004: 117f.) speculation that adjuncts inhabit another dimension: "we might intuitively think of α as attached to β on a separate plane, with β retaining all its properties on the 'primary plane', the simple structure" (cf. Irurtzun & Gallego 2007). If there is a separate plane for adjuncts, then phrase structure would have to be three-dimensional. This conflicts with the standard assumption that phrase structure is two-dimensional,³³ and an extra dimension would surely increase computational burden. Therefore, by SMT, extra dimensions should not be postulated unless there is strong empirical evidence that they are required by the interfaces.

Third, I do not believe that such strong empirical evidence exists. On Hornstein's intuitions adjuncts are not hierarchically structured in relation to one another. Adjuncts are only labelled where necessary; that is, where an operation treats them as part of the phrase they are modifying (Hornstein 2009: 89ff.). Hornstein claims that such operations, which include topicalisation, ellipsis and *do-so* anaphora, can target discontinuous objects. Illustrating with *do-so* anaphora, Hornstein (2009: 99) claims that *did (so)* replaces the linearly discontinuous *ate the cake ... with a fork* in (38):

- (38) John ate the cake in the yard with a fork in the afternoon ...
but Bill did (so) in the kitchen in the morning.

For Hornstein (2009: 99), Bill can be understood to have eaten the cake with a fork, because (38) is associated with the labelled structure in (39):

- (39) [V [V ate^the-cake] ^ with-a-fork]
 ^ in-the-yard
 ^ in-the-afternoon

According to Hornstein, therefore, adjuncts can be linearised in any order. Despite *with a fork* being incorporated into the verb phrase by labelling, *in the yard* can surface in between. However, I do not share Hornstein's intuition. In (36), I take *did (so)* to replace only *ate the cake*; it does not follow that Bill ate the cake with a fork. My intuition accords with an analysis of adjuncts as hierarchically structured. For me, (38) would be associated with the structure in (40):

³³ Indeed, Chomsky (2014: 4) retreats from his previous (Chomsky 2004: 117f.) speculation with an exhortation to "view with caution the resort to multidimensionality".

(40) [VP [VP [VP [VP ate the cake] in the yard] with a fork] in the afternoon]

For these three reasons of interface interpretation, the proliferation of dimensions, and empirical evidence, [Hornstein's \(2009\)](#) analysis, like Chomsky's, is entirely unappealing.

5 AN ALTERNATIVE: PARALLEL DERIVATIONS DURING ACQUISITION

The previous section showed that proposed theories of labelling for adjuncts are unsatisfactory. Furthermore, adjuncts could not be subsumed under the problematic theories of labelling in Section 3. Under a Chomskyan minimal search algorithm (Section 3.1), {XP, YP} ambiguities would be rife. Meanwhile, a theory of labelling by selectional asymmetries in Merge (Section 3.2) is untenable, as demonstrated by the analysis of free relatives in Section 3.2.3; but, in any case, such a theory could never extend to adjuncts, because they are not involved in selection.

In view of this, I sketch an alternative theory of syntactic labelling. I claim that labelling is acquired from the results of running parallel derivations (Section 5.1). I consider the labelling of free relatives and adjuncts in this light. In Section 5.2, I offer circumstantial support for my analysis.

5.1 *The role of acquisition*

In this subsection, I outline a course for the acquisition of labelling, before applying it to three cases considered above: {DP, v*P}, adjuncts, and free relatives.

I propose that when confronted with SOs of the form {XP, YP}, the child computes both possible labelling options in parallel, projecting X in one and Y in the other. Only one of the parallel derivations will converge. Since the same option will converge each time, the Primary Linguistic Data (PLD) will provide overwhelming evidence in favour of one of the labelling options. This knowledge is established at the interfaces through the application of domain-general, third factor ([Chomsky 2005](#)) principles of data analysis; that is, the strategy for dealing with {XP, YP} ambiguities need not be UG-given. The child then imposes this knowledge on the workings of narrow syntax.

I further hypothesise that these emergent instructions for which element provides the label are encoded in the features of LIs. In this, I follow the Borer-Chomsky Conjecture (BCC). The BCC was introduced by [Borer \(1984\)](#), adopted by [Chomsky \(1995b\)](#), and coined by [Baker \(2008: 353\)](#):

(41) “All parameters of variation are attributable to differences in the features of particular items (e.g. the functional heads) in the lexicon.”

More specifically, labelling information could be encoded by second order ([Adger & Svenonius 2011](#)) projection diacritic features, “>”, associated with categorial features. For example, consider {DP, v*P}. Faced with this SO in the early stages of acquisition, the child computes two parallel derivations. In one derivation, {DP, v*P} will be labelled D, in the other, v*. Only the latter derivation will converge, because T, the next item to be merged, selects v*, not D. Thus, the categorial feature

of v^* would come to be associated with a second order [$>D$] feature. This would cause v^* to project over D, meaning that only the convergent derivation would be computed in steady state grammars.

Labelling of adjuncts could be acquired in much the same way. When first faced with {AP, NP} SOs, the child computes parallel derivations, one where A projects and another where N projects. Only projecting N yields the relevant nominal interpretation at the interfaces, and meets the selectional requirements of D, which is next to be merged. This information is encoded as a second order [$>A$] feature associated with the categorial feature of N.

I argued in Section 3.2.3 that the distinction between free relatives and indirect questions reduces to labelling optionality. Both structures involve {DP, CP} SOs. On the approach here, the child computes parallel derivations when faced with {DP, CP} SOs. Only one labelling option converges in each context: DP for free relatives and CP for indirect questions. More generally, C projects over A' -moved APs and PPs, as well as DPs. This suggests that C comes to be associated with a generalised [$>X$] feature, whereby it projects over its specifier. I claim that this generalised feature is overridden by a specific [$>C$] feature which comes to be associated with free relative D. In this case, the relevant PLD are broader, encompassing a range of A' -movements, but the contextual variation in labelling is still acquirable.

Finally, the approach here has implications for the crash-proof framework. Syntax is not entirely crash-proof during acquisition, since only one of the parallel derivations converges. Instead, the data impose themselves on syntax over time, until it becomes crash-proof. Still, parallel derivations do not drastically reduce computational efficiency in child grammars. As I argued in Section 3.2.2, labels constrain selection under sisterhood. The wrong labelling option will therefore cause a failure of selection in the attempt to form the next syntactic object.

5.2 Circumstantial evidence

In the previous subsection I argued that labelling can be acquired from running parallel derivations. Here I offer circumstantial support for my proposal, comprising previous suggestions in the literature, the significance of binary Merge, and arguments from plausibility.

My proposal has antecedents in suggestions from Chomsky and Citko. Chomsky (2008: 145, fn. 34) notes that a possible way to label the external argument structure {DP, v^*P } “is that either label projects, but only v^* -labeling will yield a coherent argument structure at C-I”. Similarly, Citko (2008: 916) “assume[s] that labeling is essentially free... However, only a subset of possible choices will yield convergent derivations.”³⁴ My claim is that repeated (in)coherence/(non)convergence leads the child to encode the correct labelling option in their syntax.

Conceptually, this proposal further motivates binary Merge. There is strong empirical evidence for binary branching phrase structure (Kayne 1984). In MP, SMT urges the conceptual argument that the minimal operation required for recursion should also be the maximal (Hornstein, Nunes & Grohmann 2005: 209). In addi-

³⁴ See also Narita (2014: 80).

tion, the proposal here suggests that Merge may be restricted to improve the child's chances of resolving binary rather than n-ary labelling ambiguities.

Finally, I offer three arguments from plausibility in support of my approach to labelling: the speed of acquisition, the nature of the PLD, and the possibility for the data to impose themselves on narrow syntax.

First, children acquire knowledge of their language incredibly quickly. [Wexler \(2003\)](#) cites a wealth of evidence in favour of the thesis of Very Early Parameter Setting by around eighteen months. This suggests that the PLD could provide evidence for overcoming labelling ambiguities with similar speed and success to parameter setting.

Second, the nature of the PLD facilitates speedy acquisition. Child Directed Speech (CDS) is remarkably grammatical. [Broen \(1972\)](#) found that CDS shows nine times fewer false starts, mispronunciations, and hesitations than Adult Directed Speech (ADS). Moreover, sentences in CDS tend to be very short. [Phillips \(1973\)](#) found that the mean length of utterance was two words for CDS to two-year-olds, compared with eight words for ADS. Shorter utterances significantly restrict the search space for the child in establishing the outcomes of labelling ambiguities.

Third, the idea that the PLD can impose themselves on narrow syntactic labelling ambiguities gains plausibility from a parallel suggestion from Hornstein. [Hornstein \(2009: 160ff.\)](#) speculates that constructions, such as passive formation or raising of the subject, may be cognitively real epiphenomena. Where syntactic operations apply frequently in particular patterns, there could be an efficiency payoff in compiling them into a construction-specific format. Thus, cases of the PLD imposing themselves on narrow syntax plausibly extend beyond the resolution of labelling ambiguities.

6 CONCLUSION

This paper argued for syntactic labelling but against current theories, before sketching an alternative approach grounded in acquisition. Since both interfaces require headedness information, this information would be most efficiently established in narrow syntax. I pursued labels as the means of providing the interfaces with headedness information, because a label-free syntax could not. However, the Chomskyan minimal search LA fails to overcome {XP, YP} ambiguities, due to problems with modifying by movement and feature sharing. In addition, while it is conceivable that labels could play a role in selection from a crash-proof perspective, it should not be stipulated that the selector projects, because this does not happen in free relatives. Meanwhile, there is no satisfactory theory for labelling adjuncts, where {XP, YP} ambiguities are rife and there is no selection. Instead, I suggested that labelling ambiguities could be overcome by computing parallel derivations during acquisition, in an analysis which subsumed free relatives and adjuncts.

REFERENCES

Adger, D. 2003. *Core Syntax*. Oxford: Oxford University Press.

- Adger, D. & P. Svenonius. 2011. Features in Minimalist Syntax. In C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, 27–51. Oxford University Press.
- Baker, M. 2008. The macroparameter in a microparametric world. In T. Biberauer (ed.), *The limits of syntactic variation*, 351–74. Amsterdam: John Benjamins.
- Bianchi, V. 1999. *Consequences of antisymmetry: Headed relative clauses*. Berlin: Mouton de Gruyter.
- Boeckx, C. 2008. *Bare syntax*. Oxford: Oxford University Press.
- Boeckx, C. 2010. A tale of two minimalisms: Reflections on the plausibility of crash-proof syntax, and its free-merge alternative. In M. Putnam (ed.), *Exploring crash-proof grammars*, 105–124. Amsterdam: John Benjamins.
- Boeckx, C. & K. K. Grohmann. 2007. Putting phases into perspective. *Syntax* 10. 204–222.
- Borer, H. 1984. *Parametric syntax*. Dordrecht: Foris.
- Broen, P. A. 1972. *The verbal environment of the language-learning child*. Washington, DC: American Speech and Hearing Association.
- Cecchetto, C. & C. Donati. 2010. On labeling: Principle C and head movement. *Syntax* 13(3). 241–278.
- Cecchetto, C. & C. Donati. 2015. *(Re)labeling*. Cambridge, MA: MIT Press.
- Chametzky, R. 2000. *Phrase structure*. Oxford: Blackwell.
- Chomsky, N. 1970. Remarks on Nominalization. In R. Jacobs & P. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 184–221. Waltham, Mass.: Ginn.
- Chomsky, N. 1975. *The logical structure of linguistic theory*. New York: Plenum.
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, N. 1993. A minimalist program for linguistic theory. In K. Hale & S.J. Keyser (eds.), *The view from building 20*, 1–52. Cambridge, MA: MIT Press.
- Chomsky, N. 1995a. Bare phrase structure. In G. Webelhuth (ed.), *Government and binding theory and the minimalist program*, 383–439. Oxford: Blackwell.
- Chomsky, N. 1995b. *The Minimalist Program*. Cambridge, Mass.: 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. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2004. Beyond explanatory adequacy. In A. Belletti (ed.), *Structures and beyond*, Oxford: Oxford University Press.
- Chomsky, N. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36(1). 1–22.
- Chomsky, N. 2007. Approaching UG From Below. In U. Sauerland & H.-M. Gärtner (eds.), *Interfaces + Recursion = Language?*, 1–29. New York: Mouton de Gruyter.
- 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. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2013. Problems of projection. *Lingua* 130. 33–49.

- Chomsky, N. 2014. Problems of projection: Extensions. Ms., MIT. To appear in E. Di Domenico, C. Hamann and S. Matteini (eds.), *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*, Amsterdam/Philadelphia: John Benjamins.
- Citko, B. 2008. Missing labels. *Lingua* 118. 907–944.
- Collins, C. 2002. Eliminating labels. In S.D. Epstein & T.D. Seely (eds.), *Derivation and explanation in the minimalist program*, 45–61. Oxford: Blackwell.
- Donati, C. 2006. On wh-head movement. In L. Cheng & N. Corver (eds.), *Wh-movement: Moving on*, 21–46. Cambridge, MA: MIT Press.
- Donati, C. & C. Cecchetto. 2011. Relabeling heads: A unified account for relativization structures. *Linguistic Inquiry* 42. 519–560.
- Epstein, S. D., H. Kitahara & T.D. Seely. 2014. Labeling by minimal search: Implications for successive-cyclic A-movement and the conception of the postulate “phase”. *Linguistic Inquiry* 45(3). 463–481.
- Epstein, S. D., H. Kitahara & T.D. Seely. 2015. What do we wonder is not syntactic? In S. D. Epstein, H. Kitahara & T.D. Seely (eds.), *Explorations in maximizing syntactic minimization*, 222–239. London: Routledge.
- Fortuny, J. 2008. *The emergence of order in syntax*. Amsterdam: John Benjamins.
- Frampton, J. & S. Gutmann. 2000. Agreement is Feature Sharing. Ms., Northeastern University.
- Frampton, J. & S. Gutmann. 2002. Crash-proof syntax. In S.D. Epstein & D.T. Seely (eds.), *Explanation and derivation in the minimalist program*, 90–105. Oxford: Blackwell.
- Fukui, N. 2011. Merge and bare phrase structure. In C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, 73–95. Oxford: Oxford University Press.
- Fukui, N. & M. Speas. 1986. Specifiers and projection. *MIT working papers in linguistics* 8. 128–172.
- Guimarães, M. 2000. In defense of vacuous projections in bare phrase structure. *University of Maryland Working Papers in Linguistics* 9. 90–115.
- Hale, K. & S.J. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In K. Hale & S.J. Keyser (eds.), *The view from building 20*, 53–109. Cambridge, MA: MIT Press.
- Halle, M. & A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In K. Hale & S.J. Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.
- Hornstein, N. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.
- Hornstein, N., J. Nunes & K. K. Grohmann. 2005. *Understanding Minimalism*. Cambridge: Cambridge University Press.
- Irurtzun, A. & A.J. Gallego. 2007. Consequences of pair-merge at the interfaces. *International Journal of Basque Linguistics and Philology* 41(2). 179–200.
- Jackendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jackendoff, R. 1977. *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

- Kaisse, E. M. 1977. On the syntactic environment of a phonological rule. *Chicago Linguistics Society* 13. 173–185.
- Kayne, R. 1984. *Connectedness and binary branching*. Dordrecht: Foris.
- Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.
- Kayne, R. 2009. Antisymmetry and the lexicon. *Linguistic Variation Yearbook* 8. 1–31.
- Kratzer, A. 1996. Severing the external argument from its verb. In J. Rooryck & L. Zaring (eds.), *Phrase structure and the lexicon*, Dordrecht: Springer.
- Lyons, J. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Marantz, A. 1997. No Escape from Syntax. In A. Dimitriadis, L. Siegel, C. Surek-Clark & A. Williams (eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium*, vol. 4 (U. Penn Working Papers in Linguistics 2), 201–225. University of Pennsylvania.
- McCloskey, J. 2000. Quantifier Float and wh-movement in an Irish English. *Linguistic Inquiry* 31. 57–84.
- Moro, A. 2000. *Dynamic Antisymmetry: movement as a symmetry breaking phenomenon*. Cambridge, MA: MIT Press.
- Narita, H. 2011. *Phasing in full interpretation*: Harvard University dissertation.
- Narita, H. 2012. Remarks on the nature of headedness and compositionality in bare phrase structure. In H. Toyoda (ed.), *Proceedings of sophia university linguistic society* 26, 81–126. Sophia University Linguistic Society.
- Narita, H. 2014. *Endocentric structuring of projection-free syntax*. Amsterdam: John Benjamins.
- Nespor, M. & I. Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Ott, D. 2010. Grammaticality, interfaces and UG. In M. Putnam (ed.), *Exploring crash-proof grammars*, 89–104. Amsterdam: John Benjamins.
- Ott, D. 2011. A note on free relative clauses in the theory of phases. *Linguistic Inquiry* 42(1). 183–192.
- Panagiotidis, P. 2014. *Categorial features: A generative theory of word class categories*. Cambridge: Cambridge University Press.
- Pesetsky, D. & E. Torrego. 2006. Probes, goals and syntactic categories. In Y. Otsu (ed.), *Proceedings of the seventh annual Tokyo conference on psycholinguistics*, 25–60. Tokyo: Hituzi Syobo.
- Pesetsky, D. & E. Torrego. 2007. Phrasal and clausal architecture. In Karimi S., V. Samiian & W. K. Wilkins (eds.), *The syntax of valuation and the interpretability of features*, 262–294. Amsterdam: John Benjamins.
- Phillips, J. R. 1973. Syntax and vocabulary of mothers' speech to young children: Age and sex comparisons. *Child Development* 44. 182–185.
- Radford, A. 2004. *Minimalist syntax: Exploring the structure of English*. Cambridge: Cambridge University Press.
- Rizzi, L. 2006. On the form of chains: Criterial positions and ECP effects. In L. Cheng & N. Corver (eds.), *On wh movement*, 97–133. Cambridge, MA: MIT Press.
- Rizzi, L. 2010. On some Properties of Criterial Freezing. In E. P. Panagiotidis (ed.), *The complementizer phase: Subjects and operations*, 17–32. Oxford: Oxford Uni-

- versity Press.
- Rizzi, L. 2013. *Criteria, the labeling algorithm and the “Halting Problem”*. Workshop at the University of Cambridge, 25 September 2013.
- Rizzi, L. 2015. Cartography, criteria, and labeling. In U. Shlonsky (ed.), *Beyond functional sequence: The cartography of syntactic structures*, 314–338. Oxford: Oxford University Press.
- Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.), *Elements of Grammar*, 281–337. Dordrecht: Springer.
- Saito, M. 2013. *Case and labeling in a language without ϕ -feature agreement*. Ms., Nanzan University.
- Schneider-Zioga, P. 2014. *Linking, predication, and symmetry: On the syntax of the linker in Kinande*. Ms., California State University.
- Seely, T. D. 2006. Merge, derivational c-command, and subcategorization. In C. Boeckx (ed.), *Minimalist essays*, 182–217. Amsterdam: John Benjamins.
- Selkirk, E. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Selkirk, E. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.
- Selkirk, E. 1995a. The prosodic structure of function words. In J. Beckman, L. W. Dickey & S. Urbanczyk (eds.), *University of Massachusetts occasional papers 18: Papers in optimality theory*, 439–469. Amherst, MA: GLSA.
- Selkirk, E. 1995b. Sentence prosody: Intonation, stress and phrasing. In J. Goldsmith (ed.), *The handbook of phonological theory*, 550–569. Cambridge, MA: Blackwell.
- Selkirk, E. 2011. The syntax-phonology interface. In J. Goldsmith, J. Riggle & A. Yu (eds.), *The Handbook of Phonological Theory*, 435–484. Oxford: Blackwell 2nd edn.
- Stockwell, R. 2014a. *On Labelling in Syntax*. BA dissertation, University of Cambridge.
- Stockwell, R. 2014b. *The End of Endocentricity?* Ms., University of Cambridge.
- Truckenbrodt, H. 1999. On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30(2). 219–255.
- Vogel, I., M. Drigo, A. Moser & I. Zannier. 1983. La cancellazione di vocale in Italiano. *Studi di Grammatica Italiana* 191–230.
- Wexler, K. 2003. Lenneberg’s dream: Learning, normal language development, and specific language impairment. In Y. Levi & J. Schaeffer (eds.), *Language competence across populations: Toward a definition of specific language impairment*, 11–61. Mahwah, NJ: Lawrence Erlbaum.
- Zwart, J. W. 2004. *Unary Merge*. Paper presented at the Tilburg University Grammaticamodellen Staff Seminar.

Richard Stockwell
 University of Cambridge
richard.stockwell@cantab.net