Syntax from above: Towards a top-down, phase-theoretic model of sentence generation

Doug Merchant
University of Georgia

Abstract The incremental derivations of phase theory (Chomsky 2001, 2005: etc.) raise the question of the directionality of derivations (den Dikken 2018). In this squib, I outline some arguments for top-down generation, arguing that grammatical theory must treat as axiomatic the notion that derivations mirror actual structure-building processes in the brain; given the limited nature of temporary memory capacity (Cowan 2015), bottom-up generation does not survive long under such scrutiny. Next, I briefly review some theory-internal evidence (Phillips 1996, 2003, Chesi 2007, 2015) with respect to conflicting results from constituency tests. Finally, I seek to characterize the phases with respect to what Chomsky (2005, 2007) calls the 'duality of semantics', arguing that roots are inserted in the upper phase (CP) for discourse purposes, but are interpreted with respect to the core proposition in the lower phase (vP).

1   Introduction

Although the general properties of syntactic computation are by now fairly well understood, two key features of the model of the grammar developed in the generative tradition are not as well elucidated: the directionality of derivations, and the timing of the insertion of phonological form. The range of possible answers to these questions is fairly limited; indeed, the choice appears to be binary in both cases. For directionality, it's either top-down or bottom-up, and for insertion, it's late or early (before narrow syntax, or at SPELLOUT). Given that the options are so few, it's a bit mysterious that we have no conclusive answers as of yet to these fundamental questions.

To some extent, this may be because the grammar seems to function fairly well with either setting for both of what one might call 'meta-parameters' (i.e., parameters not set human-language by human-language, but rather set the same way for all human languages). In the case of the directionality meta-parameter, the question was rarely even raised until relatively recently, the first detailed discussion

being Phillips (1996; see also Phillips 2003, Phillips & Lewis 2013, Chesi 2007, 2015, Zwart 2009, 2015, den Dikken 2018). One reason for this is that in early approaches, derivations were conceived of as being assembled in one fell swoop, and subsequently adjusted by 'singular transformations' (e.g., the passive; Chomsky 1955, 1957), these transformations were eventually recast as 'movement'. As the theory developed, it became apparent that almost all cases of such movement were bottom-up.[1] This was not initially a question of the directionality of derivations *per se*, but rather a question of the directionality of the adjustments necessary to transform the results of the structure-building processes of the base into the linearized, language-specific structures that are represented by actual spoken sentences. (In the parlance of the day, it was a question of the directionality of the displacement of lexical items from 'deep structure' into 'surface structure' configurations.)

Nonetheless, the bias resulting from the empirical observations about 'upward' movement was present when, with the advent of the Minimalist Program (MP; Chomsky 1993, 1995), derivations were newly conceived of as an incremental series of MERGERS. In fact, it is the precise formulation of MERGE (which I take here from Chomsky 2001: 3) as an operation which 'takes two syntactic objects α and β and forms the new object $\gamma = \{\alpha, \beta\}$' that implements this bias in the MP. But note that an isomorphic formulation of this procedure is also available; I provisionally characterize it as follows:

(1)   BRANCH (aka Bizarro Merge): Take a syntactic object γ and form two new objects α and β, where $\gamma = \{\alpha, \beta\}$.

These two variants of Merge differ only in terms of the directionality of the derivation; they produce the same structures.[2] Indeed, for most syntactic phenomena, it seems not to matter much which setting we assume for the directionality meta-parameter (exceptions will be noted in due course). As a result, many in the field are essentially agnostic on this point. This attitude may or may not be

---

[1]There were of course analyses involving 'downward' movement that persisted for some time, e.g., Chomsky's 1957 original account of affix-lowering in the English aspect system, still used (e.g.) in the first edition of Carnie's *Syntax* textbook (Carnie 2002).

[2]Associated processes will look slightly different as well. For example, the specifics of the labelling algorithm involved (Chomsky 1995: 243–246; Rizzi 2016) will be different, as a selected XP will need to 'project' its label downwards in an endocentric manner, as in principle either α or β could inherit the label of γ. I thank an anonymous reviewer for raising this point.

warranted from a theory-internal perspective, but it has the undesirable effect of shielding the field from prosperous encounters with disciplines which would otherwise be natural allies. In particular, agnosticism about directionality prevents us from formulating testable predictions concerning the psychology and neuroscience of language production,[3] a predicament that will only intensify as these areas advance. We've got to get this straight, or linguistics runs the risk of ending up an island set apart from the rest of cognitive science.[4]

In this paper, I outline some arguments for top-down structure building, and make a provisional attempt to sketch out what phase theory might look like under such a system. In section 2, I contrast bottom-up Merge and top-down Branch, demonstrating that Merge is incapable of generating structures with open positions, where a speaker has begun speaking a sentence but has not yet decided how to complete it. Indeed, Merge becomes even more untenable when we consider the modern consensus on the nature of temporary memory capacity, where only a limited number of 'chunks' can be handled at one time (Cowan 2001, 2015, Baddeley et al. 1987, Baddeley & Wilson 2002). Next, I briefly review some theory-internal evidence (Phillips 1996, 2003, Chesi 2007, 2015) with respect to conflicting results from constituency tests. In section 3, I discuss proposed implementations of top-down structure building, and seek to characterize the phases with respect to what Chomsky (2005, 2007) calls the 'duality of semantics', arguing that roots are inserted in the upper phase (CP) for functional and discourse purposes, but are interpreted with respect to the core proposition in the lower phase (vP). Finally, in section 4, I speculate on how to account for WH islands in a top-down system, and conclude.

2   Arguments for top-down structure building

2.1   The grammar as an implementation-dependent abstraction

I will argue here that a theory of linguistic competence must include a theory of real-time sentence generation, and thus go beyond the abstraction of the idealized

---

[3]Language comprehension is a different animal, with multiple pathways potentially available to decode a message. Indeed, some, e.g. Townsend & Bever (2001), argue for multiple pathways operating in parallel, the results of which are then compared ('analysis-by-synthesis'). Much work has of course been done on sentence processing, but a good deal less on sentence production, in part because of the issues involved here.

[4]For another layer to this metaphor, see the discussion of syntactic islands in section 4.

speaker, 'unaffected by such grammatically irrelevant conditions such as memory limitations...and errors (random or characteristic) in applying his knowledge of this language in actual performance' (Chomsky 1965: 3). This performance/competence distinction has long been used as a bulwark against the necessity of accounting in the grammar for every utterance of every speaker; as speech errors are frequent in actual performance, this is a desirable result. However, there is a difference between extending the grammar by incorporating these errors on the same basis as any other utterance, and using them to assess what sorts of grammatical processes could have produced such an error. Consider the following:

(2)   a. *That group of boys are playing soccer in the field.

       b. That group of boys is playing soccer in the field.

In (2a), we have a lack of phi-feature agreement between *group* and *are*; this is known in the literature as an ATTRACTION ERROR (Franck et al. 2006). Certainly we should not extend the grammar of English upon discovering such an utterance. However, we can use such an error as an assessment tool; in this instance, to investigate the functioning of the operation AGREE. The existence of this error suggests that 'boys' is in the search space of the probe, and given the appropriate hiccough in computation, is mistaken in (2a) as the subject.[5]

So although our theory need not incorporate (2a) into the grammar (i.e., it should not be a generable sentence under our model), it should be able to explain the proximate cause of the error. Furthermore, the theory should predict that (2a) is a possible error, and that (2c) is (probably) an impossible one:

(2)   c. *That group are playing soccer in the field.

We can make a similar case with respect to memory limitations. Chomsky has taken such concerns a bit more seriously in his later work:

> Proceeding further, MI [Minimalist Inquiries, i.e. Chomsky 2000]
> proposes another reduction of computational burden: the deriva-
> tion of Exp [an expression] proceeds by phase, where each phase

---

[5]There is probably a semantic element at play here, as well, as the boys are in fact playing individually as well as in a group. One would probably not make the error '*The association of lawyers were dissolved'.

is determined by a subarray LA of LA$_i$ [Numeration], **placed in ‘active memory’** (Chomsky 2001: 11–12; my emphasis).

The discussion of the relevance of ‘memory limitations’ to the grammar generally revolves around two issues: the apparently unbounded character of sentences in terms of both length and complexity, and the difficulty of parsing certain permutations of sentences.[6] The grammar need not distinguish between what is easily parsable and what is more difficult to parse, nor between singly-embedded sentences and sentences embedded *ad infinitum*. So as we should not extend the theory to incorporate speech errors such as (2a), we also need not constrict the grammar in order to bar infinitely long sentences.

However, to the extent that grammatical theory is a theory of how sentences are generated by the human brain, the question of ‘computational burden’ cannot be irrelevant. In fact, it drives much of the current work attempting to satisfy the ideal of the Strong Minimalist Thesis, i.e., the notion of language being ‘an optimal solution to interface conditions... an optimal way to link sound and meaning’ (Chomsky 2005: 3). From this perspective, grammatical theory is an abstract model of a neurological process, one that interfaces with other neurological processes, and which likely makes use of more general cognitive processes, even if it cannot be reduced to them. So as with speech errors, we can use what we know of these processes to assess whether our theory is on the right track.

Phillips & Lewis (2013: 14) put this question in terms of whether we consider grammatical theory to be an IMPLEMENTATION-DEPENDENT or an IMPLEMENTATION-INDEPENDENT abstraction. An abstraction is implementation-independent if it can be ‘realized in different ways by multiple lower-level systems’. Arithmetic is a good example; the quantity ‘twelve’ can be expressed as the result of a number of different procedures: as the product of 2×6, the product of 3×4, the sum of $5 + 7$, the square root of 144, etc. Furthermore, these procedures can be implemented by a computer, by the human brain (perhaps to a lesser extent by other animal brains), by weighing objects the weight of which is known, and so on. Arithmetic is thus implementation-independent.

---

[6]The paradigm of which is the centrally-embedded sentence, e.g., ‘Birds cats fleas bite chase fly’. The difficulty of parsing these quite valid sentences is explicable in terms of extralinguistic perception principles on which parsing supervenes (Bever 1970); this phenomenon would presumably not apply to sentence generation.

Human language is another matter. Despite great strides in artificial intelligence, we are far from implementing its full functionality in any medium save the organic. And it is probably not controversial that every sentence of the language has a unique derivation;[7] there is no analogy here to the variable procedures that all have the quantity *twelve* as a result. Our abstraction, then, is implementation-dependent. To the extent that grammatical theory aims to be a model of sentence-generation, then, it must model the process by which a human brain generates a sentence, not any other process that can generate that sentence. Otherwise, it is not a model of language *per se*, but merely a model of something *like* language.[8]

2.2   Branch v. Merge

The question of the interaction of working memory with the grammar, then, is a real issue, and not just limited to parsing. The proper setting of the directionality meta-parameter will in part turn on it. Given this backdrop, let us rephrase the question. Does the human brain construct incremental derivations from the bottom-up, or top-down? We speak from left to right,[9] more or less with the least deeply embedded terminals spoken first and the most deeply embedded spoken last, but it's not unthinkable that perhaps we construct these sentences 'backward', at least in simple sentences like (3)

(3)   Boys like dogs.

However, multiply-embedded sentences replete with long-distance dependencies are par for the course in human language; the following is probably not at the tail end of the distribution:

(4)   [The boy]$_i$ is likely t_i to be very happy t_i to have arrived t_i.

---

[7]'Ambiguous' sentences, e.g. *I killed the king with the knife*, are not in fact an exception to this generalization, as they are only ambiguous from the perspective of comprehension. In the mind of the speaker, who will always have one meaning or the other in mind, they are unambiguous.

[8]This conclusion is correct even if the results of the derivations are the same (i.e. if the same sentence is generated). The argument will be that top-down derivations are *a priori* preferable to bottom-up derivations, whether or not they cannot be distinguished empirically; they should thus be granted the status of the null hypothesis.

[9]Strictly speaking, *left*, *right*, *top*, and *bottom* are arbitrary vectors, but of course they are used here with their normal implications: *left* meaning temporally preceding, *bottom* meaning most deeply embedded, etc.

The bottom-up model predicts that the matrix subject in (4) reaches its Spell-out position after three separate displacements;[10] for it to be pronounced, therefore, the entire derivation must first be constructed. This is a rather severe deficiency of Merge; under this model, one cannot begin speaking until the derivation is complete. This is because the cyclic transfer to the interfaces associated with phase theory proceeds from the bottom-up; the structures of the higher phases are not spelled out and sent to the phonological component until this process is complete for all the phases below. Even if the structures associated with the lower phases can be 'forgotten' to some extent,[11] the phonological forms associated with terminals must be retained in their linear order. Figure 1 illustrates the series of applications of Merge necessary before the subject in (4) can be pronounced; note that this does not happen until step 12.

	Merge
1)	the boy
2)	arrived the boy
3)	have arrived the boy
4)	to have arrived the boy
5)	the boy to have arrived ~~the boy~~
6)	happy the boy to have arrived ~~the boy~~
7)	be happy the boy to have arrived ~~the boy~~
8)	to be happy the boy to have arrived ~~the boy~~
9)	the boy to be happy ~~the boy~~ to have arrived ~~the boy~~
10)	likely the boy to be happy ~~the boy~~ to have arrived ~~the boy~~
11)	is likely the boy to be happy ~~the boy~~ to have arrived ~~the boy~~
12)	the boy is likely ~~the boy~~ to be happy ~~the boy~~ to have arrived ~~the boy~~

Figure 1   A derivation involving incremental applications of Merge

Under phase-theoretic Branch, however, a speaker can begin to speak more or less as soon as the derivation begins, once elements from Numeration are in-

---

[10]Assuming cyclic movement and the Unaccusative Hypothesis (Perlmutter 1978) are both correct, of course; if only the former is correct, only two displacements are necessary.

[11]Of course, they cannot be forgotten completely if we take reconstruction effects to be both psychologically real and postsyntactic. den Dikken (2018: 65–66), building on arguments in Epstein & Seely (2002), points out that for uninterpretable features such as Case, SPELLOUT must be able to access information for displaced WH- elements from lower phases (for the purposes of valuation), and such the grammar cannot fully 'forget' this information.

serted into the structure and the highest phase is complete.[12] Once the elements of the higher phase have been transferred to the interfaces, they can be forgotten completely (although copies of such will still need to end up in theta positions for the purposes of propositional interpretation; see section 3.3). Figure 2 illustrates a series of operations of Branch, producing the same structure as does Merge in Figure 1. The subject is available for transfer as soon as the first phase boundary is encountered, immediately after step 1.[13]

Branch

1)   the boy
2)   the boy is
3)   the boy is likely
4)   the boy is likely ~~the boy~~
5)   . . .

Figure 2   A derivation involving incremental applications of Branch.

Note that the observation that an entire structure must be built by Merge before any of it is pronounced comes with a strong prediction: speakers will never (in fact, cannot!) begin a sentence without being precisely sure of how they intend to end that sentence. Indeed, sentence generation that depends exclusively on Merge for structure building predicts this phenomenon to be impossible. This flies in the face of intuition. For illustration, consider the following hypothetical exchange:

(5)   Speaker A: Is there any way you could lend me some money?
       Speaker B: Well, how much do you. . .
       (*Speaker B considers whether to ask Speaker A how much he wants, or*

---

[12]Assuming functional projections (CP, TP, vP) whose specifiers are on the left. For languages with right-headed TPs (e.g., German), the T head may be spelled out before the lower phase is built, but it is linearized to the right of the vP, so some sort of purely phonological buffer may be required.

[13]Technically, this is not true for defective v in raising structures like (4), where v is a 'weak' phase head (Chomsky 2001) and so Spellout isn't triggered until the complement of 'likely' is added after step 3. However, it's conceivable that for weak phases, the transfer to the PF interface might happen at the same juncture as in a strong phase, with the transfer to the semantic interface happening separately.

whether it would be more prudent to ask how much he needs.)
Speaker B: . . . need?

The contrast between Branch and Merge here is quite stark. Bottom-up Merge simply cannot generate a structure with an open position like this, in which the higher phase has been spelled out and sent to the PF interface, but in which the Numeration of the lower phase has not even been established. The decision to utter *want* or *need* must happen before the speaker begins speaking.

Branch, on the other hand, handles this effortlessly. Setting aside the details of *do*-support the lexical sub-array of the higher phase in (5) would contain just *how much* and *you*. This much is spelled out upon completion of the higher phase, and available to be spoken. Next, the sub-array associated with the lower phase is enumerated. At this point, of course, the speaker may pause, considering first one verb and then another before making a selection.

Note also that Branch requires only a limited number of elements in the work-space at a time, whereas Merge requires the entire Numeration as well as the entire lower structure to be dragged along (even if spelled-out). From the perspective of cognitive psychology, retaining this many words is extremely unlikely; the modern consensus on the nature of temporary memory capacity is that only a limited number of unrelated 'chunks'—perhaps only 3 to 5—can be handled at one time (Cowan 2001). When these chunks are composed of words in a sentence, the capacity increases, perhaps to 15–16 words (Baddeley et al. 1987, Baddeley & Wilson 2002). But it's not hard to find sentences much longer than that; the following has 65 words:

(6)   So the Trump administration ended up relying on, first, staff associated with Mike Pence, who had deep ties to the anti-abortion movement - and then also think tanks like the Heritage Foundation, this very conservative group in Washington, D.C., that had fallen out of some favor and really got a big break by attaching to the Trump campaign as kind of the dark horse candidate. (Reporter Dan Diamond, from an interview on NPR's *Fresh Air*, Jan. 25, 2018)

In light of the extensive psychological literature on this subject (for a recent review, see Cowan 2015 and references therein), it is probably not putting it too strongly to say that it is impossible for the above sentence to have been constructed in real time by a bottom-up derivation. We are left with purely bottom-up deriva-

tions as models not of human sentence-generation, i.e. not of language itself, but as a model that only approximates this capacity; a model, again, of something *like* language. Top-down derivations, on the other hand, are well positioned to generate such sentences.

2.3   Theory-internal evidence for top-down structure-building

The theory-external arguments for Branch over Merge, then, are quite robust. The question remains, however, whether any theory-internal evidence is available. Given the identical formulation of Branch and Merge as they stand (save for directionality), any given sentence is in theory generable by either. The only theory-internal evidence likely to be available, then, would come from tests which apply to derivations *in medias res*, targeting intermediate structures.

One example of such a test is the coordination test for constituency: if a string of words can be coordinated with another similar string, that string is generally thought to represent a constituent. Phillips (1996, 2003) examines the conflicting structural diagnoses provided by different constituency tests in some detail. Although most constituency tests point to verbs and objects forming a constituent to the exclusion of the subject, he notes that the coordination test shows conflicting results. Applying this test in (7b), it seems that the verb in (7a) forms a constituent with the object. However, the application of this test in (7c) suggests instead that it forms a constituent with the subject:

(7)   a.   John roasted the turkey.

     b.   John [[roasted the turkey] and [steamed the broccoli]].

     c.   [[John roasted] and [Mary carved]] the turkey.

This is of course puzzling; how can *roasted* form constituents both with the object to the exclusion of the subject, as in (7b), and with the subject to the exclusion of the object, as in (7c)?

Complicating the picture even further, Chesi (2007: 5) observes that in SOV languages like Japanese, Subj + Obj sequences can be coordinated to the exclusion of the V:

(8) *[[Mary-ga   ringo-o   2-tu] to [Nancy-ga   banana-o   -bon]]*
     [[Mary-NOM apple-ACC 2-CL] and [Nancy-NOM banana-ACC 3-CL]]
     tabeta (koto).
     ate

Mary ate two apples, and Nancy three bananas.

The issue in the English example in (7c) is that the subject and verb appear
to have been coordinated before the object is attached; the issue in the Japanese
example in (8) is that the subject and object have been coordinated before the
verb is attached. If constituents are built incrementally bottom-up, we have no
way to capture such a structure; the bracketed elements in both sentences do not
at any point in the derivation form a constituent. To get out of this quandary, we
must either assume multiple structural representations operating in parallel (as in,
e.g. Pesetsky 1995, which seems rather antithetical to the spirit of the SMT), or
stipulate that (7c) is merely a case of disguised clausal coordination (along the
lines of Ross 1985: 105–107), with the first occurrence of *the turkey* deleted.[14]

However, if constituents are built incrementally top-down, we can identify an
intermediate point of the derivation where the verb has been attached, but the
object has not yet been spelled out:

(9) Step 1: John
     Step 2: [John roasted]
     Step 3: [[John roasted] and [Mary carved]][15]
     Step 4: [[John roasted] and [Mary carved] [the turkey]]

In Step 4, the subject-verb constituency is destroyed; they no longer form a
constituent to the exclusion of the object (note the second closing bracket moving
from immediately after *carved* to after *turkey*). However, the coordinate structure

---

[14]Note that the Ross-style solution is even less practical for the Japanese data in (8), in which
it is the first occurrence of *tabeta* 'eat' that is deleted. This contrasts with the (idiomatic) clausal-
coordination translation in English, where the second occurrence of *eat* is deleted. See also Sato
(2009), who analyzes similar structures in Japanese as coordinated vP/VPs.

[15]The specifics of coordinate structure, i.e., whether it is ternary- or binary-branching, I leave
to the side. Note also that 'Mary carved' is syntactically complex, and in a bottom-up approach
would have to be constructed in a separate workspace (see e.g. Nunes & Uriagereka 2000), but in
a top-down approach can be constructed in parallel with 'John roasted', i.e., without requiring a
separate tree.

has already been built, so the issue of its constituency is moot. This constituency test applies only at an intermediate point in the derivation, one in which the S-V string has temporary constituency. An essentially identical solution, *mutatis mutandis*, is available for (8); the Subj and Obj in a Japanese sentence form a temporary constituent which can be coordinated with another S-O string before the V is added.

3   Theoretical characteristics of top-down derivations

3.1   Previous approaches

Considering the theory-external arguments for Branch over Merge, I believe a strong case exists for according top-down derivations the status of the null hypothesis. The question, then, is whether they're workable, and whether we can import decades of grammatical theory into this (apparently) radical new framework without discarding the baby with the bathwater. I will tread lightly here, noting that detailed proposals exist elsewhere: Phillips (1996, 2003) and Chesi (2007, 2015) both describe top-down systems, with Phillips (1996) focusing on identifying the parser with the grammar, as well as on arguments from conflicting constituency tests (see section 2.3 above) and the 'Pesetsky paradox',[16] and Chesi focusing on deriving successive cyclicity from the use of a Last In First Out (LIFO) memory buffer, as well as on subject and adjunct islands and parasitic gaps.

Recently, den Dikken (2018: chap. 2) has assessed both of these models critically and in an extended fashion (as well as proposing his own, an Agree-based approach), and I refer the reader to that work for details. The differences between the machinery involved in den Dikken's proposal and Chesi's appear to be relatively minor, however, and unlike Phillips (2003), both attempt to incorporate phase theory into top-down structure building. In what follows, I focus on characterizing the interpretive division-of-labor-by-phase, connecting it to what Chomsky (2005: 8) calls the 'dual semantics' of the CONCEPTUAL-INTENTIONAL (C-I)

---

[16]Pesetsky (1995) describes sentences such as 'John gave books to them in the garden on each other's birthdays', where the anaphor is found in an adjunct presumably adjoined high in the VP and is therefore unbound. From these sorts of observations he proposes that all VPs have two parallel structural descriptions, and that binding takes place in a 'cascading' structure in which the adjunct is lower in the VP. However, as *give the books to them* can be fronted to the exclusion of the adjuncts, it presumably forms a constituent, and so a standard layered structure is needed as well.

system, before sketching out a version of phase theory with top-down derivations, one that should be more or less functionally compatible with den Dikken's and Chesi's models, but which hews a bit more closely to mainstream assumptions in phase theory.

3.2 Phasal layers: The duality of semantics

I argued in the introduction that it is the precise formulation of Merge as a procedure which 'takes two syntactic objects α and β and forms the new object $\gamma = \alpha, \beta$' (Chomsky 2001: 3) which implements the bias towards bottom-up structure building in the MP. But it does this (implicitly) in concert with another pillar of grammatical theory: Baker's 1988 Uniformity of Theta Assignment Hypothesis (UTAH), in which "identical thematic relationships between items are represented by identical structural relationships between those items at the level of D-structure" (ibid., p. 46). In combining these two central features of the language faculty, we end up with a picture in which lexical items are inserted into standard structural positions within the tree; these positions uniformly assign thematic roles to DPs, which are then subsequently displaced to higher positions where necessary for functional (EPP,[17] Case) or discourse-related reasons (Questions, Force, Focus), or for the purposes of scope. Generally speaking, it is these higher positions which 1) are language-specific and/or 2) interface with discourse-level features; the lower positions, or theta-positions, are (presumably) universal in both function and interpretation.[18]

   A satisfactory division of labor follows arises from this. I submit that the vP layer is the locus of interpretation for the propositional core, consisting of the verb and all the arguments associated with the verb.[19] This is what I will call

---

[17]I.e., the Extended Projection Principle (Chomsky 1982), or the requirement that the subject position of every clause be filled.

[18]At least at the level of generality represented by, e.g. Dowty's 1991 proto-roles.

[19]A reviewer points out a complication with this generalization: in some theories, key features of interpretation are located outside the vP. Bianchi & Chesi (2014) argue that thetic subjects are interpreted in their theta positions at the interface, but that categorical subjects are interpreted in their derived position in the higher layer. The sort of 'interpretation' that they seem to have in mind for categorical subjects, that of a presupposition of existence, seems to me to additive with respect to the simple proposition; rather than being part of the interpretation of the proposition, the presupposition situates the proposition with respect to the discourse. Considering their (8a), 'Firemen are altruistic', the presupposition of the existence of firemen is the discourse-linked component, and the predication of altruism is the propositional component. Thus even if Bianchi and Chesi

the SIMPLE PROPOSITIONAL LAYER (SPL) of the sentence. The higher shell, which is represented in the default case by CP and TP, consists of second-order information, in which the proposition is situated in time and modality and in the discourse context of the actual utterance of the proposition; I will refer to this as the FUNCTIONAL/DISCOURSE LAYER (FDL).

In order to characterize these layers, I adopt the following definitions:

(10) a. SIMPLE PROPOSITION (SP): An interpretive complex consisting of a verb and its arguments, non-situated with respect to time and/or modality, in a non-language-specific configuration.

      b. COMPLEX PROPOSITION (CP): An interpretive complex consisting of a simple proposition, situated with respect to time and/or modality, in a language-specific functional configuration.

In essence, a simple proposition, in which cognitively basic concepts[20] are arranged with respect to one another in a limited number of (non-linearized) configurations, is embedded into a layer which situates the simple proposition with respect to the larger discourse context, and thereby the force and modality of the proposition. To illustrate, consider (11a–e):

(11) a. The President$_i$ T_{pres} [$_{vP}$ t_i eats steak with ketchup].

      b. The President$_i$ T_{past} [$_{vP}$ t_i ate steak with ketchup].

      c. The President$_i$ T_{pres} should not [$_{vP}$ t_i eat steak with ketchup]!

      d. Which President$_i$ T_{pres} [$_{vP}$ t_i eats steak with ketchup]?

      e. [Steak with ketchup]$_j$, the President$_i$ T_{pres} [$_{vP}$ t_i eats t_j]!!!

The basic proposition common to all these sentences is simply *[Det] President [eat] steak with ketchup*, and can be interpreted as such in all cases because of the uniformity of the structural positions in which the theta roles are assigned. In a strictly semantic sense, of course, sentences like (11a) and (11b) represent different propositions, as they have different truth conditions; one may be true

---

are entirely correct, I do not believe this to be an exception to the generalization that the *simple proposition* is interpreted within vP.

   [20]Presumably in the sense of Fodor (1998), although broaching that discussion here would take us afield.

while the other is false. By 'simple proposition', I refer merely to the non-time-situated proposition they have in common, i.e., '[Det] President [eat] steak with ketchup'. In (11a–e), this simple proposition is differently situated by the elements of the FDL: with respect to time by the Tense heads, with respect to modality in (11c) by the modal *should*, with respect to the discourse context in (11d–e) by the displacement of arguments into discourse positions, and with respect to language-specific features (here the EPP) by the displacement of the Agent into the derived subject position.[21]

The two domains of interpretation here are reminiscent of Chomsky's 2005, 2007 discussions of the 'duality of semantics', and how Internal (IM) and External Merge (EM) correlate with the two aspects of interpretation, propositional and non-propositional:

> [T]he two kinds of Merge should be expected to yield different interface properties. That is obviously true at the SM interface—the ubiquitous property of "displacement"—and appears to be true at CI as well. The two types of Merge correlate well with the duality of semantics ... EM yields generalized argument structure, and IM all other semantic properties: discourse-related and scopal properties. The correlation is close, and might turn out to be perfect if enough were understood. (Chomsky 2007: 10)

The Simple Propositional and Functional/Discourse layers I've characterized above correlate with EM and IM, in that (in a bottom-up theory), External Merge is responsible for introducing items into their theta positions, i.e., 'generalized argument structure', and Internal Merge is responsible for displacement to discourse-related, functional, and scopal positions. In other words, the duality of semantics is represented not just in terms of the operations involved in structure building, but in the interpretive import of the two layers of the clause, whereby simple propositions become complex propositions.

---

[21]Indeed, displacing an element for language-specific functional reasons is in a certain sense displacing it because of a 'discourse context': the context of the language in question. For an Irish-English bilingual, displacing the Agent into spec, TP occurs because the context is 'English' (with presumably English-speaking interlocutors), whereas moving a verb to adjoin to T in Irish occurs because the context is 'Irish'. Of course, these displacements are generally mandatory, whereas Topicalization, WH- movement and the like are optional.

3.3   Towards a top-down, phase-theoretic model of sentence generation

Given the fundamentally different nature of these two layers with respect to inter-
pretation, it is not surprising that a wealth of phenomena point to a computational
break between the construction of one and the construction of the other. Since
Chomsky (2001), these layers are known in this context as PHASES, and the com-
putational break between them is the PHASE BOUNDARY. Phases are constructed
incrementally, by recursive applications of Merge, and spelled out cyclically. The
interface between phases is known as the PHASE EDGE; this consists of the heads
and specifiers of phase-triggering projections: CP and vP.[22] When a phase edge is
complete, everything but the edge is spelled out, i.e., submitted to the SENSORI-
MOTOR (SM) and CI systems for articulation and interpretation, respectively. At
this juncture, the construction of another phase begins.

I demonstrate in this section that essentially this same infrastructure can be
adopted in a top-down derivation; the key differences will be in what counts as an
insertion site, and in the nature of syntactic movement. In this system, the FDL
is the locus of root *insertion*,[23] and the SPL the locus of root *interpretation*. In
the FDL, the insertion sites include 1) discourse positions (for DPs, spec, CP; for
V, C^0), and 2) language-specific, functional positions (for DPs, spec, TP; for V,
T^0). Roots are inserted from a lexical sub-array (LA) chosen phase-by-phase; the
sub-array must be exhausted for a phase to be spelled out.[24] Note that the roots
inserted in these sites are not interpreted in these sites, but rather in the SPL. I
refer to such roots as PROPOSITIONAL ELEMENTS:

---

[22]Chomsky (ibid.) distinguishes between transitive *v**, and defective *v*; the latter is not a
(strong) phase head. Except where the distinction is relevant and so noted, I use here *v* for both *v*
and *v**.

[23]Except for internal arguments, which in the default case are inserted in the SPL, either as
complements to the verb or in an AgrO or Asp projection (for theories in which accusative Case
is assigned there). By *root insertion*, I mean taking an element from the lexical array and inserting
it into a node in the tree. I assume these roots are abstract in the sense that they have syntactic
and semantic features but no phonological content, with phonological exponents added only at
SPELLOUT. However, see e.g. Harley (2014) for arguments that they cannot have semantic fea-
tures, either.

[24]See Richards (2011) for related arguments. Richards argues in part for an approach in which
the notion 'phase head' is epiphenomenal, and phases are defined in terms of the exhaustion of an
LA.

(12) *Propositional element* (PE) : the main verb of a clause, or one of its arguments, or a null copy of either.

PEs are inserted along with THETA-FEATURES, drawn from a limited universal set of possible thematic relations: Agent, Experiencer, Theme, etc. If a speaker intends to utter a sentence in which *John* is the Agent, i.e. the 'do-er' of the action, *John* will be inserted as [John$_{AGENT}$]; if John is to be the Theme, i.e. the 'undergoer', he will be inserted as [John$_{THEME}$]. The AGENT feature can be checked only in a specifier-head configuration in the vP, and the THEME feature only in a head-complement configuration in the VP. To accomplish this, we need movement; not of the roots themselves, but of their null copies, or TWINS (c.f. 'trace'):

(13) *Twin (t)*: A phonologically null, semantically full copy of a propositional element.

Twins are created wherever a propositional element is inserted outside of its interpretive position (for Chesi 2007, this is an 'unselected' position). If a PE is inserted directly in a theta position (such as will frequently be the case with Theme arguments), it can immediately check its theta-feature, and no twin is produced. If, however, it is inserted in a non-theta-position (anywhere in the FDL except the phase edge), it produces a twin which then inherits its theta-feature. This twin is added to LA, then inserted at the phase edge, which acts as a data buffer[25] between the FDL and the SPL. In the SPL, the twin reaches its theta position and checks its theta-feature.

This is where UTAH comes in (rather, its gritty reboot, *UTCH*):

(14) UNIFORMITY OF THETA-CHECKING HYPOTHESIS (UTCH):
      Identical thematic relationships between propositional elements (PE$_x$) and (PE$_y$) are represented by identical structural relationships between the copies of (PE$_x$) and (PE$_y$) which check their theta-features upon transfer to the C-I interface.

---

[25]Chesi (2015) uses a Memory Buffer separate from the LA, with the same basic restrictions (ibid., p. 80) as I ascribe to LAs, i.e., that it each phase has its own memory buffer, and that for a phase to be spelled out, the content of the memory buffer must either be exhausted or transferred to the memory buffer of a sequential phase. However, he assumes that copies are transferred directly to the memory buffer, with no need to remerge them at the phase edge, whereas my assumption that they must be remerged at the edge is standard in mainstream phase theory, and I think necessary to account for WH islands (see section 4).
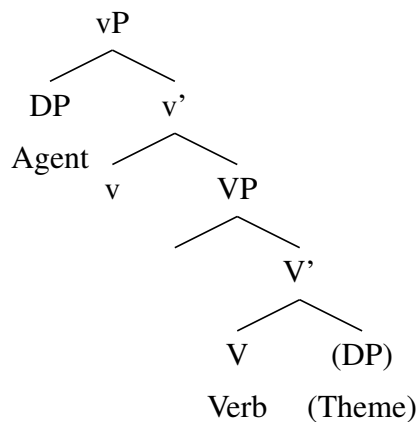
```
                              vP
                         ╱         ╲
                       DP           v'
                       Agent      ╱    ╲
                                v       VP
                                      ╱    ╲
                                           V'
                                         ╱    ╲
                                       V        (DP)
                                     Verb     (Theme)
```

Figure 3   The universal configuration associated with propositional interpreta-
tion.

In plainer English, feature-checking for a given theta role (e.g., THEME) al-
ways occurs in the same structural configuration with respect to the verb. In a
sense, this configuration *is* the (simple) proposition, schematized as in Figure 3.[26]
In this model, UTAH/UTCH is the end goal of a derivation, not a re-tracing of
movement. In other words, a derivation is a procedure that takes discourse-related
elements and arranges them into this universal configuration for the purpose of
propositional interpretation.

4 Island phenomena and conclusion

This brief sketch should hopefully demonstrate that at least some of the machin-
ery of bottom-up approaches can be incorporated into top-down derivations; it is
of course an open question whether other elements of the theory can be similarly
incorporated. This awaits further research, but my contention is that any syntac-
tic mechanism that can be described in a bottom-up system using Merge can be
described in a top-down system using Branch. For example, WH islands can be
accounted for in terms of the interaction between the buffer and the lexical sub-
arrays, where the 'island' is a restriction on adding a WH element to the LA where

---

[26]I remain agnostic as to how heads and complements are linearized cross-linguistically here.
But I think the hierarchical configuration is universal (although other projections irrelevant to this
discussion may be present, e.g. an AspP between vP and VP (MacDonald 2008).

the buffer already contains one. Consider the partial derivations of (15a) and (15b) in (15c-f):

(15)   a. *What did John think who kissed?

      b. What did John think Mary kissed?

      c. **what$_i$** did John think [$_{CP}$ ....
$$\text{LA}_2\text{: } t_i$$

      d. ***what$_i$** did John think [$_{CP}$ t_i....
$$\text{LA}_2\text{: *who}_j \text{ kissed}$$

      e. **what$_i$** did John think [$_{CP}$ t_i....
$$\text{LA}_2\text{: Mary}_j \text{ kissed}$$

      f. **what$_i$** did John think [$_{CP}$ t_i [TP Mary$_j$ [vP t_j kissed t_i]]]

In (15c), t_i (the null copy of *what$_i$*) is the last remaining element from the previous LA, and can now be remerged into the specifier of the embedded CP, which is part of the phase edge. In the case in (15d) where *who* is chosen for LA$_2$ (the sub-array for the embedded CP), its only possible insertion site is already occupied. As a result, the derivation will crash, as LA$_2$ will never be exhausted if *who* cannot be inserted.[27] In (15e), however, we see that choosing a DP without a WH feature allows its insertion in spec, TP, and the movement of its twin to spec, vP for propositional interpretation. The result of this is that when a WH element (or one of its copies) is in the buffer, other WH elements cannot be added to the same LA; this has the flavor of a general restriction on lexical sub-arrays.

Other island phenomena will hopefully yield to successful top-down analyses, and indeed Chesi, den Dikken, and others have made significant progress in this regard. Ideally, of course, island phenomena should be reducible to general computational constraints, the goal being for the theory to "deliver the 'syntactic archipelago' on a silver plate"(den Dikken 2018: 103). I leave elaboration of this model, along with the incorporation of other components of the grammar, to future work. I hope to have planted a seed, however, and I do think we shall see in the end that trees grow from their roots, not their leaves.

---

[27]The derivation will crash for an additional reason, as *who* will be enumerated along with the feature [$_{AGENT}$], a feature that is then stranded in LA and never introduced into the structure. The proposition associated with the embedded vP will be uninterpretable without this feature (i.e., without a copy of the Agent in spec, vP).

References

Baddeley, Alan, Giuseppe Vallar & Barbara A. Wilson. 1987. Sentence comprehension and phonological memory: Some neuropsychological evidence. In Max Coltheart (ed.), *Attention and performance XII: The psychology of reading*, 509–529. London: Lawrence Erlbaum Associates.

Baddeley, Alan & Barbara A. Wilson. 2002. Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia* 40(10). 1737–1743.

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Bever, Thomas G. 1970. The cognitive basis for linguistic structure. In John R. Hayes (ed.), *Cognition and the Development of Language*, 279–362. New York: Wiley & Sons.

Bianchi, Valentina & Cristiano Chesi. 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45(4). 525–569.

Carnie, Andrew. 2002. *Syntax: A generative introduction*. Oxford, UK: Blackwell Publishing.

Chesi, Cristiano. 2007. Five reasons for building phrase structures top-down from left to right. *Nanzan Linguistics: Special Issue* 3. 71–105.

Chesi, Cristiano. 2015. On directionality of phrase structure building. *Journal of Psycholinguistic Research* 44(1). 65–89.

Chomsky, Noam. 1955. *The logical structure of linguistic theory*. Bloomington, IN. Distributed by Indiana University Linguistics Club. Published in part by Plenum, 1975.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

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

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1993. A Minimalist Program for linguistic theory. In Ken Hale & Samuel Jay Keyser (eds.), *The view from Building 20: Studies in linguistics in honor of Sylvain Bromberger*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels, Juan Uriagereka & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–156. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 2005. On phases. http://www.fosssil.in/chomsky_phases.pdf.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the view from semantics*, 1–29. Berlin: Mouton de Gruyter.

Cowan, Nelson. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24. 87–114.

Cowan, Nelson. 2015. George Miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological Review* 122(3). 536–541.

den Dikken, Marcel. 2018. *Dependency and directionality*. (Cambridge Studies in Linguistics 154). Cambridge, UK: Cambridge University Press.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Epstein, Samuel D. & T. Daniel Seely. 2002. Rule applications as cycles in a level-free syntax. In Samuel D. Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the Minimalist Program*, 65–89. Oxford, UK: Blackwell Publishing.

Fodor, Jerry. 1998. *Concepts: Where cognitive science went wrong*. Oxford, UK: Clarendon Press.

Franck, Julie, Glenda Lassi, Ulrich H. Frauenfelder & Luigi Rizzi. 2006. Agreement and movement: A syntactic analysis of attraction. *Cognition* 101(1). 173–216.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276.

MacDonald, Jonathan E. 2008. *The syntactic nature of inner aspect: A minimalist perspective*. Amsterdam: John Benjamins.

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

Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 4. 157–189. https://doi.org/DOI:http://dx.doi.org/10.3765bls.v4i0.2198.

Pesetsky, David. 1995. *Zero syntax: Experiencers and cascades*. Cambridge, MA: MIT Press.

Phillips, Colin. 1996. *Order and structure*. Cambridge, MA: MIT dissertation.

Phillips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34. 37–90.

Phillips, Colin & Shevaun Lewis. 2013. Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics* 6. 11–47.

Richards, Marc. 2011. Deriving the edge: What's in a phase? *Syntax* 14. 74–96.

Rizzi, Luigi. 2016. Labeling, maximality, and the head-phrase distinction. *The Linguistic Review* 33(1). 103–187.

Ross, John R. 1985. *Infinite syntax!* Norwood, NJ: Ablex Publishing.

Sato, Yosuke. 2009. Coordination, dependency, and gapping in Japanese at the interfaces. https://ling.auf.net/lingbuzz/000809. Accessed Jan. 17, 2019.

Townsend, David J. & Thomas G. Bever. 2001. *Sentence comprehension*. Cambridge, MA: MIT Press.

Zwart, Jan-Wouter. 2009. Prospects for top-down derivation. *Catalan Journal of Linguistics* 8. 161–187.

Zwart, Jan-Wouter. 2015. Top-down derivation, recursion, and the model of grammar. In Andreas Trotzke & Josef Bayer (eds.), *Syntactic complexity across interfaces*, 25–42. Berlin: De Gruyter.