Remarks on the Nature of Headedness and Compositionality in Bare Phrase Structure

Hiroki Narita
Waseda Institute for Advanced Study

Abstract

The purpose of the present article is to highlight some of the contemporary issues surrounding the notion of headedness and compositionality in bare phrase structure. The following list enumerates questions to be discussed in this article.

[**1**] Labeling = projection = headedness?

[**2**] Why headedness?

[**3**] Is headedness relevant to narrowly syntactic computations or does it arise only at SEM/PHON?

[**4**] Is Merge 'to a head'? Is the application of Merge constrained by headedness (or the 'edge-feature' of lexical items)?

[**5**] How can headedness be assigned to {XP, YP}?

[**6**] Are traces of Internal Merge invisible?

[**7**] Does Transfer involve deletion of the phase-interior?

[**8**] Are agreement features visible at SEM?

[**9**] Does *{XP, YP} constrain narrowly syntactic derivations or apply only at SEM?

[**10**] Are symmetric structures locally unstable?

[**11**] What are the 'atoms for computation'?

[**12**] Is endocentricity universal?

This article will address these questions, step by step, in this order.

[1] Labeling = projection = headedness?

One of the main purposes of this article is to argue that the following three concepts should be carefully distinguished from each other: labeling, projection, and headedness (endocentricity). The discussion in this article will be based on the following definitions of the three terms, which are hardly controversial:

(1) a. labeling: association of nonterminal symbols (*labels*; S, NP, V', etc.) with phrasal constituents
b. projection: duplication of features of a lexical item (LI) onto (typically 'dominating') constituents
c. headedness (endocentricity):[1] centrality of a certain LI in the distribution/interpretation of a constituent

Since the advent of X-bar theory initiated by Chomsky (1970), it has become customary for generative linguists to assume the following equation:

(2) labeling = projection = headedness

Historically, the concept of labeled phrase structure developed from the earlier theory of phrase structure grammar (PSG), where phrase-markers are generated step-by-step by a system of phrase structure rules (PSRs, rewrite rules). (3) summarizes some of the familiar examples of PSRs, each of which takes the nonterminal symbol on the left hand side as its input and expands it into the sequence of symbols on the right hand side.

(3) a. S' → COMP S
b. S → NP Infl VP
c. VP → V NP
d. NP → (D) N

[1]While I will follow the conventional terminology in using the terms 'headedness' and 'endocentricity' interchangeably, it may be worth noting that the notion of endocentricity was originally proposed by Bloomfield (1933:194ff, 235-36) to refer to a concept different from the one defined here (see Wujastyk 1982, Graffi 2001; see also Lyons 1968). Therein, the endocentric construction is defined as a grammatical construction that fulfills the same linguistic function as one of its constituents: therefore, *poor John* is an endocentric constituent, as its categorial nature and distribution coincides with that of *John*, while *in Vancouver* is classified as exocentric, since its distribution is shared by neither *in* nor *Vancouver*. Endocentric constructions in the Bloomfieldian sense are pretty much what generative linguists analyze as constituents involving adjuncts. It is an open question whether it is justifiable to collapse the distinction between the Bloomfieldian notion of endocentricity and the generativist notion of headedness.

According to the theory of PSG, nonterminal symbols like S, NP, and VP are generated and buried in phrase-markers as an automatic consequence of structure-generation, since they constitute necessary input and output of PSRs. Representational labels were, in this sense, plainly indispensable in the earlier PSG-based syntax. In this respect, it was an ingenious decision for transformational grammarians to attempt to utilize the explanatory force of this hypothetical construct to its limit. As a representative example, the X-bar theory put forward by Chomsky (1970) paved the way for extending the notion of labeling to the account of headedness (endocentricity) of phrase structure. It holds that the class of possible PSRs can be radically reduced to the following two general schemata, where a lexical item X is necessarily dominated by an intermediate category X', which in turn is necessarily dominated by the maximal category X" (Chomsky 1970; see also Jackendoff's 1977 tripartite X-bar structure).

(4) the X-bar schemata

a. X' → X (Y") or (Y") X
b. X" → (Z") X'

In the X-bar theory, labels of phrasal constituents are reduced to combinations of the lexical features of the head X and bar-level indices. In this theory, there is a strong sense in which nonterminals like X' and X" are 'projections' of X: N' and N" are projections of N, V' and V" are projections of V, and so on. That is, labeling = projection. It not only imposes a severe restriction on possible forms of PSRs (see Stowell 1981, Chomsky 1986, who further maintain that the concept of (language-specific) PSRs can be eliminated altogether), but it also provides room for an easy way to capture the headedness of phrase structure, namely the hypothesis that projection is *the* basis of headedness (namely, projection = headedness). The equation in (2) was thus advocated, and it has become one of the central hypotheses in the tradition originated from the X-bar theory.

However, the theory of *bare phrase structure* advanced by Chomsky (1994 *et seq.*) changed the picture. It holds that the compositional structure of human-language expressions is exhaustively characterized by recursive application of *Merge*. Merge is a binary set-formation operation that takes two syntactic objects (SOs), say α and β, and creates a set $\{\alpha, \beta\}$. Since no known condition of set theory requires that sets are associated with 'labels' or distinctive nonterminal symbols, only stipulations can guarantee assignment of any such extraneous entities to set-theoretic objects generated by Merge. Correspondingly, the representation of labels and projections in bare phrase structure is a plain departure from the strong minimalist the-

sis (SMT), constituting an enrichment of UG (Chomsky 2007a:23; see also Collins 2002, Seely 2006).

It is commonly supposed that we still need to stipulate a distinct mechanism of labeling/projection even in bare phrase structure, in order to capture the facts about headedness/endocentricity. However, note that such an argument would go through only if it is further shown that labeling by projection is *the* necessary device of commendable theory-internal simplicity to encode endocentricity. This was arguably the case in the earlier PSR-based syntax, where labels were anyway generated as a necessary component of compositional structuring, but not any more in bare phrase structure, where labeling is bound to be a departure from the SMT.

In this context, it should be noted that there are some proposals regarding headedness that make little to no recourse to labeling by projection. For example, I contend that the series of work by Chomsky (1995) can and should be understood as paving the way for a projection-free account of headedness: Chomsky (2000) hypothesizes that determination of the head of a set-theoretic object $\{\alpha, \beta\}$ correlates with selectional or probe-goal dependency between α and β, an idea followed by a number of researchers. Under this proposal, $\{\alpha, \beta\}$ is headed by (the head of) α if α selects or probes β. Collins (2002) is right in pointing out that this sort of selection/probe-based mechanism can be adapted to encode the centrality of 'head' LIs without making recourse to labels/projections, and his locus principle is a particular attempt to pursue this line of approach. Still after the selector/probe-based mechanism of head-detection became unavailable due to the elimination of selection from the theory of narrow syntax (Chomsky 2004:112-113; see also Pesetsky 1982), Chomsky (2008) puts forward another algorithm of head-detection (still called the 'labeling algorithm') in (5):

(5) a. In $\{H, \alpha\}$, H an LI, H is the head.
b. If α is internally merged to β, forming $\{\alpha, \beta\}$ then the head of β is the head of $\{\alpha, \beta\}$. (Chomsky 2008:145,(2)-(3), restated)

Chomsky (2010a, lectures at MIT in fall 2010) further suggests to eliminate (5b) from the algorithm of head-detection, reducing it to minimal search for an LI in each phrase (5a) (see also Piattelli-Palmarini et al. 2009:52ff). Let us refer to this minimal search-based definition of headedness as *Minimal Head Detection* (MHD):

(6) *Minimal Head Detection (MHD)*:
The head of an SO Σ is the most prominent LI in Σ.

These are all attempts to account for headedness without recourse to any extraneous mechanism of labeling/projection, departing from the equation in (2). MHD is articulating the view that head-detection may be reduced to the third-factor principles of computational efficiency (Chomsky 2005), maximally respecting the SMT. See Chomsky (2010a, 2011), Ott (2011a,c), Narita (2011a,b,c, forthcoming, to appear), and Lohndal (in progress) for various explorations of MHD.

Building on the prospect of label-free and projection-free approaches to headedness, the following discussion will specifically explore some of the consequences of MHD (6). It will summarize several important questions concerning the prospect of MHD, each of which will be stated as the heading of the relevant section.

[2] Why headedness?

Even though the theory of bare phrase structure eliminates the basis of representational labels and projections, it still seems clear that 'noun phrases' are interpreted in a 'nouny' way, while 'verb phrases' in a 'verbal' way, and likewise phrases 'headed by' P, A, C, T, etc., are all interpreted differently. Whatever theoretical analyses one may entertain to describe this intuition, the prominent effect of headedness on the CI-side of interpretation seems to be simply a matter of fact.

Moreover, the effect of headedness presumably governs the computation of the phonological component as well. As pointed out by Narita (2009, 2011a,c), virtually all the past proposals on the mechanism of *linearization* (*i.e.*, mapping of SOs to a corresponding sequence of LIs readable by SM) seem to share the fundamental hypothesis that linearization makes recourse to labeling by projection. Various theories of the Kaynean Linear Correspondence Axiom (LCA) (see Kayne 1994, Chomsky 1995, Epstein et al. 1998, Moro 2000, Sheehan 2010, in press among others), Fukui and Takano's (1998) theory of 'Demerge', and directionality parameters of all kinds (Chomsky 1981 *et seq.*) are all formulated with recourse to labels and projections, and some proposals that attempt to eliminate recourse to labels still rest on various notions of headedness (see Uriagereka 1999, Richards 2004, Kayne 2011 and Narita 2011a:Ch.4, 2011c for examples).

Therefore, we are interested in asking the following *why*-question:

(7) Why does the effect of headedness arise at SEM/PHON?

What (7) is calling for is an *explanation* of endocentricity/headedness, not so much a description of it. As we mentioned, countless descriptions of endocentricity have been put forward since X-bar theory (Chomsky 1970, Jackendoff 1977 *et seq.*), prototypically making heavy use of labeling by projection. But these previous pro-

posals have largely failed to address the *why*-question, largely because they have regarded labeling by projection as an axiom of syntactic analysis and thereby trivialized that very *why*-question as a matter of course. Departing from this tradition, then, what can we say about a possible explanation of endocentricity?

Importantly, the minimalist program (MP) offers a ready-made answer to such a question, namely (8):

(8) Headedness arises at SEM because it is the simplest mode of interpretation.

There is little doubt that this is the desideratum that we are led to defend in the pursuit of the SMT. The question is how we can make sense of this answer.

Narita (2011a:§5.2) puts forward a plausible line of reasoning to this effect. Let us consider the relation between syntax and CI. Syntax is the component of FL that generates SOs and maps them to the interface with CI (SEM), and CI is the performance system that receives these SOs via SEM and exploits them for human thoughts. Let us follow Lasnik et al.'s (2005) terminology and refer to the mapping of SOs to SEM as *Interpret*. Interpret and its phonological counterpart (*Spell-Out*, the mapping of SOs to PHON) constitute *Transfer* (see Chomsky 2004 *et seq.*). The generation of SOs and their mapping to SEM via Interpret should be achieved in an efficient and optimal fashion, so long as the SMT holds.

Among other things, an important consequence of the efficient exploitation of SOs by CI should be what is often called the *strict compositionality* of semantic interpretation. Simply put, it holds that the formal semantic property of a linguistic expression is contingent on those of its parts. Captured in terms of bare phrase structure, then, it should be that the interpretation assigned to an SO $\{\alpha, \beta\}$ at SEM is determined in some way or another by the properties of α and β. This is arguably an optimal way of assigning interpretation to phrasal SOs, rooted in computational efficiency, and this should be why syntax generates SOs at all: otherwise, there is little reason for syntax to generate hierarchical structures to begin with.

Narita further points out that in order to assign compositional interpretation to each SO $\{\alpha, \beta\}$, then, Interpret will need to 'inspect' semantic properties of α and β. Note that each of α and β may be another set-theoretic object $\{\gamma, \delta\}$, whose semantic properties are then determined by γ and δ. Then, the relevant inspection may go on, recursively, all the way down to an 'atomic element' whose semantic properties are not dependent on any smaller entities. These elements are called *lexical items* (LIs), acquired from radically impoverished linguistic data from the surrounding environment and stored in the Lexicon, presumably with the help of an innately endowed capacity of concept-formation (the uniformly rapid and success-

ful acquisition of LIs is in itself a quite marvelous fact about human infancy, but maybe rendered somewhat manageable by the help of such analytic means as the one advocated by Narita 2011a,c, forthcoming; see [11] for details). Then, (9) is minimally required from the principle of strict compositionality:

(9) *Semantic features of LIs*:
Features of each LI H contain, at the very least, instructions for Interpret regarding how H contributes to the interpretation of the SO it is embedded into, the smallest one of which is $\{H, \alpha\}$.

We may refer to such features of LIs as *semantic features*.[2] In this framework, Interpret can start assigning interpretation to each SO Σ only after it inspects the internal constitution of Σ and finds an LI that can serve as a locus of compositional interpretation. Presumably, then, such inspection of LIs should be minimized, for the sake of efficient computation at Interpret.

Narita maintains that MHD is this minimal inspection operation. MHD, working in tandem with the semantic features of LIs (9), guarantees (10):

(10) Each phrasal SO $\{H, \alpha\}$ can efficiently receive compositional interpretation at SEM via the semantic features of H.

This much is empirically necessary for any theory of SEM, and thus it strictly adheres to the SMT. Based on this hypothesis, we can give well-defined theoretical content to the notion of head in these terms:

(11) The *head* of an SO Σ is the LI that determines compositional interpretation of Σ at SEM by means of its semantic features.

In this line of reasoning, then, headedness of phrase structure arises at SEM simply as a result of optimal exploitation of SOs via strict compositionality.

The thesis advocated here can be formulated as follows (cf. Narita 2011a:§5.2):

(12) *Thesis*:
MHD reduces to the minimized inspection of semantic features of LIs at Interpret.

This line of reasoning is what I take to be a promising way for making sense of the minimalist answer (8) to the *why*-question (7): *headedness is reduced to just*

[2] See McGilvray (1998, 2002, 2009) and Pietroski (2005, 2008 *et seq.*) for discussion of the conception of semantic features (or 'concepts') as instructions for CI.

a cover term for the simplest mode of compositional interpretation. Although criticisms have been raised for some of the technical aspects of Narita's proposal, to be examined in what follows, this approach to the explanation of headedness still seems promising to me in light of the SMT. I will investigate some of the grounds and consequences of (12) in the following discussion.

Notice also that if we follow Chomsky (2007a,b, 2008) in speculating that the faculty of language (FL) is designed primarily for internal thoughts (*i.e.*, for CI purposes), it is reasonable to suppose that the effect of headedness, optimal for CI, has a correlative influence on the mapping to PHON as well. Especially under the assumption that Interpret occurs concurrently with Spell-Out (Chomsky 2004 *et seq.*; see Narita 2011a:Ch.2 and Samuels forthcoming), it is reasonable to suppose that MHD, applying primarily for the sake of Interpret, also feeds information about headedness to Spell-Out as well. See Narita (2011a:Ch.4, 2011c) for exploration of MHD-based theory of linearization at Spell-Out.

[3] Is headedness relevant to narrowly syntactic computations or does it arise only at SEM/PHON?

As noted in the previous section, the effect of headedness (via MHD) clearly figures in the compositional interpretation of SEM, and presumably in phonological linearization, too. Be that as it may, it is not obvious whether headedness has any effect on the computation of narrow syntax, hence the question in the heading of this section. The answer to this question will depend on whether there are any narrowly syntactic operations whose application is based on the headedness of SOs.

The following is a list of the most obvious candidates for narrowly syntactic operations:

(13) a. Merge
b. Agree (probe-goal search)
c. Transfer
d. selection (c(ategorial)-selection, subcategorization, θ-role assignment, etc.)

None of these operations appear to rest on headedness of the input SO, as we will briefly review below. First, as noted by Chomsky (2004:113), Agree (13b) is just a relation between two LIs (or features shared by them), and what is required for the application of Agree is just that the 'probe' LI c-commands the 'goal' LI.

Consider the operation *Transfer* (13c) next. Stated in the contemporary theory

of phases developed by Chomsky (2000 *et seq.*), a *phase* is an SO to which the operation called Transfer applies upon its completion, and Transfer is an operation that cyclically strips off a certain domain of a phase, called the *interior*, subjecting it to the mappings to SEM and PHON (Interpret and Spell-Out). Notice that the domain of Transfer can be defined purely structurally: the interior of a phase is the sister of a certain set of LIs called 'phase heads' (C, *v*, P, etc.), and the edge of the phase is the domain of the phasal SO that is not (contained in) the interior (see Chomsky 2000, 2008, Richards 2007, Narita 2011a for various definitions of phases). Again, no recourse to labels or headedness is necessary in the formulation of Transfer (13c).

Furthermore, one plausible approach holds that c(ategorial)-selection and subcategorization are entirely reducible to s(emantic)-selection (Pesetsky 1982), and that s-selection plays virtually no role in narrow syntax (Chomsky 2004:112-113). Considerations about differing acceptability judgments are really pointing to the same conclusion: occasional violation of selectional restrictions is a common fact about ordinary language use, and it often does not result in clear-cut unacceptability, which suggests that CI just retroactively assigns selectional interpretations to whatever SOs it receives from syntax, yielding various degrees of deviance (see Chomsky 1955/1975, 1965, Chomsky 2004:112, Ott 2009, 2010, Boeckx 2010b for much relevant discussion of the notion of degrees of grammaticalness). Then, selection is by and large a matter of post-syntactic concern, and hence irrelevant to the discussion of narrow syntax.

This leaves us with Merge (13a), the most fundamental operation of narrow syntax. The theory of bare phrase structure imposes the simplest formulation of Merge, namely that its application is truly unconstrained. Empirically, unconstrained Merge immediately provides a principled explanation of the discrete infinity of human language. It also explains the ubiquity of displacement phenomena (with reconstruction effects), as it characterizes movement (with copy theory) as a result of free application of internal Merge (IM). Any stipulations that depart from unconstrained Merge will lose accounts of these fundamental facts about human language, and hence they will bear a heavy burden of proof. Then, the simplest assumption is that anything can be merged with anything, irrespective of the headedness of the SOs.

Although headedness has often been viewed as a prerequisite for Merge-application, the next section will provide arguments for the view that the simplest, truly unconstrained conception of Merge should be favored. That is, the negative answer to the question in [4] will be provided. It will in turn complete the negative answer to question [3], namely that the notion of headedness is purely interpretive and arises

only at SEM/PHON, meaning it is irrelevant to narrowly syntactic computations. Note that this conclusion is required for the principled account of headedness discussed in [2], which holds that it optimally reduces to the simplest mode of compositional interpretation appearing at the interfaces. Any additional conceptions will require independent justification, but fortunately, as discussed in this and the following section, there is little reason to suppose that headedness figures in narrowly syntactic operations.

Incidentally, this conclusion further undermines the rationale for representational labels/projections. Some researchers maintain that syntactic representation of labels is required on grounds of computational efficiency, insisting that it significantly contributes to reduction of the search space, since labels, bearing all relevant features of the head, provide 'closer' goals for downward probing (see, *e.g.*, Hornstein 2009:Ch.3 and Ott 2011a:§3.1.2). However, this kind of argument has empirical force only under the assumption that syntactic dependencies are established on the basis of labels/headedness, which is in itself an empirically questionable and conceptually ill-advised premise.

[4] Is Merge 'to a head'? Is the application of Merge constrained by headedness (or the 'edge-feature' of lexical items)?

Optimally, Merge is an unconstrained operation that can apply to any SOs. Importantly, however, it has been predominantly assumed in the literature that Merge is 'to a head', namely that its application rests on headedness in a certain sense.

(14) *Traditional assumption*:
Merge is attraction of an SO to a head LI.

For example, the specifier-complement distinction is commonly characterized as *first-merged* and *later-merged* to the head LI (Chomsky 2007a:11; 2008:146). Relatedly, consider the following passage from Chomsky (2007a):

(15) "If an element Z (lexical or constructed) enters into further computations, then some information about it is relevant to this option: at the very least, a property that states that Z can be merged, but presumably more, it is commonly assumed. The optimal assumption is that this information is provided by a designated minimal element of Z, a lexical item W (Z itself, if it is an LI), which is detectable by a simple algorithm; the *label* of Z, the head projected in X-bar theories." (Chomsky 2007a:9)

Chomsky is proposing here that headedness is a necessary condition for Merge-application.[3] He claims that there is a 'simple algorithm' of head-detection in narrow syntax that can identify a designated minimal LI W within Z that provides for Z, among other things, the property that states that Z can be merged with some SO. This property is what Chomsky (2008) calls the *edge-feature* (EF) of LIs. Regarding the characterization of EFs, Chomsky (2008) specifically provides the following remark:

(16) "For an LI to be able to enter into a computation, merging with some SO, it must have some property permitting this operation. A property of an LI is called a *feature*, so an LI has a feature that permits it to be merged. Call this the *edge-feature* (EF) of the LI. If an LI lacks EF, it can only be a full expression in itself; an interjection. When merged with a syntactic object SO, LI forms {LI, SO}; SO is its complement. The fact that Merge iterates without limit is a property at least of LIs—and optimally, only of LIs, as I will assume." (Chomsky 2008:139)

Here, Chomsky states that (17) is the effect that the notion of EF should capture:

(17) *Edge-features* (EF) (Chomsky's (2008) version):
   a. The EF is the feature that permits its bearer to be merged with some SO.
   b. The EF is a property only of LIs.

One way to make sense of (17a) is to assume that the EF is the feature that constitutes the locus, or trigger, of Merge-application, a formulation reminiscent of the earlier 'EPP-feature' of Chomsky (1995) that states, "I need a specifier." Generalizing to the merger of 'complement' along the line suggested by Chomsky, we may say that the EF is the property of LIs that states, "I may have a sister." Assuming that only LIs have EFs (17b), Chomsky (2007a) proposes in (15) that for a phrasal SO Z to be subjected to Merge, syntax must first inspect the internal constitution of Z and find the EF of the head LI W. That is, the EF must be 'projected' from head LIs to phrasal SOs, in order for Merge to apply to them.

(18) Application of Merge to an SO Σ is contingent on the EF of the head LI of Σ.

[3] Although the relevant notion is called the 'label' or the 'projected' head, it is clear that no notion of representational labels or projection of features as defined in (1) is implied in Chomsky's theory.

In this way, Chomsky reformulates the common conception of Merge as 'attraction' to the head LI (the traditional assumption in (14)) by making recourse to the notion of EF.

However, whether the stipulation in (14)/(18) is empirically adequate is questionable. First and foremost, this restriction on Merge is a departure from the overarching hypothesis of unconstrained Merge, and hence should be avoided unless strong evidence is provided. Moreover, Noam Chomsky (p.c.) points out that there are clear cases where the hypothesis of unconstrained Merge makes superior predictions. Thus, he notes in his comments on Narita (2011a):

(19) "In the background are two different conceptions of composition. Should Merge be attraction to the head, in which case we have first-Merge (complement), second-Merge (SPEC), questions about additional Merge? Or should Merge just be as stated in the simplest version, Merge(X,Y) = {X,Y}, heads irrelevant? The former was the conception in X-bar theory and its descendants, connected with the universal endocentricity assumption [to be discussed in [12], HN]. The latter is simpler, and also has empirical advantages I think—e.g., successive-cyclic movement, where properties of the head are irrelevant."

Indeed, there is no clear sense in which successive cyclic movement is contingent in any way on properties of head LIs: thus, internal merger of a wh-phrase to the edge of *v* (as in, *e.g.*, *what did you t buy t?*) does not depend on any lexical properties of *v*, or of the wh-phrase for that matter.[4] The existence of such 'head-independent' merger clearly fares well with the unconstrained conception of Merge. Further, it is also worth noting that the EF is originally proposed in (16) as a feature that distinguishes mergeable LIs from non-mergeable LIs like interjections or frozen expressions, and it was just a highly theory-internal decision to make use of this concept in formulating the 'Merge-as-attraction-to-the-head' hypothesis as in (15), unwarranted as it stands now. Then, for the sake of pushing further the hypothesis of unconstrained Merge, we need to eliminate the head-dependent conception of Merge in (14)/(18) (see also Boeckx 2008, 2010a and Fukui 2011 for related discussion).[5]

[4]It has been customary to describe the movement as 'head-dependent', as in Chomsky (2000, 2001), where it is proposed that the movement is dependent on (or triggered by) the optionally assigned EPP-feature of *v*. But such complication of the applicability of Merge is a departure from the SMT, as pointed out by (Chomsky (2008)).

[5] It should be noted in passing that this conclusion is incompatible with one of Narita's (2011a, forthcoming) rationalizations of what he calls the *H-α schema*, which was based on the very con-

[5] How can headedness be assigned to {XP, YP}?

Let us go back to the minimal-search based conception of head-detection, MHD repeated here:

(6) *Minimal Head Detection (MHD)*:
The head of an SO Σ is the most prominent LI in Σ.

As noted, the effect of MHD is that for any SO $\{H, \alpha\}$, where H is an LI and α an SO, H is the head of $\{H, \alpha\}$.

Note that MHD also makes the following reverse prediction: MHD cannot determine headedness for any SOs that depart from the form of $\{H, \alpha\}$. A prototypical case is $\{XP, YP\}$, where both of the merge-mates are phrasal SOs and hence no LI immediately stands as the most prominent. This state of affairs can be described as (20).

(20) MHD cannot determine headedness for $\{XP, YP\}$.

Note that as long as the notion of headedness reduces to the barest mode of compositional interpretation, as proposed by Narita (2011a:§5.2) (see [2]), (20) amounts to saying that $\{XP, YP\}$ cannot be assigned compositional interpretation. Then, such structures should be ruled out by the principle of Full Interpretation (FI).

(21) *Full Interpretation (FI)* :
Every constituent of SEM and PHON contributes to interpretation.

I will refer to this prediction as *Representational* *$\{XP, YP\}$ ('representational' since it speaks to the legibility of interface representations).

(22) *Representational* *$\{XP, YP\}$:
Headless SOs of the form $\{XP, YP\}$ are ruled out by FI at SEM.

At face value, this prediction seems to be contrary to the observation that instances of 'XP-YP structures' appear to be abundant in natural languages, and to be falsified by simple sentences like [[the man] [kissed the girl]]. Then, at first glance, the simplest conception of headedness seems unsustainable.

Note that in earlier theories of labeled phrase structure, the problem of headedness of $\{XP, YP\}$ was swept under the rug by the stipulation of labeling/projection:

ception of EFs in (17) and the Merge-as-attraction-to-the-head. However, the other of Narita's two rationalizations of the H-α schema is based solely on MHD and stay unaffected, as we will see in [9].

the headedness of {XP, YP} is simply predetermined by virtue of the structure being labeled by the head of either XP or YP, and so the other phrase is called a 'specifier' of that head. However, reference to such label-dependent notations is simply unavailable in the theory of label-free syntax (see Chomsky 2010a, 2011, Narita 2011a:§5.3 and Lohndal in progress).

There are at least three options proposed in the literature of label-free syntax that pave the way for effectively circumventing Representational *{XP, YP}. The first option, proposed by Chomsky briefly in his (2008:160, note 34) and explored in his lectures at MIT in fall 2010, is to stipulate that traces of IM 'don't count' for the purpose of head-detection (see also Ott 2011a,c).

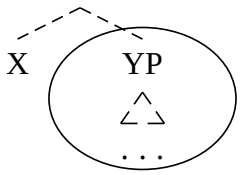
(23) *Trace invisibility* (Hypothesis 1):
Traces of IM are invisible to MHD.

(23) maintains that if a phrasal constituent XP of {XP, YP} undergoes IM, the original occurrence ('trace', informally speaking) of XP in {XP, YP} becomes invisible to head-detection, and so MHD will just 'look into' YP and single out the head of YP as the most prominent LI within {XP, YP}, determining headedness.

The second option for circumventing Representational *{XP, YP}, proposed by Narita (2011a,b,c, forthcoming, to appear), is based on a specific hypothesis on the operation of *Transfer* proposed in Chomsky (2004, 2008) (see already Uriagereka 1999 and Nunes and Uriagereka 2000). Stated in the contemporary theory of phases developed by Chomsky (2000 *et seq.*), Transfer is an operation that cyclically strips off the interior of a *phase*, subjecting it to the mappings to SEM and PHON (Interpret and Spell-Out). Chomsky (2004, 2007a, 2008) proposes that the phase-interior subjected to Transfer will be literally 'forgotten', *i.e.*, rendered invisible for further computation.

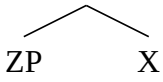
(24) *Transfer*:
Applied to a phase Σ, Transfer strips off all the structural information related to the interior of Σ from the active workspace of narrow syntax and maps it to SEM and PHON.

Under this conception, Transfer has an effect of periodically reducing computational loads of narrow syntax. For example, if a phase-head LI X takes a phrase YP as its sister, Transfer strips off YP from the workspace of narrow syntax at the completion of the phase headed by X, and subjects it to mappings to SEM and PHON.

(25) a. 

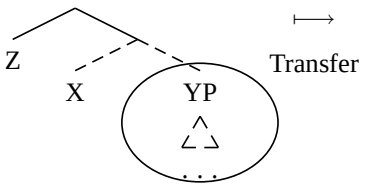
⟼ Transfer b. X

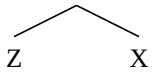
Narita (2011a:§2.5.1) put forward the hypothesis that Transfer subjects to interpretation not only YP (and elements within it) but also all the syntactic relations established with regard to it, maximally respecting the effect of periodic reduction of computational loads. He specifically maintains that the information subjected to Transfer includes the information that YP was merged with X, forming {X, YP}. Thus, applying to a phase {X, YP}, only the phase-head LI X will be left by Transfer for further computation.[6] After Transfer, nothing precludes the remaining SO X from merging with some other ZP, forming {X, ZP}.

(25) c. X ⟼ Merge with ZP d. 

Note that while this derivation effectively achieves merger of the X-phase and ZP, it does not involve any step at which an SO of the form {XP, YP} is formed. Given these considerations, Narita maintains that a phase headed by X can still be merged with some other phrasal SO, circumventing Representational *{XP, YP}, as long as it can be reduced ('atomized') to a bare LI X by Transfer.[7]

[6] The situation is different in cases where the phase headed by X assumes a 'specifier' SO Z at its edge. In such cases, Transfer cannot reduce the entire phase to an LI X, since not only X but also Z is left for further computation (Chomsky 2000 *et seq.*).

(i) a. 

⟼ Transfer b. 

[7] The idea that Transferred phases count as 'atomic' elements has been proposed in various places. For example, Uriagereka (1999) and Nunes and Uriagereka (2000) propose that Transfer/Spell-Out has the effect of eliminating all but the topmost label of an XP from the derivational workspace. This proposal has been adopted by, *e.g.*, Sheehan (in press) among others. However, it should be clear from the present discussion that I am here explicitly departing from their label-based conception of the 'atomization' effect of Transfer. Instead of making recourse to labeling/projection, I am claiming here that the relevant atomizing effect arises simply as a result of eliminating not only the phase-interior XP but also the 'is-merged-with' relation it has established with the phase head H by forming {H, XP}.

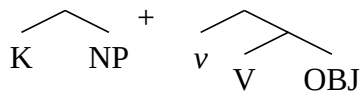
(26) *Atomization by Transfer* (Hypothesis 2):
Transfer applying to a phase of the form {X, YP} effectively reduces it to an LI X.

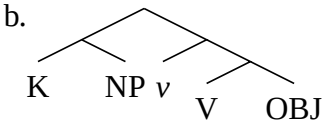
Yet another way to determine headedness for certain {XP, YP} structures is proposed by Chomsky (p.c., lectures at MIT in fall 2010, and in Chomsky 2011). The idea is summarized in (27), which I will refer to as the hypothesis of *Bifurcated MHD*:

(27) *Bifurcated MHD* (Hypothesis 3):
a. Not only an LI but also a feature can be a head determined by MHD.
b. Head-inspection via MHD can be bifurcated for SOs of the form {XP, YP}.
c. If XP and YP share a feature F as the most prominent element, then the bifurcated inspection into XP and YP by MHD can single out F as the head of {XP, YP}.

The applicability of Bifurcated MHD is limited, but it has several important consequences, as we will see below.

Having in mind these three ways to circumvent Representational *{XP, YP}, let us consider the merger of the external argument to the edge of *v* in (28). Given the prevalence of article-less languages, I will follow Bittner and Hale (1996a,b), Neeleman and Weerman (1999), and Narita (2011a) among others in assuming that nominal phrases are headed by the functional category K(ase) with an unvalued Case-feature (which may be equivalent to *n*/*n** proposed in Chomsky 2007a). The syntax of nominal phrases may involve more articulated structures, but it does not matter much for the present discussion.

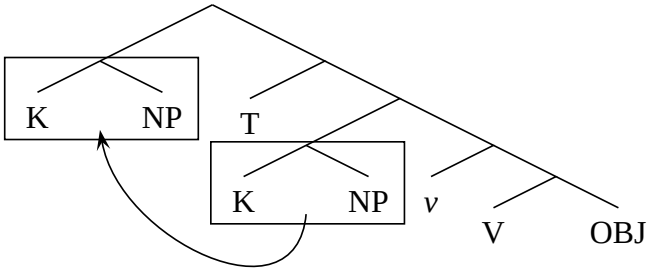
(28) a. 

⟼ b. 

The output of external Merge (EM) as such will be ruled out by Representational *{XP, YP}, but the three hypotheses mentioned above provide different ways to circumvent this problem.

Hypothesis 1 (trace invisibility, (23)) holds that movement of one of the XPs will suffice. Thus, if the subject KP is dislocated by A-movement to the 'EPP'-position, its original occurrence in (28b) will become invisible, thus MHD for (28b) can look into the structure further down and hit *v* as the most prominent LI in (28b).

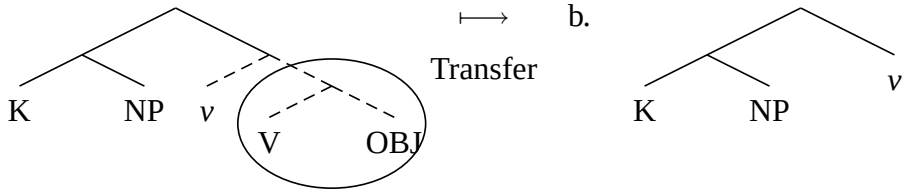
(29)



The other possibility is the derivation in which {*v*, {V, OBJ}} undergoes IM and the structure is thereby headed by K. However, advocates of Hypothesis 1 refrain from proposing this as a legitimate derivation, presumably due to the assumption that the θ-role assignment by *v* to the subject is contingent on *v* heading the relevant structure (see [10]).

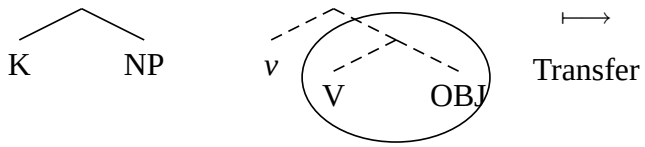
Next, Hypothesis 2 (atomization by Transfer, (26)) holds that if syntax can make use of Transfer and reduce at least one of the XPs to a simplex LI, the resultant structure can effectively be assigned a head by MHD. For example, if we adopt the standard assumption that *v* is a phase head, Transfer can strip off all the structural information related to its complement {V, OBJ} to the interfaces, and thereby reduce the phrase headed by *v* to a bare LI *v*.

(30) a.

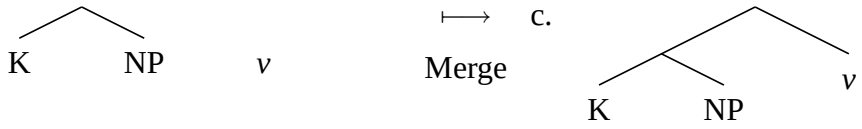


Or, without further stipulations, it may be that Transfer applies before the external merger in (28):

(31) a.



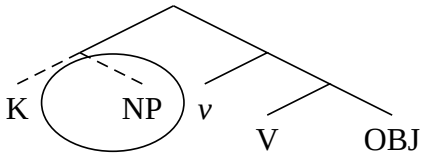
b.



In both derivations, Transfer of {V, OBJ} by *v* in effect enables *v* to become the head of the structure.

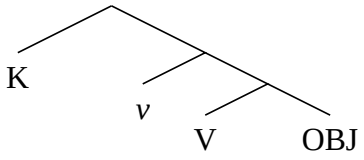
Narita also explores the possibility of atomizing the subject KP by Transfer for the relevant external merger (see also Uriagereka 1999). Assuming that K can constitute its own phase (see Narita 2011a), Transfer renders the K-phase a bare LI K, and so the K can sit in the sister of the verbal phrase in conformity with Representational *{XP, YP}. Again, Transfer may apply to KP before or after the external merger with the verbal phrase.

(32) a.

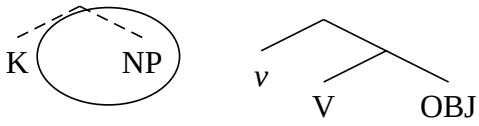


⟼ Transfer

b.

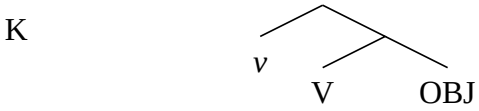


(33) a.



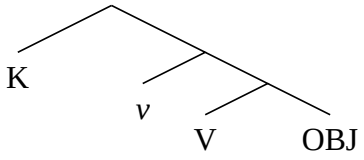
⟼ Transfer

b.



⟼ Merge

c.



In a nutshell, Hypothesis 2 holds that apparent cases of XP-YP (external) merger are still susceptible to derivations complying with Representational *{XP, YP}, as long as at least one of the two phrases is a phase that can be reduced by Transfer to a simplex LI.[8] This hypothesis is grounded on the insight that cyclic Transfer is crucially in service of recursive structure-embedding in label-free syntax (see Narita 2011a:Ch.3, forthcoming). Needless to mention, this approach crucially hinges on the proper formulation of what can count as phases: readers are referred to Chomsky

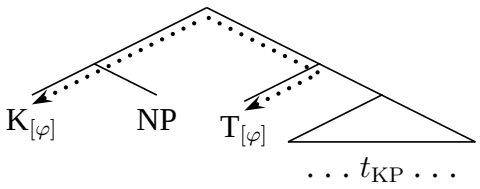
[8]See note 6.

2000, 2008, Richards 2007, Narita 2011a and many others for various proposals.

Incidentally, the external merger in (28) is a typical instance of XP-YP structure that cannot be dealt with by Bifurcated MHD (Hypothesis 3), as noted by Chomsky (2011). No feature is shared by the subject KP and the verbal phrase, and hence MHD cannot determine its head, even if the search is bifurcated for these two phrases.

Rather, Bifurcated MHD is proposed primarily for cases involving Agree and IM. Consider again the case of A-movement in (29). As noted above, Chomsky proposes that headedness for the SO {{K, NP}, {*v*, {V, OBJ}}} is successfully determined by MHD via trace invisibility (23). However, the same strategy does not suffice to assign headedness to the topmost node created by IM (29), since the higher occurrence of {K, NP} in (29) is not a trace. The XP-YP structure in (29) is instead dealt with by Bifurcated MHD in Chomsky's theory: it is traditionally assumed that the relevant 'EPP'-movement involves φ-feature agreement between the subject and T. Thus, the matching φ-features stand as the most prominent elements in both {K, NP} and {T,. . . }, and so they can stand as the head of {{K, NP}, {T,. . . }} via Bifurcated MHD.

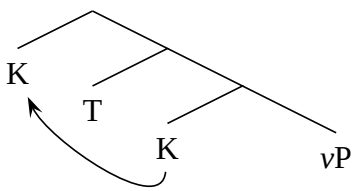
(34)



This way, Bifurcated MHD can single out featural heads if the relevant XP-YP structure is generated by way of Agree plus IM.

In Chomsky's analysis, Hypotheses 1 and 3 work in tandem in deciding what becomes the head of structures created by IM. How can Hypothesis 2 (based on atomization by Transfer) deal with cases of IM? Narita (2011a,b,c, forthcoming, to appear) maintains that Transfer applying to the moving category effectively makes the structure conform to MHD. Recall that at the completion of the subject K-phase, Transfer reduces it to an LI K. Then, EM of T to (32b)/(33c) followed by IM of K will result in the following SO.

(35)



This structure conforms to Representational *{XP, YP}, since the relevant internal merger in effect applies only to an LI K.[9]

The three hypotheses each raise quite significant empirical questions:

[6] Are traces of Internal Merge invisible?

[7] Does Transfer involve deletion of the phase-interior?

[8] Are agreement features visible at SEM?

The following sections will address the questions in this order.

[6] Are traces of Internal Merge invisible?

Clearly, Hypothesis 1 (23) rests on the positive answer to this question. What kind of rationale can be provided for the idea of trace invisibility? Chomsky suggests that IM creates a discontinuous element, namely a *chain*, and that the notion of discontinuity has something to do with the invisibility of the lower copy, *i.e.*, the tail of the chain. A more meticulous exposition of the idea can be found in Ott (2011c):[10]

(36) "[C]onsider what it means for YP to undergo IM. While IM does not manipulate the original object {XP, YP}, it does yield a configuration in which YP occupies two positions; more specifically, YP after IM is a set (chain) containing two occurrences, one within the original set {XP, YP}, and another one outside of it. This implies that after movement, discontinuous YP is no longer properly contained in the original set, since it now only contains one of YP's occurrences (but not YP as a whole). This leaves XP ... as the only remaining object properly contained in the original set, which

[9] Narita (2011a,b,c, forthcoming, to appear) goes on to propose that the atomization of the moving element is a general feature of IM: recall that IM is defined as Merge applying to two SOs α and β one of which is part of the other. If β undergoes IM to the edge of α, then α is necessarily a phrasal SO, given the very fact that it contains an occurrence of β. Thus, unless further assumptions (like Hypothesis 3) are added to UG, Representational *{XP, YP} predicts that β should be a simplex LI, which derives the consequence that only LIs can undergo IM. Narita proposes that apparent cases of phrasal-movement should be reanalyzed, along this line of reasoning, as IM of a phase-head LI that has subjected its complement to Transfer. Note that this application of Transfer renders the relevant phase-interior an island, yielding the so-called freezing effect (Culicover and Wexler 1980, Uriagereka 1999 and many others). See also (48) below.

[10] Although Ott presents the idea of headedness in terms of labeling, little seems to hinge on his assumption of labeling = projection = headedness, as far as the idea of trace invisibility is concerned. Some restatement was made in (36) to avoid terms relating to representational labels.

> is consequently labeled by XP (X) [*i.e.*, headed by the head of XP; HN]." (Ott 2011c:§4.1)

Thus, trace invisibility is related to the necessarily discontinuous nature of chain objects created by IM, as Chomsky and Ott claim.

Notice that the hypothesis of trace invisibility is coupled with the empirical claim that the higher occurrence of a chain is visible to MHD. Empirically, this idea provides room for endocentricity determined by internally merged constituents, as instantiated by wh-free relatives according to some analyses (Donati 2006, Chomsky 2008, Cecchetto and Donati 2010, Ott 2011b). Moreover, a related effect can be observed in Icelandic quirky agreement phenomena, often referred to as the *defective intervention effect*. Thus, φ-feature probing by T (or C) can 'skip' a trace of dative KP as in (37a), while it is blocked by an in-situ dative KP as in (37b):

(37) *Icelandic*: (Holmberg and Hróarsdóttir 2003:(1),(2))

a. Mér$_i$ virðast t_i [hestarnir vera seinir]
me.DAT seem.PL the-horses.NOM be slow
'It seems to me that the horses are slow.'

b. það virðist/*virðast einhverjum manni [hestarnir vera
EXPL seems/seem some man.DAT the-horses.NOM be
seinir]
slow
'It seems to some man that the horses are slow.'

Building on these facts, Chomsky claims that the visibility of chain heads and invisibility of traces are empirically motivated as general features of syntactic computation, not restricted to MHD.

However, the Icelandic data are far more complicated than the exposition just outlined. For example, it is known that while the intervention effect does not arise for traces created by A-movement, it does for traces created by wh-movement:

(38) *Icelandic*: (Holmberg and Hróarsdóttir 2003:(3))

Hvaða manni$_i$ veist þú að virðist/*virðast t_i [hestarnir vera seinir]
which man/DAT know you that seems/seem the-horses be slow
'To which man do you know that the horses seem to be slow.'

See Holmberg and Hróarsdóttir (2003, 2004), Hiraiwa (2005) among others for the relevant data. While the effect of the A/A' distinction on movement is in itself a difficult fact to capture in the theory of generalized IM (see Hiraiwa 2005, Chomsky 2008, Obata 2010, Sorida 2011, in progress), these data clearly suggest that merely

being a trace is not a sufficient condition for invisiblity, at least as far as probe-goal relations (Agree) are concerned.

Further evidence for this conclusion comes from head-movement. The standard analysis maintains that V moves to *v* in languages like English (and further to T in languages like French; see Pollock 1989), which may be partially motivated by the 'categorization' requirement, and/or by the affixal nature of *v* (see the literature of Distributed Morphology). It is clear that the movement will render the lower occurrence of V in {V, OBJ} part of a discontinuous chain object, and thus it is predicted to be invisible to MHD under Hypothesis 1. However, we have strong reasons to suppose that {V, OBJ} is headed by V, irrespective of whether V moves or not: the compositional interpretation assigned to {V, OBJ} is primarily configured by the semantic features of V (eventhood, aspect, etc.), and the object is interpreted relative to the θ-role assigned by V (typically Theme). Therefore, we have to construct a theory of MHD where V's trace can still determine headedness and compositional interpretation of {V, OBJ}. The simplest approach to this effect is to eliminate Hypothesis 1.

This counterargument to trace invisibility is based on the assumption that head-movement is syntactic. Thus, it may be circumvented by adopting Chomsky's (2001) hypothesis that head-movement is phonological (applying at PF) and has no effect at SEM (see also Boeckx and Stjepanović 2001). But we have reasons to believe that this option is theoretically ill-advised, because if we take that path, we will correspondingly lose the account of a number of syntactic properties of head-movement, such as its structure-dependence (linear-independence), minimal search requirement and phase-level application (see Berwick and Chomsky 2008, Chomsky 2010a,b for contemporary discussion). See also Lechner (2006) for instances of semantically active head-movement.

As a final remark, it is plainly uncontroversial that traces enter into various aspects of CI-interpretation, such as predicate-argument structure, operator-variable formation, and reconstruction effects with respect to binding and scope. Therefore, traces must be visible, to various degrees, for the sake of compositional interpretation. Then, so long as we pursue the null hypothesis that MHD can be entirely reduced to minimal inspection of semantic features of LIs at Interpret (12), it is simply contrary to much evidence to speculate that traces are invisible to MHD. Coupled with the fact that there is little evidence for the relevance of headedness in narrowly syntactic computation (as we saw in [3] and [4]), then, we are forced to withdraw, or reconsider the basis of, Hypothesis 1 (23), both by data and by the SMT.

[7] Does Transfer involve deletion of the phase-interior?

In his exploration of MHD, Narita (2011a,b,c, forthcoming, to appear) refrained from adopting Chomsky's hypothesis of trace invisibility (23), while he was in full agreement with the significance of MHD. As noted in [5], the approach he took was to tackle Representational *{XP, YP} solely on the basis of Hypothesis 2, repeated here.

(26) *Atomization by Transfer* (Hypothesis 2):
Transfer applying to a phase of the form {X, YP} effectively reduces it to an LI X.

Incidentally, it should be noted that the Transfer operation, as formulated in (24) repeated here, seems to involve a certain form of deletion, reducing the structural content of {X, YP}.

(24) *Transfer*:
Applied to a phase Σ, Transfer strips off all the structural information related to the interior of Σ from the active workspace of narrow syntax and maps it to SEM and PHON.

Note that after the elimination of the interior YP from the workspace of narrow syntax, no syntactic operation (IM, Agree, etc.) should apply to the elements internal to YP. This is what is often referred to as the *Phase-Impenetrability Condition* (PIC).

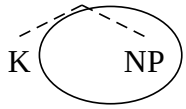
(39) *Phase-Impenetrability Condition (PIC)* (adapted from Chomsky 2000:108):
In a phase α with head H, the interior of H is not accessible to syntactic operations outside α. Only H and its edge are accessible to such operations.

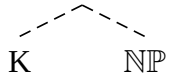
I follow Chomsky (2004, 2008) in assuming that this Transfer-as-forgetting hypothesis is the best way to characterize the PIC (see Chomsky 2008:143). The PIC is empirically corroborated by a number of observations on the locality of syntactic operations (see Chomsky 2000, Uriagereka 1999 and much subsequent work; see also Narita (2011a, forthcoming), to appear).

Readers might wonder if the deletion operation involved in Transfer should count as a violation of the *No-Tampering Condition* (NTC) (40).

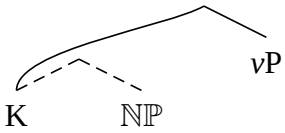
(40) *No-Tampering Condition (NTC)*:[11]
No elements introduced by syntax are deleted or modified in the course of linguistic derivation.

I would like to maintain that elimination from the active workspace of narrow syntax does not necessarily entail tampering with the relevant interior. Thus, it is perfectly possible to suppose that Transfer leaves the phase unchanged, while it only terminates further access to the interior SO. For example, Transfer may render the NP interior of a K-phase invisible for further computation in the active workspace, but the NP may occupy the complement of K throughout the derivation and reach SEM as such. That is, the derivation sketched in (33) may in fact involve the following form, where the LI K is in effect both the sister of the 'forgotten' NP and of *v*P:

(41) a. 

*v*P ⟼ b. 

*v*P ⟼

c. 

Even though it departs from conventional phrase-markers, the availability of such 'multi-dominance' structures is just a simple consequence of unconstrained Merge (Citko 2005, Kasai 2007, de Vries 2009 and references cited therein). The two partially intersecting SOs each satisfy Representational *{XP, YP} just as desired, and the interior NP reaches SEM and PHON without any loss. Thus, no violation of the NTC is implied in the approach based on Hypothesis 2.[12] If the occurrence of K in (41c) further undergoes IM, the resultant SO corresponding to (35) would be (41e).

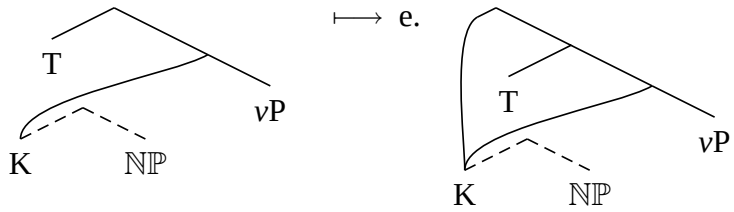
[11]The original NTC in Chomsky (2008:138), given in (i), was formulated specifically in terms of Merge:

(i) Merge of α and β leaves the two SOs unchanged (cf. Chomsky 2008).

Indeed, if we restrict our attention to the simplest formulation of Merge, namely an elementary binary set-formation, it is just an operation that does nothing more than combining SOs, and therefore it cannot modify the elements that constitute the input to this operation. The NTC as formulated in (40) is thus trivially satisfied. Here I am reformulating the NTC as a more general ban on tampering with elements introduced by syntax.

[12]It is not unreasonable to describe the state of affairs by saying that the phase-interior SO is shifted onto a 'different plane'. Cf. Chomsky (2004) for a different usage of the term 'plane'.

(41) d. ⟼ e.



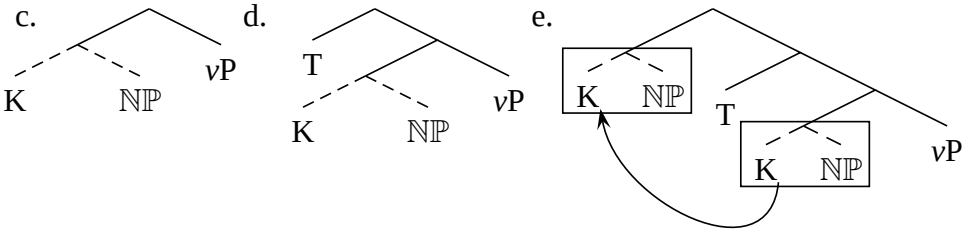
Again, no violation of the NTC is necessary.[13]

Incidentally, it is possible that we eventually find (yet undiscovered) reasons not to allow multi-dominance or structural reduction of {X, YP} to an LI X. Then, for the sake of sustaining the basic tenet of Hypothesis 2, all we need to do is to assume that what Transfer does to the phase {X, YP} is to terminate further access to YP and also to somehow let the phase-head X stand as the most prominent element *in place of* {X, YP} for the purpose of MHD. This version of Hypothesis 2 can be formulated as follows:

(42) *Atomization by Transfer* (Hypothesis 2; second version):
Transfer applying to a phase of the form {X, YP} terminates further access to YP and let the phase-head X stand as the most prominent element in place of {X, YP}.

Under this assumption, the multi-dominance representations in (41c,d,e) can justifiably be replaced with more familiar single-root diagrams in (43c,d,e).

(43)



[13]Readers might wonder why the NP interior of the K-phase is prototypically spelled-out at the moved position. I contend that NP gets interpreted as a sister of K by CI and SM (41a), prior to the application of IM (41c), which accounts for the fact that NP gets pronounced in proximity to an occurrence of K (see Narita 2011a:§4.4, 2011c). What remains to be explained is the fact that the highest occurrence is typically chosen to be pronounced in association with the NP interior (see Nunes 2004 for some related discussion). We might speculate that the choice of the highest occurrence is in part motivated by reduction of the burden of processing: all but one copy of a chain is to be erased at PHON, presumably due to minimization of computation and production (cf. Chomsky 2008:146), and processing would be further eased if the highest (which is typically the leftmost) is chosen to be pronounced, provided that the problem of finding the gap(s)/trace(s) of moved elements can rely on the directionality of left-to-right processing.

(42) is not an unreasonable assumption, and similar proposals have already been proposed in the literature. The concept of computational prominence of the phase-head LI owes its origin to Chomsky's (2000, 2008) derivational notion of 'label' (which is free from representational labeling/projection; see [1]) or Collins's (2002) notion of 'locus' in his label-free system. These notions were identified with the LI constituting the locus of the immediately preceding operation (probes triggering Agree and IM, selectors triggering EM, and so on). In the framework of phase theory where virtually all operations apply at the phase-level characterized by phase-head LIs (Chomsky 2007a, 2008; see also Narita 2011a:Ch.2), then they are prototypical instances of Chomsky's label or Collins' locus, and thus count as the most prominent element for further computations, just as proposed in (42).

The question of the nature of cyclic reduction by Transfer is related to what Boeckx and Grohmann (2007) call the 'recombination problem': according to Boeckx and Grohmann, 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. To take the case in (41) as an example, CI and SM should be able to know that the NP interior and the phase-head K were once in a sister relation with each other, even though they got Transferred separately and reach SEM and PHON in different cycles. However, the recombination problem arises only if Transfer involves deletion and separation of phase chunks. Thus, if the somewhat 'weaker' formulation of Transfer mentioned above is sufficient to capture the basic tenet of Hypothesis 2, we can safely conclude that Hypothesis 2 does not necessarily pose the recombination problem.

In short, the hypothesis of Transfer as elimination of the interior domain *from the active workspace of narrow syntax* does not entail deletion and tampering of the relevant phase structure. This avoids the recombination problem, while it also keeps to a straightforward characterization of the PIC.

[8] Are agreement features visible at SEM?

We now turn to Hypothesis 3, (27) repeated here.

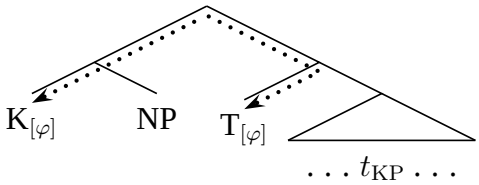
(27) *Bifurcated MHD* (Hypothesis 3):

a. Not only an LI but also a feature can be a head determined by MHD.
b. Head-inspection via MHD can be bifurcated for SOs of the form {XP, YP}.
c. If XP and YP share a feature F as the most prominent element, then the bifurcated inspection into XP and YP by MHD can single out F as

the head of {XP, YP}.

Recall the treatment of the structure involving A-movement in (34), repeated here.

(34)



According to Hypothesis 3, the XP-YP structure in (29) is effectively headed by the φ-features shared by K and T via Bifurcated MHD. The structure is created by IM of KP upon φ-feature agreement with T. Thanks to the Agree-relation holding between K and T, T's unvalued φ-features get valued by their interpretable counterparts on K. Therefore, as Hypothesis 3 goes, the matching φ-features of K and T stand as the most prominent elements in both {K, NP} and {T,. . . }, and so they can stand as the head of {{K, NP}, {T,. . . }} via Bifurcated MHD. This way, Bifurcated MHD can single out featural heads if the relevant XP-YP structure is generated by way of Agree plus IM.

The effectiveness of Hypothesis 3 crucially hinges on the assumption in (44):

(44) Agreement features (unvalued features that get valued by Agree) are visible to MHD.

For the concrete case in (34), it is critical for Bifurcated MHD to be able to 'see' the valued φ-features on T. However, note that this set of φ-features constitute a typical instance of 'uninterpretable' features, receiving no legitimate interpretation on T (tense cannot be meaningfully specified with respect to person, number or gender). Hence its presence at SEM is a violation of Full Interpretation, which should be avoided by some prior application of deletion to it. It is typically assumed that the relevant deletion applies simultaneously with the application of Agree (see Richards 2007, Chomsky 2007a, 2008, Narita 2011a:Ch.2).

(45) Unvalued features are uninterpretable and should get deleted immediately upon valuation by Agree.

The two conceptions of agreement features in (44) and (45) directly speak against each other. On the one hand, if MHD applies before Agree, then the φ-features on T are just unvalued and hence unable to determine headedness via Bifurcated MHD. They may be the most prominent elements of T in the relevant sense, but they are

certainly not shared by KP. On the other hand, if MHD applies after Agree, the φ-features are already deleted upon valuation for the sake of Full Interpretation, and hence they should become invisible to any operations including MHD.[14] Then, how can MHD 'see' the valued agreement features at any point of linguistic derivation?

In this relation, recall the discussion in [3] that there is little evidence that any syntactic operations are contingent on headedness. In particular, the most fundamental syntactic operation, namely Merge, should be able to apply essentially freely, without any requisite trigger (see [4]). The lack of any 'headedness-dependent' operation is in itself a piece of strong evidence against the view that MHD applies in narrow syntax.[15] Moreover, recall also that the barest conception of MHD reduces to efficient inspection of semantic features applying at Interpret (12).

(12) *Thesis*:
MHD reduces to the minimized inspection of semantic features of LIs at Interpret.

Both of the considerations thus point to the same conclusion, namely that MHD should apply only at Interpret, where no agreement features should remain due to Full Interpretation.

Furthermore, another aspect of (12) lies in the requirement that the inspection by MHD should be strictly restricted to *semantic* features, *i.e.*, features relevant to compositional interpretation by CI. Do φ-features on T qualify as 'semantic features', contributing to CI-interpretation in any meaningful way? Or, does it really matter to 'hit' matched φ-features in the first place, while Interpret will presumably need to inspect the entire sets of semantic features in KP and TP anyway? Again, the hypothesis in (44) bears a heavy burden of justification.

The contemporary theory of phases builds heavily on the assumption that unvalued features are 'crashingly' uninterpretable at SEM. For example, Richards (2007) and Chomsky (2007a, 2008) maintain that the requirement of valuation/deletion of unvalued features is the driving force of cyclic Transfer (synchronized with Agree/Value, according to their theory). The justification of such 'imperfections'

[14] Another possibility may be that MHD could apply 'simultaneously' with Agree/Value at the phase-level. But even if it were the case, the *prima facie* incompatibility of (44) and (45) persists, so long as we keep to the assumption (12) that MHD reduces to minimal inspection of semantic features at Interpret.

[15] Just like the lack of 'linear-dependent' operations is a strong piece of evidence against the idea that information about linear order is established in narrow syntax. See Berwick and Chomsky (2008), Chomsky (2010a,b).

as unvalued features in human language has been a long-standing problem for the minimalist program, but Chomsky (p.c.) speculates that the existence of uninterpretable/unvalued features may be rationalized, along this line of reasoning, as the device to efficiently detect and demarcate phase cycles. To the extent that cyclicity is a highly desirable feature of syntactic computation, contributing to periodic reduction of computational load, then, we may even regard the uninterpretability of unvalued features as indirectly contributing to the desideratum of computational efficiency (see Narita 2011a:Ch.2).

Therefore, if we were to pursue any version of Hypothesis 3, it would be necessary to reconsider some of the foundational hypotheses regarding the visibility of agreement features: in particular, the apparent incompatibility of (44) and (45) must be resolved somehow, and the ground for the view that agreement features have anything to do with compositional interpretation should be scrutinized.

[9] Does *{XP, YP} constrain narrowly syntactic derivations or apply only at SEM?

We saw that MHD immediately leads to the prediction called Representational *{XP, YP} (repeated here):

(22) *Representational* *{*XP, YP*}:
Headless SOs of the form {XP, YP} are ruled out by FI at SEM.

Two options are proposed by Chomsky to circumvent the problem posed by (22), Hypothesis 1 (trace invisibility) and Hypothesis 3 (Bifurcated MHD). However, recall that these hypotheses are based on empirically questionable premises, invisibility of traces and visibility of valued agreement features to MHD, respectively.

Now, suppose that we decide not to adopt these stipulations for the sake of (22), barring reference to chain properties of IM and uninterpretable features valued via Agree. Without IM and Agree, then, it would be very difficult to imagine how MHD can assign headedness to structures of the form {XP, YP}. A further constraint is that we must still keep to the Inclusiveness Condition and the NTC, thus no insertion and deletion of elements are available in the course of syntactic derivations. Then, it seems reasonable to conclude that any structures of the form {XP, YP} are simply intolerable for MHD and hence ruled out by Full Interpretation. As a consequence, then, it simply follows from MHD that only derivations that are free from any XP-YP structure would yield convergent results. That is, merger of two phrases should be precluded altogether. This constraint may be referred to as *Derivational* *{*XP,*

YP}:

(46) *Derivational* *{*XP, YP*}:
Merger of two phrases is unavailable.

This constraint is 'derivational' in the sense that it strongly affects the mode of Merge-application: it effectively requires that Merge can generate only structures of the form {H, α}, where H is an LI and α is an SO (lexical or constructed). That is, each application of Merge is required to take at least one LI as its input, a restriction that Narita (2011a,b,c, forthcoming, to appear) calls the *H-α schema*.

(47) *The H-α schema*: Merge(H, α) $\rightarrow$ {H, α}.
Each application of Merge must take at least one LI as its input.

This constraint, equivalent to Derivational *{XP, YP}, simply follows from Representational *{XP, YP}, barring stipulations concerning chains and agreement features.[16]

Narita (2011a,b,c, forthcoming, to appear) claims that Derivational *{XP, YP} is a sustainable conclusion if we still allow recourse to Hypothesis 2 (26), namely atomization of phases by Transfer. Indeed, Merge can effectively combine XP and YP while satisfying Derivational *{XP, YP}, if at least one of the phrases is reduced to an LI via Transfer *prior to* the application of Merge. Then, while Derivational *{XP, YP} rules out derivations like (30) and (32), it still rules in derivations like (31) and (33).

Narita maintains that the interplay of Derivational *{XP, YP} and Hypothesis 2 can derive a number of empirically favorable consequences. For reasons of space I can only enumerate the list of the results Narita has alluded to his H-α schema, so readers are referred to the papers cited below for details of the analysis:

(48) If we adopt Derivational *{XP, YP} and Hypothesis 2, then:
a. The complementarity of ph(r)asal-movement and head-movement is straightforwardly derived (see Narita 2011a,b; see note 9).
(i) Only phases can undergo 'XP'-movement (viz. movement of a phase head that has subjected its complement to Transfer); that is, non-phase phrases can never undergo 'XP'-movement.
(ii) Non-phase-head LIs can only undergo X^{o}-movement.

[16]Incidentally, Kayne (2011) stipulated a constraint equivalent to Derivational *{XP, YP} as an axiom of syntax. Of course, Kayne can also entertain the same hypothesis that the effect of Derivational *{XP, YP} can be deduced from MHD.

(iii) Remnant movement of XP is impossible if the head X has moved out of XP (Takano's (2000) generalization; see Narita 2011b).

b. The H-α schema and the PIC yields a unifying account of Huang's (1982) CED, even with better empirical coverage (Narita 2011a, forthcoming, to appear).
   (i) All moved ph(r)ases constitute islands (the freezing effect; Culicover and Wexler 1980, Uriagereka 1999 and many others).
   (ii) CED effects arise for moved subjects but not for in-situ subjects.
   (iii) CED effects arise for 'high' adjuncts adjoining above *v* but not for 'low' adjuncts adjoining below *v*.

c. The H-α schema strongly contributes to the typology of prosodic major phrases and I(ntonational)-phrases (Narita and Samuels 2009; see also Samuels 2012:Ch.4)

d. The H-α schema accounts for the distribution of free positions in phrasal idioms (external arguments, indirect objects, and possessors apparently involve {XP, YP} structures and are not possible parts of phrasal idioms) (Narita 2011b).

e. MHD is 'failure-proof'; *i.e.*, all SOs are efficiently assigned headedness via MHD (Narita 2011a, forthcoming).

f. The H-α schema provides a strong analytic means for the acquisition of LIs (Narita 2011a, forthcoming; see [11] below).

To the extent that these results receive support from empirical data, the strength of Derivational *{XP, YP} is corroborated accordingly. Any approach that departs from Derivational *{XP, YP} will thus need to reconsider what the empirical bases for these generalizations are.

In passing, I would like to comment on Ott's (2011c) evaluation of Narita's approach based on Derivational *{XP, YP}. In his attempt to defend Hypothesis 1, he notes,

(49) "This view of free and unrestricted application of Merge contradicts recent proposals by Kayne (2011), Zwart (2011), and Narita ([2011a]), among others. ... According to these authors, the generation of {XP, YP} structures is barred altogether; however, this result is achieved by stipulation. ... Unless some principled reason is provided for intrinsic constraints on Merge, and hence for the corresponding enrichment of UG, such proposals should be met with skepticism."

As Ott correctly points out, Narita's proposal (among others) contradicts Ott's representational approach to *{XP, YP}, in that only the former effectively precludes Merge from applying to two phrases.[17] However, it is not the case that Narita puts forward Derivational *{XP, YP} as an 'intrinsic constraint on Merge'. Rather, Narita (2011a) was devoted to defending the hypothesis that the effect of Derivational *{XP, YP} could be straightforwardly derived from independent principles of UG, relating to the total absence of representational labels/projection and the reduction of head-detection to minimal search for LIs (MHD). In order to derive the effect of Derivational *{XP, YP} from Representational *{XP, YP}, it simply suffices *not* to incorporate any stipulations on (in)visibility of chains and agreement features, keeping to a simpler theory of UG. Thus, Narita's approach is perfectly compatible with the theory of unconstrained Merge: the claim along this line is that Merge can in principle combine any two SOs, but only certain choices will receive appropriate endocentric interpretation at SEM.

Moreover, it is not clear whether Narita's approach is more stipulative than Ott's approach based on Hypothesis 1. Ott's approach adopts Representational *{XP, YP} *with* the stipulation of trace invisibility (23), while Narita's approach adopts Representational *{XP, YP} *without* that very stipulation, straightforwardly yielding the effect of Derivational *{XP, YP}. This line of deduction of Derivational *{XP, YP}, entertained by Narita (2011a, forthcoming) and based solely on MHD, hence stays unaffected by Ott's criticism. Thus, we can still hope for incorporating the major empirical import of Derivational *{XP, YP} into the minimal theory of head-detection, while keeping closely to the SMT.[18]

Each analysis has its own strength and weakness, aiming at different empirical coverages, so a direct comparison would not be very informative at the current stage of understanding. Only considerations from further empirical data will be able to foster the reconciliation of these two prospective approaches.

[10] Are symmetric structures locally unstable?

Hypothesis 1 and Hypothesis 3 both face some fundamental problems concerning the visibility of traces and agreement features, as we reviewed in [6] and [8]. However, it should nevertheless be acknowledged that these proposals were put forward

[17] As for Kayne (2011), see note 16.

[18] However, I agree with Ott's assessment as for the other of the two deductions of the H-α schema entertained by Narita. As noted in note 5, it speaked to the 'Merge-as-attraction-to-the-head' hypothesis (14) and its EF-based reformulation (18) entertained by Chomsky (2008), which should be eliminated in favor of the theory of unconstrained Merge.

by Chomsky to capture some important insights about linguistic structure.

First, behind Hypothesis 1 is the following idea, essentially originating from Moro's (2000) 'dynamic antisymmetry':

(50) SOs of the form {XP, YP} are locally unstable, requiring 'symmetry-breaking' IM of either XP or YP.

Recall that Hypothesis 1 resolves the problem of Representational *{XP, YP} by moving either XP or YP and letting the other phrase determine the head, assuming trace invisibility. For example, Chomsky (2008:160, note 34; lectures at MIT in fall 2010) maintains that (50) recaptures the essence of Moro's (2000) analysis without recourse to Kayne's LCA.[19] Thus, assuming that small clauses are symmetric {DP, DP} structures, obligatory movement of either one of the DPs as shown in (51) can be regarded as a representative example of (50).

(51) *Italian*: (Moro 2007)

a. [molte foto del muro]$_i$ sono [t_i [la causa della rivolta]]
many pictures of-the wall are the cause of-the riot
'many pictures on the wall are the cause of the riot.'

b. [la causa della rivolta]$_i$ sono [[molte foto del muro] t_i]

c. *sono [[molte foto del muro] [la causa della rivolta]]

See also Ott (2011a,c) for an extension of this kind of analysis to German split topicalization.

Moreover, recall the discussion of EPP-movement in [5]. EM of the subject KP and the verbal phrase {*v*, {V, OBJ}} results in an unstable XP-YP structure (28), but A-movement of the subject KP will stabilize it by the invisibility of the subject's trace, according to Hypothesis 1. Along this line of reasoning, Chomsky hints the possibility that the obligatoriness of so-called EPP-movement can be partially reduced to the effect of (50).[20]

Further, what seems to be the most important consequence of (50) is the insta-

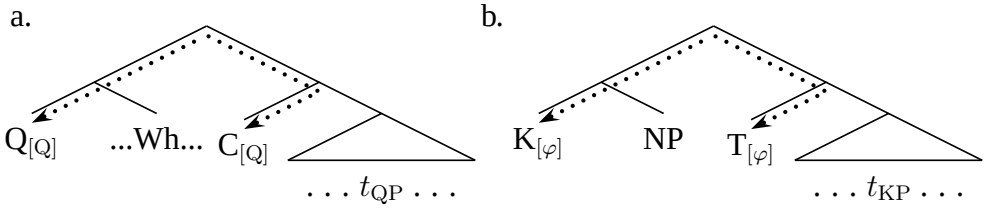
[19] See Narita (2011a:§4.2; to appear) for the argument that the LCA has no place in label-free syntax (see already Uriagereka 1999; Kayne 2009:note 8).

[20] As noted in [5], IM of {*v*, {V, OBJ}} should also be able to remedy the instability of the relevant XP-YP structure, but such movement seems unattested in natural languages, presumably because some aspect of θ-role assignment by *v* to the subject is contingent on *v* heading the relevant part of the structure (see Narita 2011a:Chs.4-5 for some discussion). Incidentally, Chomsky (2011) also explores the possibility of extending his mechanism of head-detection in such a way that movement of OBJ from {*v*, {V, OBJ}} may also stabilize (28) (cf. Alexiadou and Anagnostopoulou's (2001) subject-in-situ generalization). I will put this proposal aside.

bility of intermediate occurrences of successive cyclic movement. For example, if a wh-object undergoes successive cyclic movement to the edge of matrix CP, as in *Which picture$_i$ did you t_i think t_i (that) John t_i saw t_i yesterday?*, any intermediate occurrence of the wh-phrase should necessarily be part of an XP-YP structure, and hence qualify as locally unstable, according to Hypothesis 1. Hence, the wh-phrase must keep moving. The fact that the wh-phrase can never stay in intermediate positions thus follows directly from (50), a highly desirable result of Hypothesis 1.

Under Chomsky's analysis, the ultimate landing site of IM is instead characterized by Hypothesis 3 (Bifurcated MHD). Thus, successive cyclic Wh-movement should end at the edge of C that agrees with the wh-phrase with respect to Q-features (52) (cf. Cable's 2010 argument that wh-phrases are headed by a covert functional category Q; see also Narita 2011a, forthcoming), and A-movement (which can also be successive cyclic) should end at the edge of T that agrees with the moving KP with respect to φ-features ((34) repeated here as (52b)).

(52) a. b.



Importantly, the theory of Bifurcated MHD is articulating the following intuition (cf. Fukui 2011).

(53) SOs of the form {XP, YP} that are 'featurally symmetric' are stable.

Indeed, (53a) and (53b) are salvaged by symmetrically distributed matching features: [Q] for (53a), [φ] for (53b).

Interestingly enough, then, Chomsky's theory is putting forward two contrastive notions of symmetry: it holds that certain structures require IM due to the local instability of the XP-YP symmetry, while IMed SOs are incorporated into other {XP, YP} structures via the symmetry of matching features. In this hypothesis, then, symmetry is characterized on the one hand as something syntax marks intolerable, and on the other as something syntax is rather looking for. The dichotomy between the two kinds of symmetry is a worrisome complication of UG, and hence to be adopted only if forced by empirical evidence.

In this relation, I would like to point out that we can rather eliminate (50) while incorporating the major effect of (50) into (53). Note that the set of XP-YP structures inducing symmetry-breaking IM seems to correspond to the set of

featurally asymmetric {XP, YP} structures: the external argument and the verbal phrase share no matching features, and thus presumably count as featurally asymmetric; and XP and YP in small clauses are often analyzed as involving no direct feature-matching/Agree (see Moro's 2000 counterargument to the AgrP analysis of small clauses). Thus, we may suggest that local instability arises not because of the symmetry at the level of LI/phrase distinction, but rather because of the asymmetry at the level of features.

The idea that syntax looks for structural symmetry seems to have a certain conceptual appeal. Importantly, the characterization in (53) seems to hinge little on the notion of semantic headedness, thus it may be that the principle of symmetry-orientation in (53) is in its nature syntactic, arising for the sake of simplicity and economy (see also Fukui 2011).

Does syntax regard symmetric structures intolerable? Or does it rather look for them? I have to leave this important question for future research, but note that any proposal that attempts to deal with these issues will have to provide a theory of stable {XP, YP}'s and unstable {XP, YP}'s. The following is the list of empirical domains that would be relevant to the inquiry.

(54) In what environment does {XP, YP} count as locally (un)stable?
- a. small clauses (unstable?)
- b. {[external argument], {*v*, ...}} (unstable?)
- c. intermediate positions of successive-cyclic movement (unstable)
- d. subject-predicate {KP, {T, ...}} (stable)
- e. {WhP, {C, ...}} (stable)
- f. object shift (stable?)
- g. merger of indirect object (stable?)
- h. merger of possessor (stable)
- i. adjunction (stable)
- j. coordination (stable)
- k. {X, Y} (?)

Notice that the notion of unstable {XP, YP} is a difficult thing to capture in the framework of Derivational *{XP, YP} (but see note 21). As noted above, this approach precludes formation of {XP, YP} in the first place. Then, SOs will always take the form {X, YP} and can be dealt with by MHD without any problem. Therefore, everything should be stable in terms of headedness, contrary to facts about successive cyclic movement, for example. Therefore, to the extent that it is plausible to assume the existence of 'endocentrically unstable' {XP, YP} structures,

advocates of Derivational *{XP, YP} (notably Narita 2011a,b,c, forthcoming, to appear and Kayne 2011) will have to provide certain alternative characterizations of such SOs.[21]

[11] What are the 'atoms for computation'?

The theory of bare phrase structure presumes that Merge is the necessary device for structure-generation in human language. Moreover, in order to let any sort of Merge-based computation work at all, there must also exist some finite set of 'atomic elements' that serve as input to Merge-based recursive structuring but are not created by Merge. Throughout the present article, I have used the term 'lexical item' (LI) to refer to the 'atomic' elements of computation that are stored in the mental lexicon. Each LI may further contain some formal features that differentiate one LI from the other in their functions and distributions in syntactic derivation, as well as some other intrinsic features that can contribute to interpretation either at CI or SM (semantic and phonetic features).

The atomicity of LIs also crucially figures in the mechanism of MHD. MHD is understood as efficient inspection of primitive elements that are relevant to interpretation, and the discussion in this paper has been based on the assumption that the atoms for Merge, namely LIs, are also atoms for MHD. That is, we are subscribing to the following equation:

(55) Elements stored in the lexicon (LIs) = atomic elements for Merge
= atomic elements for head-detection.

This equation, though traditional, is open to scrutiny. Consider, *e.g.*, Hypothesis 3. It proposes that not only LIs but also features of LIs can determine headedness. This hypothesis clearly goes against (55), in that it denies the atomicity of LIs for MHD. We cast some doubt on the premises that Hypothesis 3 was based on, but whatever the ultimate fate of Hypothesis 3 may be, it may well be a fruitful line of research to disentangle the three concepts encapsulated in (55). I will continue to adopt (55), but it should be noted that no known principle of computation requires (55), so it would be an important empirical discovery if corroborated by empirical considerations.

It is uncontroversial that there exist LIs that are stored in the lexicon and constitute input to narrow syntax. That said, we really don't know what the internal

[21] Note that Derivational *{XP, YP} still rules in symmetric structures of the form {X, Y} (54k), where both of the constituents are LIs. Thus, Narita (and in fact Kayne 2011, too) may pursue the hypothesis that endocentrically unstable structures (if any) are actually of the form {X, Y}.

composition of LIs can be. How 'complex' are LIs? Does UG specify any constraint on the possible composition of LIs? How can the composition of each LI be learned/acquired by the child through experience? Opinions vary as to what the smallest unit of linguistic computation is, or where the computation 'bottoms out'. Linguists have proposed all sorts of different conceptions of the lexicon, but there is no sign that the controversy will ever find a point of agreement. However, despite linguists' continuous failure to pin down a reasonable and agreeable set of universal linguistic primitives, human infants still acquire the lexicon of one or another I-language with remarkable speed and uniformity, a familiar poverty-of-the-stimulus fact. How could this ever be possible? If there is indeed some set of primitives that are readily accessible to infants, why are trained adult linguists bound to fail to find the slightest trace of them?

Narita (2011a,c, forthcoming) claims that the H-α schema (Derivational *{XP, YP}) might hint at a clue to this learnability problem ((47) is reproduced here).

(47) *The H-α schema*: Merge(H, α) $\rightarrow$ {H, α}.
Each application of Merge must take at least one LI as its input.

Consider, for example, recent experimental results discussed in Yang and Gambell's work (see Yang 2002, 2004, Gambell and Yang 2003 among others), according to which the general mechanism of statistical data analysis (presumably a constituent of the third factor of FL design) provides a reasonable first-cut segmentation of words in primary linguistic data, when it works in tandem with the principle (presumably determined by UG) that each phonological word bears a single primary accent. Given the relative ease of detecting phonological words in primary data (say, the three phonological words in /ðəˈboɪzˈkistəˈgərl/, *the-bóys kíssed a-gírl*), it seems reasonable to suppose that children acquire these readily detectable units as the first provisional candidates for LIs. From there, the H-α schema provides a preliminary analysis of sentential structures comprised of these words, synthesizing them in accordance with the H-α schema (e.g., {*the-bóys*, {*kíssed*, *a-gírl*}} or the like).

In addition to such a 'bottom-to-top' application, the H-α schema might also give a clue to a 'top-down' decomposition of words to smaller units, as Narita (2011a,c, forthcoming). For example, the phonological word *the-bóys* may be eventually analyzed as {the, bóys}, further to {the, {boy, s}}, and even to {the, {-s, {*n*, $\sqrt{boy}$}}} under the guidance of the H-α schema. Each of such H-α schema-based reanalyses of phonological word structures may lead to a corresponding revision of the list of provisionally analyzed LIs. For example, a learner of English may start

with the provisional list of unanalyzed phonological words as his first lexicon, say {*the-bóys*, *kíssed*, *a-gírl*, ... }, but continuous revisions will be made to this list as the acquisition proceeds.

Various cues from experience (distributional or semantic) may be taken as evidence for such decomposition. Among other things, the H-α schema predicts that any instance of a movable element should involve a topmost phase-head LI, sometimes covert: for example, provided with the H-α schema, the distribution of nominals, some of which involves A- or A'-movement, might be sufficient to indicate that there is some category that appears at the topmost edge of the relevant nominal structure, namely K. Whereas learners of languages like Japanese will identify it with overt case-particles (like *o* 'ACC' and *ga* 'NOM'), learners of languages like English may assign a morpheme with zero phonetic content to this category. Arguably, such functional categories are relatively easy to detect via overt evidence, such as morphological manifestation and movement. So to speak, they satisfy what Fukui and Sakai (2003:327; see also Thráinsson 1996) calls the *visibility guideline for functional categories*, which holds that a functional category has to be visible (i.e., detectable) in the primary linguistic data, thus simplifying the problem of learnability.

Narita (2011c) further speculates that Cinque's (1993) null theory of nuclear stress might be of great advantage for the process of H-α schema-based decomposition. In simplified terms, Cinque's theory holds that the most deeply embedded element has the primary stress in the structure. If the H-α schema independently guarantees that each of the constituent structures is analyzed as {H, α}, then the Cinquean nuclear stress rule always assigns the primary stress to α. Therefore, the stressed-unstressed distinction will be made to correlate with the XP/X°-distinction, which greatly facilitates the H-α schema-based decomposition of linguistic structures, not only at the sentential- and phrasal levels but maybe also at the word-internal level, too (recall that Cinque's null theory was proposed as a unifying account of stress patterns at these levels).

In this respect, the H-α schema is both a very 'soft' and a very 'hard' constraint on the possible form of linguistic structure. It is quite 'soft' in the sense that it allows both leftward and rightward branching of the H-α structure, as in, e.g., {the, {{*n*, $\sqrt{lecture}$}, yesterday}} in English, or {{that-GEN, {yesterday-GEN, {*n*, $\sqrt{lecture}$}}}, ACC} (*so-no kinoo-no koogi-o*) in Japanese. More importantly, the analysis parses the relevant H-α branching structure without involving any extra notions like 'head-of', 'adjunct-of', or 'specifier-of'. These projection-based notions are just unnecessary and stipulative residue of the earlier phrase structure

grammar that are simply unavailable in projection-free bare phrase structure. On the other hand, the H-α schema is so 'hard' a constraint that it disallows any vacant structural slots like unfilled 'specifiers' and 'complements', which cannot be excluded from X-bar-theoretic phrase structures. Moreover, it strongly restricts the possible instances of 'displacement' in natural languages, and it instructs the child that any apparent instances of 'XP'-movement should involve cyclic application of Transfer, specifically triggered by (probably covert) phase-heads that encapsulate Transferred domains for phonological purposes (see note 9). This way, the H-α schema provides a strong restriction on the space of possible syntactic structures, and hence a strong bias for the initial linguistic analysis entertained by the child. The problem of explanatory adequacy is correspondingly simplified, approaching a principled explanation of human language.[22]

In this approach, the atomicity of LIs is regarded not as an unanalyzable axiom of UG but rather as a consequence of the H-α schema-based analysis of lexical entries. This seems to me to be a very important conclusion, in light of the unresolved controversy on the set of linguistic primitives.

To the extent that this line of analysis can be characterized only in terms of the H-α schema, namely Derivational *{XP, YP}, we may conclude that Derivational *{XP, YP} leads to a more promising picture of lexical acquisition than other approaches to Representational *{XP, YP}. In that the H-α schema/Derivational *{XP, YP} results straightforwardly from *not* adopting visibility stipulations on traces and agreement features, then, considerations from lexical acquisition, among many others listed in (48), may constitute reasons to be cautious about the scope of these unwarranted complications of UG.

[12] Is endocentricity universal?

The earlier PSG-based conception of phrase-markers holds that each phrase is associated with labels, and the X-bar theory further maintains that all labels are projections of head LIs. Then, under the assumption that projection imposes headedness, the X-bar-theoretic approach in effect subscribes to the *universal endocentricity hypothesis*:

(56) *The universal endocentricity hypothesis*:
Every phrase is headed by an LI.

[22] See also Narita (2011b) for the idea that the H-α schema also constrains the possible structure of idiom chunks (phrasal 'listemes').

(56) has become a standard assumption since the advent of the X-bar theory, followed by the majority of subsequent theories in the generative framework. However, it should be noted that once PSG-based syntax is replaced with the theory of Merge, the concept of universally labeled phrase structure is correspondingly exorcised from the framework of bare phrase structure. (56) thus loses its theorematic status, and it becomes subject to empirical investigation to evaluate its validity. Does (56) receive real support from empirical data, or should it be regarded as an unwarranted residue of PSG that is to be discarded as well?

It is worth noting in conjunction with this that Narita's theory of the H-α schema (Derivational *{XP, YP}), while successfully eliminating recourse to representational labels/projection, still subscribes to the universal endocentricity hypothesis, in that it still rests on the assumption that head-assignment via MHD is an absolute and inviolable requirement. However, Noam Chomsky (p.c.) raised a question concerning Narita's (2011a) projection-free but still universally endocentric conception of phrase-markers. Thus, regarding the long-standing universal endocentricity claim (56), he commented:

(57) "That's true of X-bar theory and its descendants, but not earlier approaches. Thus the central PSG rule was S $\rightarrow$ NP VP, non-endocentric. That may well be correct, given impossibility of SPEC in a pure Merge system allowing free merge of SOs."

He further notes in his comment on Narita (2011a):

(58) "Also, it's a highly theory internal decision to call the result of IM endocentric, also probably wrong (indirect questions vs successive cyclicity, for example)."

Indeed, putting forward the idea of universal endocentricity (56) has been "a highly theory internal decision," essentially originating in X-bar theory, but not required for earlier approaches and also for the contemporary framework of bare phrase structure. Especially for SOs of the form {XP, YP} that were traditionally described as involving specifiers, we no longer have any strong reason to stick to headedness stipulations (recall the discussion in [4]). Therefore, regardless of whether they are created by EM or IM, any {XP, YP} structures are now open to non-endocentric characterizations.[23]

[23]It is worth noting that Chomsky first put forward the X-bar schemata in *Remarks on Nominalization* (Chomsky 1970) in such a way that they could be interpreted not as strict formats for PSRs but as a kind of evaluation measure that merely sets preference for (unmarked) X-bar-theoretic

Note that Hypothesis 3 (27) characterizes the subject-predicate structure {KP, TP} as, in a sense, 'doubly headed' by φ-features, owing to the probe-goal agreement between KP and TP. This feature-based conception of headedness certainly departs from the traditional conception of labeling/projection, in that the latter maintains that only *an LI* (not features) can determine headedness of an SO. As noted in [8], the hypothesis that agreement features should be visible to MHD is problematic, in that the barest form of MHD is just the minimal inspection of semantic features at Interpret (recall (12)). However, the very idea of bifurcated search, embedded in Hypothesis 3, has a certain conceptual appeal, in that it provides room for non-endocentric compositional interpretation of *{XP, YP} while still keeping to the minimal theory of head-detection. Then, whatever the ultimate fate of Hypothesis 3 may be, the existence of bifurcated inspection of semantic features seems to be a matter of course. Little is known about the formal properties of such inspection, but the conceptual barrier that the stipulation of labeling/projection has been imposing on us is now removed in the theory of bare phrase structure. Inquiry into the nature of compositional interpretation of nonendocentric structures appears to be a potentially quite fruitful research topic, to my eyes.

Concluding Remarks

This paper was an attempt to highlight some of the contemporary issues surrounding the notion of headedness and compositionality in bare phrase structure. It enumerated several important research questions pertaining to these issues, which readers can review by skimming the headings of the sections above. I deliberately leave many of the questions open for future research, trying to present the prospects for a minimal theory of head-determination from various viewpoints. None of the issues are resolved, and undoubtedly there will come numerous others untouched in this paper or unrecognized at the current stage of inquiry. Only time will tell if any of the cursory remarks expressed in this contribution are to be justified.

Acknowledgments:

I am grateful to Cedric Boeckx, Noam Chomsky, Koji Fujita, Naoki Fukui, Hidehito Hoshi,

projections, leaving the possibility of (marked) nonendocentric structures. I thank Naoki Fukui for bringing this point to my attention.

C.-T. James Huang, Takaomi Kato, Yasuhiko Kato, Masakazu Kuno, Terje Lohndal, Dennis Ott, Bridget Samuels, Masanobu Sorida, Mihoko Zushi, and the audience of the 26th annual meeting of the Sophia University Linguistic Society for the enlightening discussions that led to various revisions and extensions of this paper. I am solely responsible for all the remaining inadequacies.

References

Alexiadou, Artemis, and Elena Anagnostopoulou. 2001. The subject-in-situ generalization and the role of case in driving computations. *Linguistic Inquiry* 32:193–231.

Berwick, Robert C., and Noam Chomsky. 2008. 'poverty of the stimulus' revisited: Recent challenges reconsidered. ms. MIT. A short abstract appeared in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Washington, DC, p. 383.

Bittner, Maria, and Ken Hale. 1996a. The structural determination of Case and agreement. *Linguistic Inquiry* 27:1–68.

Bittner, Maria, and Ken Hale. 1996b. Ergativity: Toward a theory of a heterogeneous class. *Linguistic Inquiry* 27:531–604.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Boeckx, Cedric. 2008. *Bare syntax*. Oxford: Oxford University Press.

Boeckx, Cedric. 2010a. Defeating lexicocentrism. ms. ICREA/UAB. Downloadable at lingBuzz/001130.

Boeckx, Cedric. 2010b. A tale of two minimalisms: Reflections on the plausibility of crash-proof syntax, and its free-merge alternative. In *Exploring crash-proof grammars*, ed. Michael Putnam, 105–124. Amsterdam: John Benjamins.

Boeckx, Cedric, and Kleanthes K. Grohmann. 2007. Putting phases into perspective. *Syntax* 10:204–222.

Boeckx, Cedric, and Sandra Stjepanović. 2001. Head-ing toward PF. *Linguistic Inquiry* 32:345–355.

Cable, Seth. 2010. *The grammar of Q: Q-particles, wh-movement, and pied-piping*. Oxford: Oxford University Press.

Cecchetto, Carlo, and Caterina Donati. 2010. Relabeling heads. a unified accout for relativization structures. ms. University of Milan-Bicocca and University of Rome-La Sapienza. lingBuzz/001014.

Chomsky, Noam. 1955/1975. *The logical structure of linguistic theory*. ms. Harvard University, 1955. Published in part in 1975, New York: Plenum.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, ed. Roderick A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, MA: Ginn.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1994. Bare phrase structure. *MIT Occasional Papers in Linguistics* 5:Department of Linguistics and Philosophy, MIT. Reprinted in G. Webelhuth, ed. (1995), Malden: Blackwell, *Government and Binding Theory and the Minimalist Program*, 383-439.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond: The cartography of syntactic structures*, ed. Adriana Belletti, volume 3, 104–131. New York: Oxford University Press.

Chomsky, Noam. 2005. Three factors in the language design. *Linguistic Inquiry* 36:1–22.

Chomsky, Noam. 2007a. Approaching UG from below. In *Interfaces + recursion = language? Chomsky's minimalism and the view from semantics*, ed. U. Sauerland and H.-M. Gärtner, 1 –29. Berlin and New York: Mouton de Gruyter.

Chomsky, Noam. 2007b. Biolinguistic explorations: design, development, evolution. *International Journal of Philosophical Studies* 15(1):1–21.

Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory*, ed. Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.

Chomsky, Noam. 2010a. Introduction to the collection ed. by Naoki Fukui. ms. MIT.

Chomsky, Noam. 2010b. Poverty of stimulus: Unfinished business. Lecture presented at Johannes-Gutenberg University Mainz, March 24, 2010.

Chomsky, Noam. 2011. Problems of projection. Paper presented at the CUNY Linguistics Colloquium, the City University of New York, February 17, and at the Leiden University Centre for Linguistics (LUCL), March 14, 2011.

Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24:239–298.

Citko, Barbara. 2005. On the nature of Merge: External Merge, Internal Merge, and Parallel Merge. *Linguistic Inquiry* 36:475–496.

Collins, Chris. 2002. Eliminating labels. In *Derivation and explanation in the minimalist program*, ed. Samuel David Epstein and T. Daniel Seely, 42–64. Oxford: Blackwell.

Culicover, Peter W., and Kenneth Wexler. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Donati, Caterina. 2006. On wh-head-movement. In *Wh-movement: Moving on*, ed. Lisa L.-S. Cheng and N. Corver, 21–46. Cambridge, MA: MIT Press.

Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima, and Hisatsugu Kitahra. 1998. *A derivational approach to syntactic relations*. Oxford: Oxford University Press.

Fukui, Naoki. 2011. Merge and bare phrase structure. In *The Oxford handbook of linguistic minimalism*, ed. Cedric Boeckx. Oxford: Oxford University Press.

Fukui, Naoki, and Hiromu Sakai. 2003. The Visibility Guideline for Functional Categories: Verb raising in Japanese and related issues. *Lingua* 113:321–375. Reprinted in Fukui (2006).

Fukui, Naoki, and Yuji Takano. 1998. Symmetry in syntax: Merge and Demerge. *Journal of East Asian Linguistics* 7:27–86. Reprinted in Fukui (2006).

Gambell, Timothy, and Charles D. Yang. 2003. Scope and limits of statistical learning in word segmentation. In *Proceedings of NELS 34*, 29–30.

Graffi, Giorgio. 2001. *200 years of syntax: A critical survey*. Amsterdam: John Benjamins.

Hiraiwa, Ken. 2005. Dimensions of symmetry in syntax: Agreement and clausal architecture. Doctoral Dissertation, MIT.

Holmberg, Anders, and Throbjörg Hróarsdóttir. 2003. Agreement and movement in Icelandic raising constructions. *Lingua* 113:997–1019.

Holmberg, Anders, and Throbjörg Hróarsdóttir. 2004. Agreement and movement in Icelandic raising

constructions. *Lingua* 114:651–673.

Hornstein, Norbert. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. MIT.

Jackendoff, Ray. 1977. *X'-syntax*. Cambridge, MA: MIT Press.

Kasai, Hironobu. 2007. Multiple dominance in syntax. Doctoral Dissertation, Harvard University.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kayne, Richard S. 2009. Antisymmetry and the lexicon. *Linguistic Variation Yearbook* 8:1–31. Reprinted in Anna Maria Di Sciullo and Cedric Boeckx (eds.), *The biolinguistic entreprise: New perspectives on the evolution and nature of the human language faculty*. Oxford: Oxford University Press.

Kayne, Richard S. 2011. Why are there no directionality parameters? In *Proceedings of WCCFL 28*, ed. Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz, 1–23. Somerville, MA.: Cascadilla Proceedings Project.

Lasnik, Howard, Juan Uriagereka, and Cedric Boeckx. 2005. *A course in minimalist syntax: Foundations and prospects*. Oxford: Blackwell.

Lechner, Winfried. 2006. An interpretive effect of head movement. In *Phases of interpretation*, ed. Mara Frascarelli, 45–70. Berlin: Mouton de Gruyter.

Lohndal, Terje. in progress. Without specifiers: Phrase structure and events. Doctoral Dissertation, University of Maryland, College Park.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

McGilvray, James. 1998. Meanings are syntactically individuated and found in the head. *Mind and Language* 13:225–280.

McGilvray, James. 2002. MOPs: the science of concepts. In *Belief and meaning: Essays at the interface*, ed. Wolfram Hinzen and Hans Rott, 73–103. Frankfurt: Ontos.

McGilvray, James. 2009. Introduction to the third edition. Noam Chomsky (2009) *Cartesian linguistics: A chapter in the history of rationalist thought, Third Edition*. Cambridge: Cambridge University Press, pp.1-52.

Moro, Andrea. 2000. *Dynamic antisymmetry*. Cambridge, MA: MIT Press.

Moro, Andrea. 2007. Some notes on unstable structures. ms. Universita San Raffaele, Milano.

Narita, Hiroki. 2009. Full interpretation of optimal labeling. *Biolinguistics* 3:213–254.

Narita, Hiroki. 2011a. Phasing in Full Interpretation. Doctoral Dissertation, Harvard University (Downloadable at *http://ling.auf.net/lingBuzz/001304*).

Narita, Hiroki. 2011b. The H-α schema, the complementarity of ph(r)asal/head-movement, Takano's generalization, and the distribution of free positions in phrasal idioms. Paper presented at GLOW in Asia: Workshop for Young Scholars, Mie University, September 7-8, 2011.

Narita, Hiroki. 2011c. Phasing in linearization of unlabeled phrase structure. Paper to be presented at the 29th Annual Meeting of the English Linguistic Society of Japan, Workshop on the Syntax-Phonology Interface, Niigata University, November 12-13, 2011.

Narita, Hiroki. forthcoming. Phase cycles in service of projection-free syntax. In *Phases: developing the framework*, ed. Ángel J. Gallego. Mouton de Gruyter.

Narita, Hiroki. to appear. *{XP, YP}, the independence of the LCA and antisymmetry, and the LCA-free phase-based account of the CED effect. To appear in Balazs Suranyi, ed., *Minimalist Approaches to Syntactic Locality*.

Narita, Hiroki, and Bridget Samuels. 2009. The H-α schema and phonological ph(r)asing. Paper

presented at the Formal Approaches to the Phonology-Morphology-Syntax Interfaces workshop at ConSOLE XVIII, Universitat Autònoma de Barcelona (UAB), December 16, 2009.

Neeleman, Ad, and Fred Weerman. 1999. *Flexible syntax: A theory of case and arguments*. Dordrecht: Kluwer.

Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.

Nunes, Jairo, and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3:20–43.

Obata, Miki. 2010. Root, successive-cyclic and feature-splitting Internal Merge: Implications for feature-inheritance and Transfer. Doctoral Dissertation, University of Michigan.

Ott, Dennis. 2009. The conceptual necessity of phases: Some remarks on the minimalist enterprise. In *Explorations of phase theory: Interpretation at the interfaces*, ed. K. K. Grohmann. Mouton de Gruyter.

Ott, Dennis. 2010. Grammaticality, interfaces, and UG. In *Exploring crash-proof grammars*, ed. Michael Putnam, 89–104. Amsterdam: John Benjamins.

Ott, Dennis. 2011a. Local instability. Doctoral Dissertation, Harvard University. Published online in *GAGL (Groninger Arbeiten zur germanistischen Linguistik)*: *http://irs.ub.rug.nl/dbi/4e8027471fa04*.

Ott, Dennis. 2011b. A note on free relative clauses in the theory of phases. *Linguistic Inquiry* 42:183–192.

Ott, Dennis. 2011c. Symmetric Merge and local instability: Evidence from split topics. ms. University of Groningen. To appear in *Proceedings of ConSOLE XIX*.

Pesetsky, David. 1982. Paths and categories. Doctoral Dissertation, MIT.

Piattelli-Palmarini, Massimo, Juan Uriagereka, and Pello Salaburu, ed. 2009. *Of minds and language: A dialogue with Noam Chomsky in the Basque Country*. Oxford: Oxford University Press.

Pietroski, Paul M. 2005. *Events and semantic architecture*. Oxford: Oxford University Press.

Pietroski, Paul M. 2008. Minimalist meaning, internalist interpretation. *Biolinguistics* 2:317–341.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry* 20:365–424.

Richards, Marc D. 2004. Object shift, scrambling, and symmetrical syntax. Doctoral Dissertation, University of Cambridge.

Richards, Marc D. 2007. On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry* 38:563–572.

Samuels, Bridget. 2012. *Phonological architecture: A biolinguistic approach*. Oxford: Oxford University Press.

Samuels, Bridget. forthcoming. Consequences of phases for morphophonology. In *Phases: developing the framework*, ed. Ángel J. Gallego. Mouton de Gruyter.

Seely, T. Daniel. 2006. Merge, derivational c-command, and subcategorization in a label-free syntax. In *Minimalist essays*, ed. Cedric Boeckx, 182–217. Amsterdam: John Benjamins.

Sheehan, Michelle. 2010. Labelling, multiple spell-out and the final-over-final constraint. In *Incontro di Grammatica Generativa XXXV 2010*, ed. Vincenzo Moscati and Emilio Servidio. Siena.

Sheehan, Michelle. in press. The resuscitation of CED. In *NELS 40 Proceedings*.

Sorida, Masanobu. 2011. Cyclic Transfer, chains, and the A/A' problem of Japanese scrambling. Paper presented at the Chains in Minimalism conference, Yokohama National University, February 12-13, 2011.

Sorida, Masanobu. in progress. Inclusiveness: Its implications for chain formation. Doctoral Dis-

sertation, Sophia University, Tokyo.
Stowell, Tim. 1981. Origins of phrase structure. Doctoral Dissertation, MIT.
Takano, Yuji. 2000. Illicit remnant movement. *Linguistic Inquiry* 31:141–156.
Thráinsson, Höskuldur. 1996. On the (non-)universality of functional categories. In *Minimal ideas: Syntactic studies in the minimalist framework*, ed. Werner Abraham, Samuel D. Epstein, Höskuldur Thráinsson, and Jan-Wouter Zwart, 253–281. Amsterdam: John Benjamins.
Uriagereka, Juan. 1999. Multiple Spell-Out. In *Working minimalism*, ed. Samuel David Epstein and Norbert Hornstein, 251–282. Cambridge, MA: MIT Press.
de Vries, Mark. 2009. On multidominance and linearization. *Biolinguistics* 3.4:344–403.
Wujastyk, Dominik. 1982. Bloomfield and the Sanskrit origin of the terms 'exocentric' and 'endocentric'. *Historiographia Linguistica* 9:179–184.
Yang, Charles D. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
Yang, Charles D. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8:451–456.
Zwart, Jan-Wouter. 2011. Structure and order: asymmetric merge. In *The Oxford handbook of linguistic minimalism*, ed. Cedric Boeckx, 96–118. Oxford: Oxford University Press.