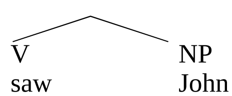
Dependency Formation
and Directionality of Tree Construction

Norvin Richards
MIT

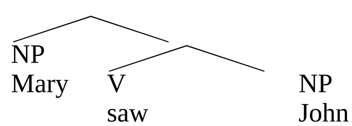
Cyclicity effects are implemented in the Minimalist program, in part, by interleaving tree-building operations with other syntactic operations. The order of operations in the derivation is linked to the hierarchical structure of the tree, since this structure is created in the course of the derivation in a way which makes the hierarchical structure a good guide to the order in which material was introduced. Making this theoretical move raises an important question: in which direction is the tree constructed?

The standard answer in Minimalism to this question (see, e.g., Chomsky 1995, 1998) is that the tree is created from the bottom up, with material lower in the tree being introduced before material higher in the tree. For instance, a VP with NPs as its specifier and complement[1] would be created as follows. First, the V would be merged with its complement, to create a constituent like that in (1a). The resulting constituent would then be merged with the NP specifier, creating the new constituent (1b):

(1) a.



b.



Phillips (1996, to appear) has proposed that the tree should be created in essentially the reverse order, from the top down[2]. In Phillips' theory, the derivation of a VP like the one in (1) would involve first merging the specifier with the verb, to yield the constituent in (2a). Next, the complement NP is

---

[*] For helpful comments on this paper I would like to thank Ben Bruening, Noam Chomsky, Kyle Johnson, Cornelia Krause, Martha McGinnis, David Pesetsky, Colin Phillips, Uli Sauerland, Kazuko Yatsushiro, and audiences at WCCFL 18, the University of Arizona, the University of California at Irvine, Kanda University of International Studies, and Keio University. Many thanks to Roumyana Izvorski, Ani Petkova, and Roumyana Slabakova for their help with the facts of Bulgarian, and to Lisa Cheng, Hooi Ling Soh, and Wei-tien Dylan Tsai for their assistance with Chinese; all examples from those languages which are not otherwise cited are from them. Responsibility for any errors is mine alone.

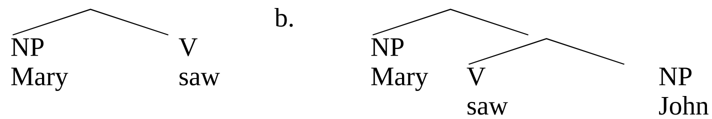
[1] This is purely for the sake of illustration; whether such a VP can exist is obviously immaterial here.

[2] Actually, Phillips' proposal is that trees are constructed from left to right; the difference will (hopefully) not be relevant for the proposal discussed here.

MIT Working Papers in Linguistics 33, 000–000

merged with the verb, destroying the constituent in (2a) and creating the new constituent structure in (2b):

(2) a. b.



The end result of this derivation is the same as the one in (1), but the history is different in interesting and, Phillips argues, important ways. One important difference is that the derivation in (2) involves the creation of constituents which do not survive in the final representation; the constituent [Mary saw] exists in (2a) but has ceased to exist in (2b). No similar changes occur in derivations like the one in (1), where the constituent created in (1a) is still a constituent in (1b). Phillips argues convincingly that these changes in constituent structure over the course of the top-down derivation offer a natural account of cases in which different constituency tests yield different results; these cases are discussed very extensively in Pesetsky 1995, and in Phillips 1996, to appear.

In this paper I will try to provide some additional arguments that the tree should be constructed from the top down. The arguments will have to do with the way in which dependency formation is to be represented in such a model; I will try to show that a top-down derivation yields a better account of certain kinds of interactions between and within dependencies.

1. Argument #1: The Principle of Minimal Compliance

The first argument will have to do with a type of additional-wh effect. In Richards (1997, 1998) I developed a theory of certain kinds of interactions between dependencies which I claimed served as a diagnostic for the order in which dependencies are created. In this section I will use this diagnostic to argue that wh-movements to positions higher in the tree are created earlier than wh-movements to positions lower in the tree. If the argument is valid, and if the order in which movements takes place is an indication of the order in which the tree is constructed, the arguments to be developed will give us evidence that the top of the tree exists before the bottom of the tree; that is, that the tree is constructed from the top down.

1.1 The PMC: multiple movements to a single head

There appear to be a number of cases in which a particular constraint need only be satisfied once within a certain domain. The English facts in (3) are a case in point:

(3)

a.* Which book did the senator deny
the rumor that he wanted to ban ___ ?

b. Which senator ___ denied the rumor that he wanted to ban War and Peace ?

c. Which senator ___ denied the rumor that he wanted to ban which book ?

The contrast between (3a) and (3c) is surprising on a theory which posits covert movement. Both of these examples, in such a theory, involve extraction from an island, yet (3c) is well-formed. We might in principle account for these facts by drawing a distinction between moved wh-phrases and wh-in-situ; for instance, by not positing covert movement as a way of dealing with wh-in-situ, or by claiming that overt and covert movement are not equally sensitive to island effects (for approaches of the first type, cf. Reinhart 1995, Cole and Hermon 1998, and references cited there; approaches of the second type include Huang 1982 and much subsequent work). Alternatively, we might draw a distinction between multiple-wh constructions like (3c) and single-wh constructions like (3a). I have claimed in other work (cf. Richards 1997, 1998) that this is the correct type of approach to these data; once an attractor has attracted a wh-phrase in a way which obeys Subjacency, it is free to trigger Subjacency-disobeying movements for the rest of the derivation. We may think of Subjacency, I have suggested, as being like a "tax" that must be paid once; once one wh-phrase has paid the Subjacency Tax, subsequent wh-movements need not pay it again.

Evidence that the Subjacency Tax approach is the correct one comes from the fact that data like those in (3) can be found in languages in which the distribution of overt and covert movement is not the same as it is in English. Watanabe (1992) discovered parallel facts in Japanese, a language in which all wh-phrases are left in situ:

(4)

a. *Taroo -wa Hanako -ga **nani** -o katta ka dooka
Taroo TOP Hanako NOM what ACC bought whether
tazuneta no?
asked Q
'What did Taroo ask [whether Hanako bought]?'

b. Taroo-wa Hanako -ga kuruma-o katta ka dooka
Taroo TOP Hanako NOM car ACC bought whether
dare -ni tazuneta no?
who DAT asked Q
'Who did Taroo ask [whether Hanako bought a car]?'

c. Taroo -wa Hanako -ga **nani** -o katta ka dooka
Taroo TOP Hanako NOM what ACC bought whether
dare -ni tazuneta no?
who DAT asked Q
'Who did Taroo ask whether Hanako bought what?' (Watanabe 1992)

In Japanese, as in English, wh-extraction from certain kinds of islands is blocked, as we can see in (4a). Japanese also exhibits an additional-wh effect like the one found in English, however; adding a wh-phrase outside the offending island, as in (4c), redeems the island violation.

A similar set of facts can be found in Bulgarian, a language in which all wh-movement is overt:

(5)

a. ***Koja kniga** otreče senatora*t
which book denied the-senator
malva*ta če iska da zabrani ___
the-rumor that wanted to ban
'Which book did the senator deny [the rumor that he wanted to ban t]?

b. **Koj senator** ___ otreče
which senator denied
malva*ta čeiska da zabrani *Vojna i Mir*
the-rumor that wanted to ban war and peace
'Which senator denied the rumor that he wanted to ban *War and Peace*?'

c. **Koj senator** **koja kniga** ___ otreče
which senator which book denied
malva*ta čeiska da zabrani ___
the-rumor that wanted to ban
'Which senator denied the rumor that he wanted to ban which book?'

These sentences are translations of the English examples in (3), and the judgments are identical, even though all of the movements in question are overt.

Apparently, then, the contrast between overt and covert movement is not a relevant one for the facts in (3), and something like a Subjacency Tax approach is called for; in (3c), (4c), and (5c) above, wh-extraction from a position outside the island "pays the Subjacency Tax," making it possible to subsequently violate Subjacency by extracting from inside the island.

Note that this effect does seem to be crucially sensitive to properties of the derivation, as the contrast in (6) shows:

(6)

a. **<u>Which senator</u>** ___ said that
[the rumor that he wanted to ban **<u>which book</u>**]
had been spread by Communists?

b. *<u>**Which book**</u> did the senator say that
[the rumor that he wanted to ban ____]
had been spread **by whom**?

(6a) is another Subjacency Tax case; Subjacency-obeying movement of *which senator* makes it possible to subsequently violate Subjacency by covertly moving *which book* out of the island. (6b) is an attempt at paying a Subjacency Tax retroactively; Subjacency is first violated by overt movement of *which book* out of an island, and then subsequently obeyed by covert movement of *by whom*. Apparently this is ineffective; Subjacency must be obeyed by the <u>first</u> wh-movement triggered by a given C.

Bearing this result in mind, consider the Bulgarian Subjacency Tax example in (5c), repeated as (7):

(7) **<u>Koj senator</u>** **<u>koja kniga</u>** ___ otreče
which senator which book denied
[malva*ta čeiska da zabrani ___]
the-rumor that wanted to ban
'Which senator denied the rumor that he wanted to ban which book?'

In (7) *koj senator* 'which senator' is paying the Subjacency Tax for *koja kniga* 'which book'. If it is generally the case that it is the first wh-movement in the derivation which must obey Subjacency, the well-formedness of (7) is evidence that *koj senator* moves earlier in the derivation than *koja kniga*. In other words, we apparently have evidence that in cases of movement to multiple specifiers, movement to the highest specifier occurs first in the derivation, and subsequent movements "tuck in" to specifiers below the existing ones. See Richards 1997, to appear, Mulders 1997 for discussion (these works crucially assume a bottom-up derivation; for one way of getting a similar result in a top-down approach, see appendix 1).

In Richards (1997, 1998) I suggested that the Subjacency Tax phenomenon is a special case of a general phenomenon in which dependencies may "assist" one another, a phenomenon for which I tried to give a unified account using what I called the Principle of Minimal Compliance (PMC). A simplified version of the PMC is given in (8); the simplification will not be relevant for the concerns of this paper.

(8) **The Principle of Minimal Compliance**
Given a dependency D that obeys a constraint C, the participants in D can be ignored for the rest of the derivation for purposes of determining whether any other dependency D' obeys C.

For purposes of (8), we must understand the "participants" in a dependency as including the head and tail of the dependency, as well as the attractor in cases of dependencies created by feature attraction. What the PMC claims is essentially that elements of syntactic structure need only obey constraints once; after they have obeyed a constraint, they are no longer considered for purposes of determining whether a constraint has been obeyed.

Consider how the PMC would interact with the following version of Subjacency to yield Subjacency Tax effects:

(9) **Subjacency**
An attractor cannot attract across an island.

Of course, we would like to have a definition of what constitutes an "island", but no such definition is necessary for our purposes here. The derivation involved in Subjacency Tax effects is given in (10):

(10) a. C^o wh_1 [**island** wh_2]

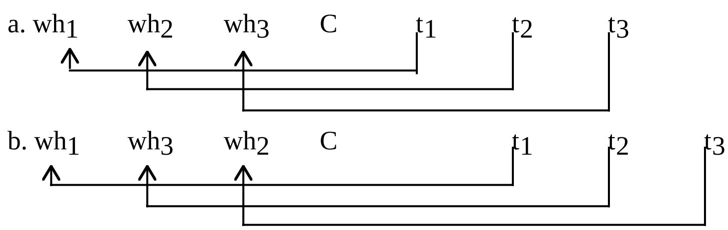
b. wh_1 C^o t_1 [**island** wh_2]

In (10a), the attractor attracts a wh-phrase in a way which obeys Subjacency. The attractor has now participated in a Subjacency-obeying dependency, and, by virtue of the PMC, may be ignored for the rest of the derivation for purposes of determining whether other dependencies obey Subjacency. In (10b) the attractor attracts a wh-phrase out of the island, violating Subjacency. According to the PMC, however, we are now entitled to ignore the presence of the attractor while evaluating the structure for compliance to Subjacency; since it is the attractor that has violated Subjacency, the Subjacency violation effectively does not occur.

In previous work I have tried to show that the PMC applies to a number of different constraints. Here I will briefly describe one other PMC effect, having to do with the interaction of the PMC with Shortest Attract. Suppose we assume a version of Shortest Attract like that in (11):

(11) **Shortest Attract**
An attractor must attract the closest possible attractee.

The PMC is predicted to interact with Shortest Attract in a way which will become apparent in cases where more than two movements are triggered by a single attractor. The prediction is that just like Subjacency, Shortest Attract will have to be obeyed by the first instance of Attract, but that after this attractees will be able to move in any order. This prediction is illustrated in (12) for multiple wh-movement:

(12)


Recall that we have seen evidence to the effect that in cases of movement to multiple specifiers of a single head, the linear order of the specifiers reflects the order in which movement took place; the highest specifier moved first, followed by the second specifier, and so forth. The prediction for a case like the one illustrated in (12), then, is that the highest wh-phrase will have to move first, but that the second and third wh-phrases will be able to move in either order.

This prediction is borne out, as data from Bulgarian discovered by Bošković (1995) show. The examples in (13) show a pair of wh-phrases which must ordinarily occur in a particular order; by hypothesis, this order reflects the underlying c-command relation between the wh-phrases:

(13) a. Kogo kakvo e pital Ivan
whom what AUX asked Ivan
'Who did Ivan ask what?'

b.?* Kakvo kogo e pital Ivan
what whom AUX asked Ivan (Bošković 1995, 13-14)

In (14) we see the effects of the PMC; when a third, higher wh-phrase is added, the wh-phrases in (13) become freely ordered, as predicted:

(14) a. Koj kogo kakvo e pital
who whom what AUX asked
'Who asked whom what?'
b. Koj kakvo kogo e pital

Here, as in the Subjacency case, there is evidence that the well-formed move must precede the ill-formed move in the derivation. If it were possible to remedy a Shortest Attract violation with a subsequent move which obeyed Shortest Attract, then we would expect the example in (15) to be well-formed, contrary to fact:

(15) * Kogo koj kakvo e pital
whom who what AUX asked
'Who asked whom what?'

The derivation in (15) would involve first violating Shortest Attract by attracting *kogo* 'whom' past the subject wh-phrase *koj* 'who', and then obeying Shortest Attract by attracting *koj*, the highest wh-phrase. If it were possible to apply the PMC retroactively, we would expect this example to be well-formed. Subjacency and Shortest Attract both interact with the PMC in the same way, then; the well-formed move must occur first, if it is to redeem a move which would be ill-formed in isolation.

1.2 The PMC revisited: Superiority

Armed with these beliefs about the nature of movement, let us move on to consider some new cases involving the PMC. In the previous section I tried to show that the PMC can be used as a diagnostic for the order in which operations take place in the derivation; if a well-formed move α is to redeem an ill-formed move β, then α must precede β in the derivation. The PMC cases we have examined up until now have involved multiple movements triggered by the same attractor. In this section I will discuss a PMC case involving two different attractors, which gives us evidence, I will argue, that the tree should be constructed from the top down. The logic of the argument will be as follows. The PMC case discussed here will involve a well-formed move triggered by an attractor high in the tree which allows an attractor lower in the tree to trigger an ill-formed movement with impunity. Suppose that the PMC is in fact a reliable guide to the order in which operations take place. Then the PMC case discussed here will crucially involve movement triggered by a higher attractor which precedes movement triggered by a lower attractor in the derivation. If we are to derive cyclicity effects by forcing the tree to be constructed in a particular direction and allowing the operations involved in constructing the tree to be interleaved with other syntactic operations, then the top of the tree will have to exist before the bottom of the tree; that is, the tree will have to be constructed from the top down.

Sentences like the one in (16) are discussed by Huang (1982) and by Lasnik and Saito (1992, 118):

(16) Who t wonders what who bought t ?

(16) has several surprising properties. One is that it is unambiguous; the wh-in-situ *who* in the embedded clause may only take matrix scope, not embedded scope. That is, this sentence can only be interpreted as a pair-list question about pairs of people, with an embedded single-wh question (the meaning roughly

sketched in (17a); it cannot have the other logically possible interpretation, involving an embedded multiple-wh question (17b):

(17) a. For what person x and what person y does x wonder what y bought?

b. Who wonders for what person y and thing z y bought z?

This is at odds with the ordinary behavior of wh-in-situ in English, which can typically take scope at any c-commanding interrogative C; compare the minimally different (18), which is ambiguous in the expected way:

(18) Who t wonders who t bought what?

Not only is the behavior of the *who* in situ in (16) surprising given the ordinary behavior of wh-in-situ in English, but it is rather surprising on more general theoretical grounds. Moving objects very commonly cannot move too far, but this is a case of a moving object which cannot move to a local position; it must skip the closest scope position and land in the matrix clause.[3]

Another surprising property of (16) is that it is well-formed, despite the fact that it contains a Superiority violation in the embedded clause. Apparently the short wh-movement in the matrix clause is crucial for licensing this Superiority violation, as we can see if we remove the matrix wh-phrase:

(19) *John wonders what who bought t

This looks like a job for the PMC; we have here another case of a movement which would be ill-formed in isolation, but which is permitted because of the presence of a well-formed movement. In this case, the ill-formed movement is a Superiority violation in the embedded clause, and the well-formed movement is a local wh-movement in the matrix clause. In the previous section we saw that well-formed movements must precede ill-formed movements if the PMC is to apply. If this generalization holds here as well, the wh-movement in the matrix clause must occur earlier in the derivation than the Superiority violation in the embedded clause; that is, the tree must be constructed from the top down. I will also be making non-trivial assumptions about the nature of covert movement, which space considerations will prevent me from defending or discussing here.

Suppose we consider the derivation of (16). Throughout I will be assuming a locality restriction Shortest (cf. Richards 1997), meant to subsume the notions of Shortest Attract and Shortest Move:

[3] Constraints on "overly short movement" are not unattested in the literature, however; cf. Tada 1993, Saito and Fukui 1998.

(20) **Shortest**
An attractor cannot create a movement relation between A and B if there is some X c-commanded by A and c-commanding B such that X could participate in the movement relation (either as a landing site or as a mover).

(20) simply requires attractors not to trigger movements that skip potential landing sites or elements that could undergo the relevant kind of movement.

Let us first consider a top-down derivation of (16). The derivation would begin from the top of the tree, constructing the tree until the trace position for the wh-phrase *who* was reached:

(21) Who$_1$ C

Next a trace of *who* would be created in the trace position:

(22) Who$_1$ C ~~who~~

Tree construction would then continue, down past the embedded CP with its wh-phrase specifier to the third wh-phrase *who* in situ:

(23) Who$_1$ C ~~who~~ wonders what C

(24) Who$_1$ C ~~who~~ wonders what C who$_2$

At this point in the derivation covert movement of *who* in situ can take place into the matrix clause. The move violates Shortest, since *who* skips the Spec of the embedded CP, a possible landing site. However, the matrix C is no longer required to obey Shortest, since it has participated in one Shortest-obeying move (that of the *who* in the matrix clause). Thus, Shortest is effectively obeyed:

(25) Who$_1$ ~~who$_2$~~ C ~~who~~ wonders what C who$_2$

We continue constructing the tree until we reach the trace position of *what*, where a trace of *what* is inserted:

(26) Who$_1$ ~~who$_2$~~ C ~~who~~ wonders what C who$_2$ bought ~~what~~

This move violates Shortest again, since *what* skips another wh-phrase (*who$_2$*). Because this wh-phrase has participated in a well-formed dependency, however (the one created in step (25)), the PMC allows us to ignore it from now on for purposes of evaluating Shortest; if we ignore *who*, *what* has not violated Shortest in (26), and (26) is therefore well-formed, as desired.[4]

This approach to the well-formedness of (16) also successfully accounts for the surprising fact that the wh-word in situ must take wide scope. Suppose we consider a top-down derivation for the (ill-formed) reading for (16) represented in (27):

(27) Who ___ wonders what who bought ___

The derivation would begin just like the one described above, beginning with the creation of the dependency headed by the *who* in the matrix clause:

(28) Who$_1$ C

(29) Who$_1$ C ~~who~~

By the PMC, since the matrix C has now attracted a wh-phrase in a way which obeyed Shortest, it is now free to disobey Shortest for the rest of the derivation. In the previous derivation, this ability was crucial in rendering the derivation well-formed; a wh-phrase was subsequently attracted from the embedded clause into the matrix clause in a way which violated Shortest, and the PMC was responsible for rendering the Shortest violation well-formed. But in this derivation the matrix C only attracts one wh-phrase; no wh-phrases are attracted from the embedded clause into the matrix clause.

The rest of the derivation, then, will involve creating the embedded clause, which will contain a Superiority violation; that is, a violation of Shortest. The PMC will be irrelevant; the C of the matrix clause is the only one which has participated in a well-formed dependency, and it will not trigger any further operations in the derivation. Thus, the PMC will be unable to redeem the Shortest violation in the embedded clause, and the derivation is correctly predicted to be ill-formed. We correctly derive the fact, then, that a sentence like (16) has only one reading, one in which the wh-in-situ has matrix scope.

Let us consider how a bottom-up derivation would fare. The tree would be constructed from the bottom up until the most embedded C was reached:

(30) C who bought what

At this point in the derivation no wh-movements have yet taken place, so the PMC is irrelevant; C must attract *who*, thus obeying Shortest:

---

[4] This use of the PMC requires us to understand the PMC as acting recursively; the move in (25), which is only well-formed because of the PMC, apparently is itself capable of triggering the PMC. See Richards 1997, 1998 for other cases of this kind of recursive application of the PMC.

(31) who C ~~who~~ bought what

Once this move has been performed, however, we have no hope of deriving the sentence in (16) (repeated as (32)), at least on standard assumptions:[5]

(32) Who t wonders what who bought t ?

In this section we discussed an example of an additional-wh effect in which a wh-phrase in a higher clause appears to license a Superiority violation in a lower clause. I have made the general claim that additional-wh effects always involve the additional wh-phrase moving before the movement which is to be "saved" takes place; this is one of the properties built into the Principle of Minimal Compliance. If that general claim is correct, then in this specific case the wh-phrase in the higher clause will have to move before the Superiority violation in the embedded clause takes place; that is, the tree will have to be constructed from the top-down.

2. Argument #2: Expletive-Associate Relations; Move over Merge

Chomsky (1995, 1998) offers an account of the facts in (33) which is based on a bottom-up derivation:

(33) a. There seems __ to be a man in the room

b. *There seems a man to be __ in the room

Chomsky's account proposes that the derivation runs as follows; first, the tree is built from the bottom up until the lower of the two EPP positions (Spec of the embedded infinitival TP) is reached:

(34) ___ to be a man in the room

At this point the EPP must be satisfied. Two options are available; either *a man* must Move to check EPP, or *there* must be taken from the Numeration and Merged. Chomsky suggests that the operation Merge is preferred to the operation Move, on the grounds that Merge is a "simpler" operation, being one of the subparts of the complex operation Move. Thus, Merge of *there* is preferred here to Move of *a man*:

(35) there to be a man in the room

---

[5] Adopting Single Output Syntax (Pesetsky 1998, Bobaljik 1995, 1999, Groat and O'Neill 1996, Fox and Nissenbaum 1999) might in fact allow this problem to be circumvented; for reasons of space, I will not discuss this any further here.

Construction of the tree continues until the higher Spec TP is reached:

(36) ___ seems there to be a man in the room

At this stage, too, Merge would be preferred over Move, but no Merge is possible; the Numeration has been exhausted. There is therefore no option but to perform a Move operation, Attracting *there* into Spec TP:

(37) there seems ~~there~~ to be a man in the room

The well-formedness of (38) is problematic for this account (as pointed out by Alec Marantz and Juan Romero):

(38) There was heard a rumor [that a man was in the room]

The Numeration for (38) apparently contains an instance of *there*. Bearing this in mind, consider a bottom-up derivation. First the tree will be built up until the lowest EPP position is reached:

(39) ___ was a man in the room

At this point we again have two ways, in principle, of satisfying the EPP; Merge *there*, or Move *a man*. By the reasoning outlined above, we should expect to be forced to Merge *there*, giving (40):

(40) there was a man in the room

But (40) is not a possible source for (38); for one thing, locality restrictions on A-movement will prevent movement of *there* from its position in (40) into the matrix clause of (38). Chomsky (1998) proposes a solution to this problem involving division of the Numeration into smaller sub-Numerations, such that *there* is not in fact contained in the Numeration for the embedded clause, and thus fails to block Move of *a man* at the step in the derivation shown in (39). Here I will consider an alternative approach to these facts, based on a top-down approach.

Consider, first, the top-down approach to the contrast in (33), repeated as (41):

(41) a. There seems __ to be a man in the room

b. *There seems a man to be __ in the room

The structure would be built from the top down until the higher of the two EPP positions is reached:

(42) There seems ___

At this point we have the same choice we had before; we must decide whether to Merge *a man* or "Move" *there* into the EPP position (where "Move", in this framework, involves creating a trace of the moved element). Apparently the preference is for movement of *there*:

(43) There seems ~~there~~

This approach thus needs to assume that Move is preferred over Merge, rather than vice versa. To put it another way, apparently there is a preference for copying material already on the "workspace", as opposed to introducing new lexical material. We might understand this as evidence that access to the lexicon is computationally costly (cf. Chomsky 1998); searching the lexicon involves interacting with a large, computationally complex object, and the computational system avoids this operation, in favor of manipulating objects already in working memory, whenever possible[6].

The derivation continues, eventually reaching the theta-position where *a man* must be Merged. Move of the expletive into the theta-position, though preferred by the general principle banning avoidable lexical access, is presumably prevented by conditions on theta-assignment.

(44) There seems ~~there~~ to be a man...

Now consider how this approach accounts for the well-formedness of (38), repeated as (45):

(45) There was heard a rumor [that a man was in the room]

Again, the tree is constructed from the top down until the first choice point is reached; in this case, the EPP position in the embedded clause:

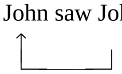
(46) There was heard a rumor that ___

At this point, since lexical access is costly, the computational system would prefer to Move *there*. This kind of movement is impossible, however, for whatever general reasons ban A-movement out of tensed clauses. Lexical access is therefore unavoidable; *a man* must be Merged in this position:

---

[6] "Wherever possible" will have to be defined; I will return to this issue several times in the paper, though I will be unable to completely resolve the issue. As a first approximation, however, we will apparently need to view movement as being subject to all the restrictions that are typically imposed on it in a bottom-up derivation; it will have to be subject to locality, will be motivated only be feature checking, and so forth. Such restrictions will be especially necessary in a top-down derivation, in order to avoid, for instance, movement from one theta-position to another; otherwise, (i) would block (ii), given the preference for Move over Merge in this system:

(i) John saw John



(ii) John saw Mary

(47) There was heard a rumor that a man...

The derivation then proceeds in the desired way.

The top-down theory, then, offers an account of the possible relations between an expletive and an associate which is arguably simpler than that of Chomsky (1995, 1998); the theory does not need to postulate the existence of sub-Numerations, or even of a Numeration.[7] The only crucial components are a lexicon and a general principle making lexical access costly; both of these seem well-motivated, and the latter will do further theoretical work for us in Appendix 1.

3. Argument #3: Sinking

A third advantage of top-down approaches to tree construction is that they make available a kind of movement operation which is not available in bottom-up approaches and which seems to be useful.

In a top-down approach there must be a constraint on Merge saying something like (48):[8]

(48) Merge α to a position which c-commands as few nodes as possible.

Assuming (48), let us consider again how dependencies are created in this kind of derivation, using as an example the embedded clause of (49):

(49) I wonder what John bought

Given the account developed thus far, the derivation of the embedded clause might be expected to proceed as follows. First *what* and the embedded C would be Merged to form a constituent:

---

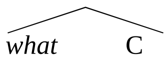
[7] One of the empirical uses of the Numeration in Chomsky (1995, 1998) is to prevent examples like (i-ii) from competing:

(i) A man is in the room

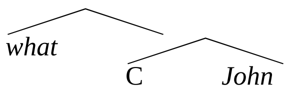
(ii) There is a man in the room

Given the preference in Chomsky's theory for Merge over Move, (ii) would block (i) if the two sentences competed; since they have different Numerations, in his approach, they do not compete. In a top-down approach the problem does not arise; the well-formedness of (i-ii) is a consequence of the lack of look-ahead properties in the computational system. A derivation can begin by Merging either *a man* or *there*; Merge of *there* will force further lexical access later in the derivation, making it a "bad choice" from the point of view of this theory, but since the computational system cannot look ahead in the derivation it has no way of avoiding such bad choices. The two structures in (i-ii) are thus both equally available, as desired.

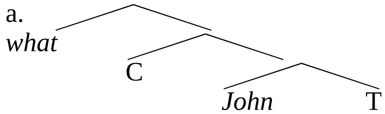
[8] More or less equivalent, ignoring for the moment the existence of complex left branches, would be a requirement that α be c-commanded by as many nodes as possible. I use the formulation in (48) largely to make it easier to explain a point which is soon to follow, but nothing very crucial hinges on this. (48) is similar in spirit to Phillips' (1996) requirement of Branch Right.

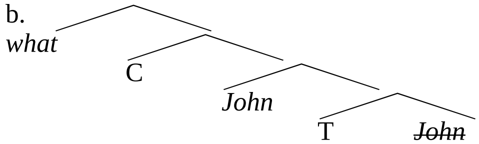
(50) 

Next *John* would be merged to the structure, destroying the constituent in (50) and creating a new constituent structure:

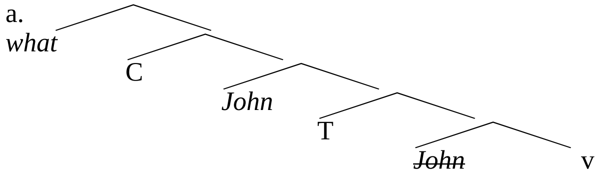
(51) 

Construction of the tree would continue until the theta-position for *John* was reached, at which point a copy of *John* would be created:[9]

(52) a. 

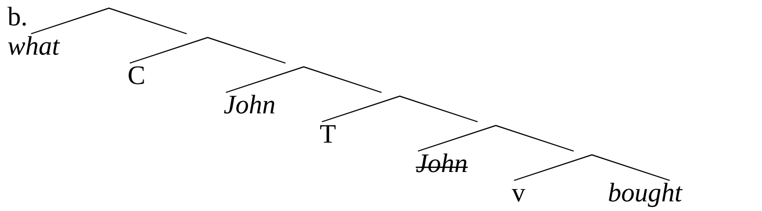
b. 

The top-down derivation would continue, again, until the theta-position for *what* was reached, at which point a copy of *what* would be inserted:

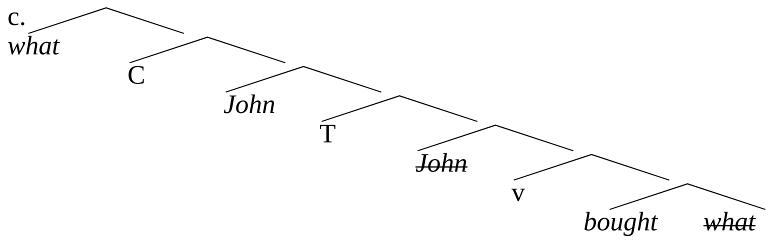
(53) a. 

[9] Here I insert the copy of *John* before inserting the theta-role assigner v, but this is not necessary; if v were inserted instead of *John* in (52b), subsequent insertion of a copy of *John* would still yield the structure in (53a), assuming that linear order is not crucially represented in these trees, at least in this case. I will demonstrate this point more fully in Appendix 2.

b.



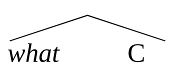
c.



In this derivation, "movement" involved creation of copies of the moving elements at the point in the derivation at which a position for the copy was introduced; the theta-positions, in the movement examples given above. I will reserve the term "movement" for this kind of operation.

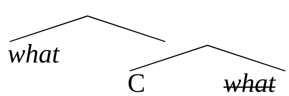
Suppose it were possible to create a copy of an object much earlier, perhaps immediately after it was introduced into the structure. This kind of early copying would allow a formally different kind of movement relation with certain arguably desirable properties, which I will refer to here as "sinking". Suppose we consider a derivation for the structure above involving Sinking of *what*. The derivation would begin as before, with a constituent consisting of *what* and the interrogative C:

(54)



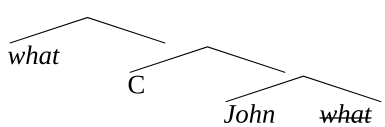
Next we create a copy of *what* immediately:

(55)



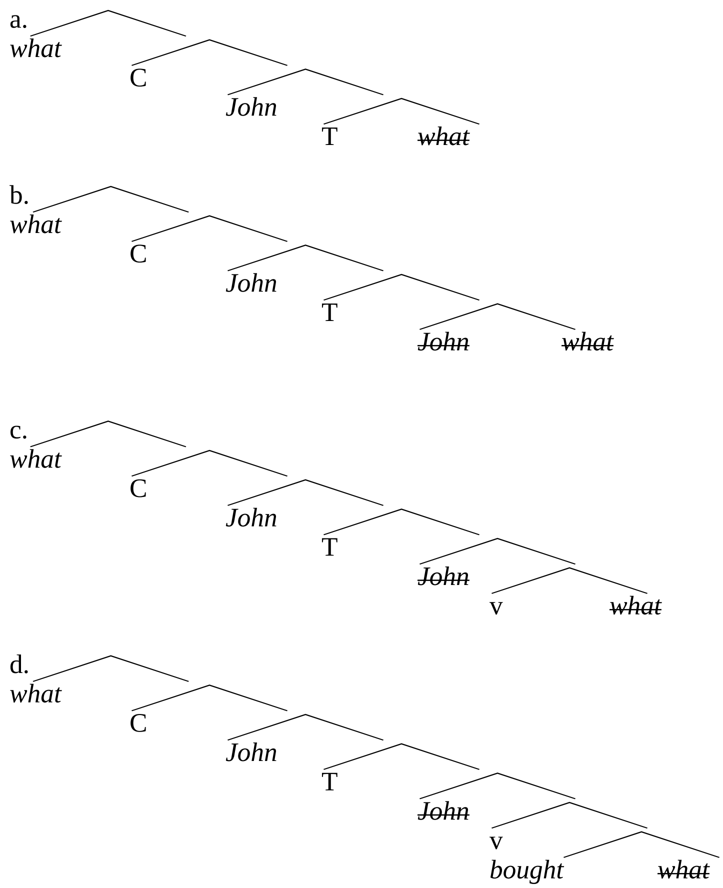
The copy of *what* has been Merged in a way that obeys the constraint on Merge in (48); it c-commands the smallest possible number of existing nodes of the tree (namely, one). Next we Merge *John*, again obeying (48) by having *John* c-command only one node (in this case, the copy of *what*):

(56)



The derivation continues as before, with every new node being Merged in a position c-commanding exactly one node; namely, the copy of *what*. At the end of the derivation, the theta-position of *what* is reached; further Merge, if any were to take place, would extend the tree below *what*'s copy:[10]

(57)



In this derivation the word *what* undergoes a kind of very successive-cyclic wh-movement; it enters into a local c-command relation, at some point in the derivation, with every syntactic object along the path between Spec CP and its theta-position. Such "extreme successive-cyclic movement" has been posited in a number of approaches, for a variety of reasons, some of which I will describe below (cf. Chomsky 1998, Saito and Fukui 1998, Agbayani 1997, 1998,

[10] In this derivation the copy of *John* could not be Merged after v without violating (48) (cf. the preceding footnote). Either the grammar must tolerate this degree of look-ahead, or violations of (48) must be allowed under certain circumstances (perhaps when phonologically null nodes are involved, as in this case; such an exception might be made to look natural if the derivational approach developed here were to be linked strongly with a theory of parsing).

Takahashi 1994, Fox 1999 for approaches of this kind in a Minimalist framework; other parallels include Manzini's (1992) address-based dependencies, Kayne's (1983) notion of g-projection, Chomsky's (1986) system of movement, TAG approaches to movement (e.g., Kroch 1989), and the HPSG notion of slash percolation (Borsley (1996) and references cited there). In the Minimalist program, however, extreme successive-cyclic movement, like successive-cyclic movement generally, has seemed strangely out of place. Since movement in that program is generally assumed to be driven by feature checking, advocates of successive-cyclic movement must decide whether to invent a feature to drive the intermediate movements or to develop some way for successive-cyclic movement to be exempt from the requirement that movement be feature-driven. Both measures seem rather ad hoc. In the top-down approach, on the other hand, Sinking is actually very minimally distinct from movement, the only difference being the point in the derivation at which the copy is created. The successive-cyclic property of the movement follows naturally from the way in which Merge is performed in this theory, which seems like a desirable result. Of course, if we do decide to postulate Sinking, we would like to know which examples of displacement involve movement and which involve Sinking; I will make a very preliminary attempt at answering this question, while leaving many issues open[11].

3.1. Arguments for very successive-cyclic movement

In this section I will describe briefly some of the uses to which very successive-cyclic movement has been put in the literature; these phenomena will be useful as diagnostics for Sinking in the remainder of the paper.

One possible use for Sinking has to do with the licensing conditions on parasitic gaps. Nissenbaum (1998) and Richards (1997) both develop accounts of parasitic gap licensing in which the licensing operator must enter into a local syntactic relation with the island containing the parasitic gap in the course of the derivation. The derivation of a PG example like (58), for example, would involve the wh-phrase *what* adjoining to the maximal projection containing the subject island:

(58) Which book do [[people that read PG] ___ usually like ___ the best]?

Nissenbaum (1998) and Richards (1997) both claim that adjunction of the wh-phrase *what* to a maximal projection containing the subject island is the operation responsible for making it possible to create a dependency between the operator and the parasitic gap in the subject island. If these theories of parasitic

[11] The question about the distribution of Sinking is clearly linked to the question of what kinds of features we should decide to posit as driving forces for movement; one property of Sinking is that it allows us to characterize movement past certain kinds of positions which is not obviously driven by features in those positions. Defining the class of syntactically active features is beyond the scope of this paper; the hope of the above discussion is simply that this class can be reduced by eliminating those features which have no purpose other than to drive successive-cyclic movement.

gaps are correct, then, wh-movement must target some position between its base position and its landing site.

Another case in which very successive-cyclic movement would appear to be useful was discovered by Lebeaux (1991) (and see also Fox 1999 for discussion). The phenomenon in question has to do with reconstruction of a relative clause which contains both a pronoun that must be bound by a quantifier outside the relative clause and an R-expression that must be free from binding by a pronoun outside the relative clause. The interesting contrast in question is given in (59):

(59) a. [Which paper that he_i gave to the $teacher_j$] did every $student_i$ ask her_j to read t?

b.*[Which paper that he_i gave to the $teacher_j$] did she_j tell every $student_i$ to fix t?

The contrast in (59) can be accounted for on the assumption that reconstruction can only treat the entire relative clause as a unit. In (59a), there are positions in the tree to which reconstruction of the relative clause could occur in a way which would both allow *the teacher* to obey Condition C and allow *he* to be bound by the quantifier *every student*. The positions in question are those that are c-commanded by *every student* and not by *her*. In (59b), by contrast, there are no positions that are c-commanded by *every student* but not by *she*; if the relative clause is to be reconstructed as a single unit, there is no position to which the relative clause can be reconstructed in a way which will satisfy the conditions on *he* and *the teacher.*

On the further assumption that reconstruction can take place only to a position which has been occupied by the moving element at some point in the derivation, the well-formedness of (59a) is evidence that the NP *which paper that he gave to the teacher* occupies some position between *her* and *every student* at some point in the derivation, since a well-formed sentence can only result from reconstruction to such a position. Thus, movement must be very successive-cyclic, targeting not only Spec CP but also other intermediate positions along the path of movement (including, for instance, a position between *her* and *every student* in (59a); such a position could not be a Case-checking position for the wh-phrase, since it is insufficiently local to the theta-position, nor is it a Spec CP).

The particular mechanism I have proposed above for dealing with extremely successive-cyclic movement forces us to take a particular attitude towards reconstruction, if these facts are to be accounted for. Recall that this account does not posit features to drive movement to intermediate positions; rather, successive-cyclic “movement” takes place when Merge is repeatedly performed in such a way as to leave the “moving” XP as one of the two lowest nodes in the tree. There is no reason to expect Sinking to leave copies (traces) of the Sinking XP; the formation of copies, in the account described above, is driven by feature-checking. The explanation of Lebeaux’s facts, in this view of successive-cyclic movement, therefore cannot make use of theories of reconstruction which rely on the Copy Theory of movement, at least for the kinds of reconstruction involved in Lebeaux’s cases. It cannot be the case, on this theory, that (59a) is well-formed because at LF, there is a copy of the relative clause which is c-commanded by *every student* but not by *her* , and this is the copy which LF interprets, any other copies having been deleted.

Instead, the theory under development here might use an approach to reconstruction somewhat like that developed by Sauerland (to appear). On this approach, Sinking of the relative clause to an intermediate position in the clause in (59a) yields a structure which is “sent” to LF and interpreted; subsequent further sinking is responsible for moving the relative clause’s head to its theta-position, at which point the head is sent to LF for further interpretation. This way of dealing with reconstruction does not require that there be a copy of the moving phrase in the “reconstruction site”, only that there be a point in the derivation at which the moving phrase is in the reconstruction site, and that it is possible for that structure to be sent to LF but not to be pronounced. The representation sent to LF is clearly incomplete; the point in the derivation in question, in this framework, is one in which the sinking NP is at the bottom of a tree which contains *every student* but in which *her* has not yet been Merged into the structure. Dealing with reconstruction in this way would therefore force us to adopt a Cyclic Spell-Out approach.

Notice that this way of dealing with reconstruction is not available for TAG approaches to movement, which Sinking otherwise closely resembles. A TAG derivation proceeds by starting with the moving object in what will be its lowest position in the tree, and then inserts material between the moving object and the material around it, causing it to “rise” through the tree. (60) shows a simplified derivation for a case of long-distance wh-movement:

(60) a. Who left?

b. Who **do you think** left?

The derivation in (60) starts with a monoclausal question, and then adds material between *who* and *left*, either in one operation or in multiple smaller operations; in this case, the added material is the boldfaced *do you think* (or rather, the syntactic structure associated with those words). This derivation is like Sinking in that it does not need to postulate additional features to drive successive-cyclic movement and does not postulate intermediate traces. But it differs from Sinking in that there is no point in the derivation at which the moving phrase is c-commanded by the material it “moves past” in the course of the derivation. In (60), *who* begins the derivation in a position where it is not c-commanded by anything relevant; in (60b) it remains at the top of the structure, all the new material having been introduced in positions underneath *who*. The TAG approach therefore does not have the option of making use of the derivational approach to reconstruction sketched above. TAG also cannot make use of the Copy Theory of movement, of course, since the relevant movements do not involve the creation of copies.

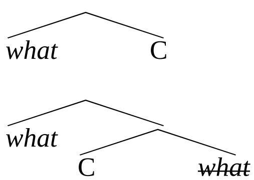
3.2 Restrictions on Sinking

In this section I will outline some plausible restrictions on the Sinking operation that might be made to follow from general considerations. We will see that these restrictions have arguably desirable empirical consequences.

3.2.1 Argument/adjunct distinctions

Sinking involves a certain "look-ahead" property; a copy of a movable object is created before the landing site of movement (and, presumably, the feature that is checked by movement) appears. The first two steps of the Sinking derivation in (54-57) are repeated here:

(61) a. b.



In step (61b), when the copy of *what* is created, the attractor bearing the next feature which will drive movement of *what* (presumably a ϕ-feature on a functional head) has not yet appeared.

On the other hand, it might be determinable by inspection that *what* will be undergoing movement later in the derivation. *What* presumably has its own Case feature, for instance, which will be checked against a feature on some other head later in the derivation. Knowing that this Case feature cannot be checked in (61b), the computational system might be able to justify making a copy of *what* immediately, since it will clearly need to do so at some point.

Consider, on the other hand, wh-movement of an adjunct. If the wh-phrase in (61) were *why* instead of *what*, it would not have a Case feature to check later in the derivation. In fact, current theories are consistent with the possibility that *why* might not have any features to check at all, other than the wh-feature which is already in a checking relation at the beginning of the top-down derivation. In the case of adjunct wh-movement, then, Sinking might be unavailable; the look-ahead problem would be insoluble in these cases.

If this is correct, then we expect adjuncts not to exhibit the properties of Sinking. Parasitic gaps, for instance, were one of the phenomena presented above as a diagnostic for extremely successive-cyclic movement. We thus predict that adjuncts will be unable to license parasitic gaps:

(62) *How well did he behave ___ [after fixing the lawnmower ___]?

We also expect adjuncts to be unable to participate in Lebeaux's (1991) phenomenon. Here judgments are not at all clear, but it seems to me that the prediction is correct:

(63) a. *[How much better than the teacher$_j$ said he$_i$ was likely to behave]
did every student$_i$ tell her$_j$ he had behaved?

b. *[How much better than the teacher$_j$ said he$_i$ was likely to behave] did she$_j$ tell every student$_i$ he should behave?

Finally, this contrast between arguments and adjuncts might give us an account of some of the differences between arguments and adjuncts with respect to island effects. Consider an island effect which has the form in (64):

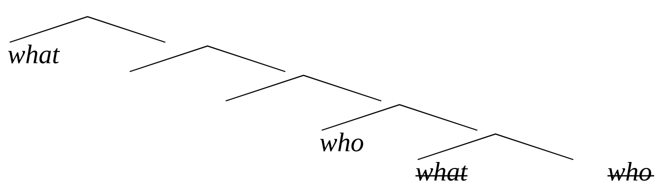
(64) Given a constituent XP and a movable object α which c-commands XP, do not create a copy of α inside XP.

A constraint like (64) would ban ordinary movement into a constituent XP. On the other hand, Sinking into XP would not violate (64); the copy of α would be created early in the derivation, before XP was formed. Assuming that (64) is a condition on the copy operation, and not on the representational relation between the head and tail of the chain, Sinking would not violate it. Since Sinking, by hypothesis, is available only to arguments and not to adjuncts, the prediction is that any conditions like (64) would apply to adjuncts only. The notion of Sinking might give us a new way, then, of capturing the difference between adjuncts and arguments with respect to island sensitivity.

3.2.2 Nested paths

It follows from the assumptions made thus far that if two paths intersect, only one of them can involve Sinking. Consider a tree that has two copies at the bottom:

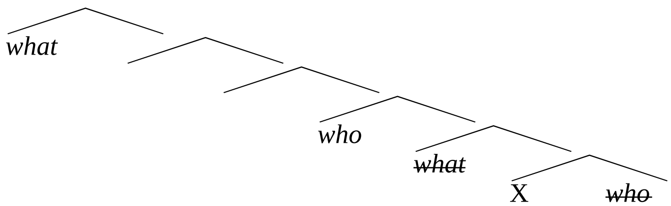
(65)



The next step in the derivation will involve introducing another head. Recall that Merge is constrained by a requirement that the newly Merged object be introduced in as low a position in the tree as possible, formalized as (48), repeated here as (66):

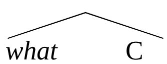
(66) Merge α to a position which c-commands as few nodes as possible

(66) would require the new head to be Merged in a position asymmetrically c-commanded by one of the two wh-word copies in (65):

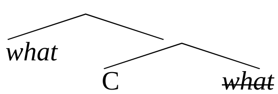
(67)


At this point in the derivation, *what* is no longer Sinking; additional material will have to be introduced in positions below the copy of *what*. In other words, intersecting paths will require that only one of the two paths be a Sinking path.

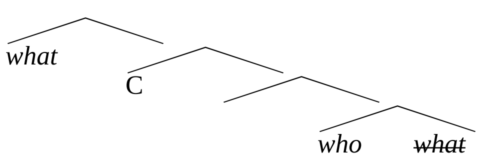
In fact, it will have to be the path which comes from the higher position in the tree which Sinks. Consider a derivation involving multiple wh-movement paths, building the tree from the top down. The derivation would begin with the highest wh-phrase and the C of which it will eventually be a specifier:

(68)


Now a decision must be made; do we create a copy of *what* immediately and begin Sinking, or do we Merge new material under C? We saw an argument in section 2 above for the assumption that Move is preferred to Merge, in this kind of derivation; I suggested that this might be a consequence of some kind of cost associated with the operation of accessing the lexicon, adapting ideas of Chomsky (1998). Avoiding lexical access in this case means making a copy of *what* and beginning a Sinking path:

(69)


Now additional material is (unavoidably) Merged, until the lower wh-phrase is reached:

(70)


We have just seen that it is impossible for both wh-phrases to Sink. Beginning a Sinking path for *who* would therefore entail leaving a copy of *what* in this position in the structure. It seems reasonable to think that leaving a copy of *what* here might be ruled out; this is not *what*'s scope position, and it has no features to check in this position. *What* will therefore have to continue to Sink, leaving *who* behind.

When wh-paths intersect, in other words, we expect that only the path which starts at the higher Spec CP will be able to sink. The "inner path" in an

English nested-paths wh-movement construction will have to undergo ordinary movement, just as an adjunct does. We therefore expect such movement paths to behave like adjuncts with respect to the diagnostics for very successive-cyclic wh-movement. Inner paths should be unable to license parasitic gaps, a fact discussed by Pesetsky (1982):

(71) a. This Volvo is the kind of car [OP_i that I know who_j
[to persuade [owners of _$_j$] to talk to __$_i$ about __$_i$]

b.*This Volvo is the kind of car [OP_i that I know who_j
[to persuade [friends of _$_j$] to talk to __$_j$ about __$_i$]

We also expect inner paths to be especially sensitive to islands, a claim which has been defended by Marantz (1994):

(72) a. *[Which congressman]$_k$ did you wonder
[which lobbyist]$_i$ to inquire whether to send ___$_i$ to ___$_k$?

b. [Which congressman]$_k$ did you wonder
whether to inquire [which lobbyist]$_i$ to send ___$_i$ to ___$_k$?

We also expect inner paths not to participate in Lebeaux's (1991) phenomenon. Here the relevant sentences are too complicated for me to judge, but I include them for completeness' sake; the prediction of this theory is that (73a) should be better than (73b)[12]:

(73) a. [Which paper that he_s gave to the $teacher_t$]$_p$ do you need to know
[which article]$_a$ every $student_s$ asked her_t to point to __$_a$ with __$_p$?

b. [Which article]$_a$ do you need to know
[which paper that he_s gave to the $teacher_t$]$_p$ every $student_s$
asked her_t to refer to __$_p$ in __$_a$?

To summarize, then; we have seen that a top-down model of the derivation makes it possible to posit a phenomenon which I have referred to here as "Sinking". Sinking is a way of dealing with successive-cyclic movement which lacks some of the conceptual drawbacks of other approaches to successive-cyclicity; in particular, there is no need to posit features to drive movement to the intermediate positions. The availability of Sinking is a natural consequence of the way Merge works in this system. Moreover, positing certain fairly

---

[12] Ben Bruening (p.c.) informs me that he gets this judgment.

natural-sounding limits on Sinking allows us to make a range of empirical predictions, which seem not obviously false.

4. Conclusions

This paper has sketched several arguments for the idea that trees should be constructed beginning at the top, following the work of Phillips (1996, to appear). The first argument involved an instance of interactions between dependencies, of a type which I argued in Richards (1997, 1998) to be useful as diagnostics for the order in which operations occur in the derivation; if these diagnostics are to be taken seriously, they seem to indicate that movements triggered by heads which are higher in the tree occur earlier in the derivation than movements triggered by heads which are lower in the tree. The second argument had to do with expletive-associate relations; I tried to show that a top-down derivation, along with a principle requiring that lexical access be avoided whenever possible (that is, that Move be preferred over Merge), yields the facts having to do with the relative placement of arguments and expletives without the need to posit Numerations, phases, or sub-Numerations. Finally, I discussed a new way of capturing the facts of successive-cyclic movement, called "Sinking", which captured some new generalizations about such movement and seemed conceptually comparatively straightforward; Sinking, as it was described here, is only available in a top-down model of the derivation. To the extent that these arguments are compelling, then, we are apparently driven to a top-down model. The consequences of the move to such a model are quite extensive, and are not adequately explored here; I must leave many problems to future research. In the appendices I will briefly touch on two unresolved issues, which I will leave just as unresolved as I have found them.

Appendix 1: Crossing and Nested Paths

The previous discussion has mentioned a number of phenomena involving intersecting movement paths which are forced to either cross or nest. Some of these phenomena have been discussed in terms of bottom-up derivations; here I will briefly try to provide an account of them using the top-down derivation for which I have argued here.

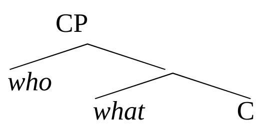
One instance in which multiple movement paths are required to cross is the case of movement to multiple specifiers; this is exemplified by Bulgarian wh-movement in (74):

(74) a. Koj kogo vižda ?
who whom sees
'Who sees whom?'

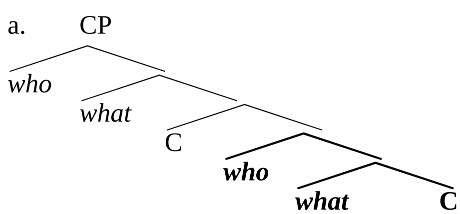
b.* Kogo koj vižda ?
whom who sees (Rudin 1988, 472-473)

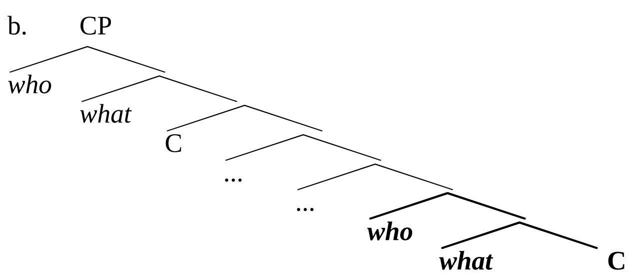
Mulders (1997) and Richards (1997, to appear) develop accounts of this fact which make crucial use of a bottom-up derivation. Here I will try to capture the same facts while constructing the tree from the top down.

The relevant derivation will involve the notion of Sinking, introduced in section 3 above. Suppose we consider a derivation that begins with a CP with multiple wh-phrases in its specifiers:

(75)


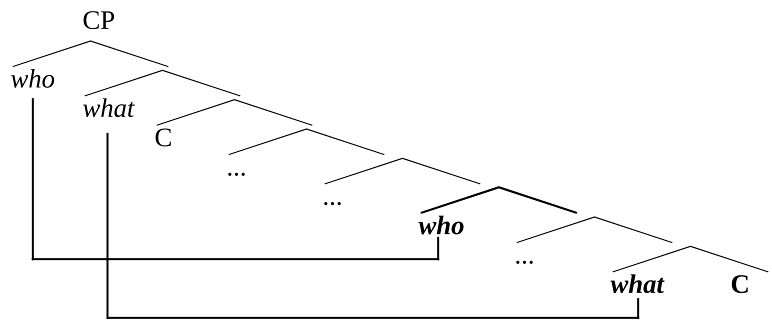
Suppose that the next step of the derivation could involve Sinking of the entire CP; a copy of the CP is created and begins to Sink:

(76) a.


b.


The next step involves introducing material between the two wh-phrases; the higher wh-phrase thus reaches its landing site, while the lower one continues to Sink:

(77)



This derivation has the desired result; the higher wh-phrase has the higher landing site, and the lower one has the lower landing site (in other words, the paths cross). The movement paths are drawn in for ease of processing.

There are a number of wh-in-situ languages which offer a bit of support for the idea that the derivation works in this way. The languages in question have a particle associated with the wh-complementizer attached to wh-phrases:

(78) mokak- **də** wætun**e**?
what Q fell -Q
'What fell?' (*Sinhala* (Hagstrom 1998, 36))

(79) taa -ga -**GA** ringo kamta**ra**
who NOM Q apple ate -Q
'(I wonder) who ate the apple' (*Shuri Okinawan* (Sugahara 1996, 250))

The derivation sketched above offers a natural interpretation of the origin of the question particles attached to the wh-words above; they could be interpreted as a morphological reflex of the copy of C posited in (75-77). Interestingly, in multiple-wh questions there is only one particle, and it is preferentially associated with the lowest wh-phrase:

(80) kauru mokak-**də** kieuw**e**?
who what Q read-Q
'Who read what?' (*Sinhala* (Hagstrom 1998, 65))

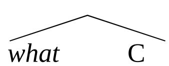
(81) taa -ga nuu -**GA** kamta**ra**
who NOM what Q ate-Q
'(I wonder) who ate what' (*Shuri Okinawan* (Sugahara 1996, 247))

Again, this is as we expect; the derivation in (75-77) ends with a representation involving a copy of C attached to the lowest wh-phrase.

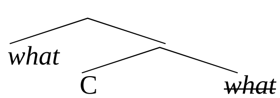
Suppose we now consider cases involving multiple attractors. In English, wh-movement to distinct attractors is required to yield nested paths; this is Pesetsky's (1982) Path Containment Condition:

(82) a. Who do the police want to know who __ killed __?
b. * Who do the police want to know who __ killed __?

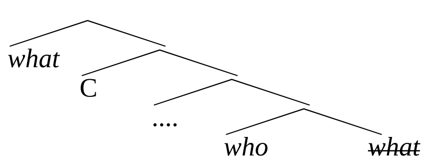
An elegant way of deriving the PCC in terms of a bottom-up derivation has been proposed (by Kitahara (1997)). Ideally, we would like to be able to show that the PCC can still be derived, even when the derivation is from the top down. In fact, the contrast in (82) can be largely accounted for by notions invoked in the earlier parts of this paper. We saw in section 3 (the section on Sinking) that when movement paths intersect, only one of them can Sink (this was a consequence of the requirement that new material Merge as low in the tree as possible). We also saw, in sections 2 and 3 above, that lexical access is avoided whenever possible; that is, operations involving copy of material already in the tree (i.e., Move) are preferred to Merge of new material from the lexicon. Suppose we consider a derivation for examples like those in (82), bearing these requirements in mind. The derivation will begin at the top of the tree, with a complementizer and the wh-word which is in its specifier:

(83)


Because lexical access has a cost, the best thing to do at this point, if it is possible, is to create a copy of something already in the tree. This is the beginning of a derivation involving Sinking; because we can determine by inspection that *what* will check features lower in the tree (Case features, if nothing else), it is in fact possible to create a copy of *what* immediately:

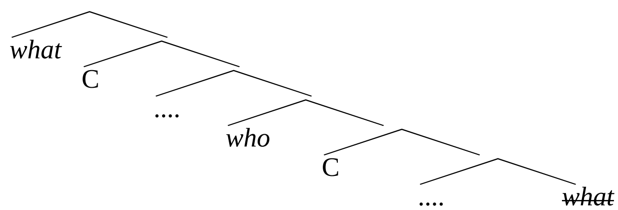
(84)


Next we must Merge new material (since there is no longer anything to copy); the new material can be merged in such a way that *what*'s copy will Sink through the tree. Eventually the next wh-phrase is reached:

(85)


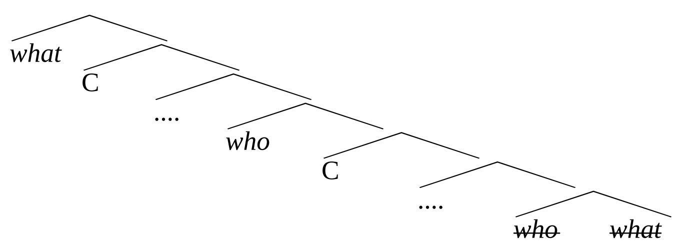
As we saw in the previous section, it is not possible for both wh-phrases to Sink; the next head to be introduced will split up the copies of *what* and *who*. Thus, we must choose between leaving a copy of *what* in this position and beginning to Sink *who*, or continuing Sinking *what*. Since this is not a position in which *what* will check features, it seems reasonable to rule out leaving a copy of *what* here[13]. *What* therefore continues to Sink.

(86)



Now consider the point in the derivation at which the first wh-trace position is reached. Here we must make a decision between two options. One option would be to use the copy of *what* as the wh-trace; the next operation, then, would be to access the lexicon and Merge new material under the copy of *what*. Another option would be to create a copy of *who* in the trace position. Since lexical access is avoided whenever possible, the second of these options is taken; a copy of *who* is created, thus delaying having to access the lexicon to get new material, and the copy of *what* continues to Sink:

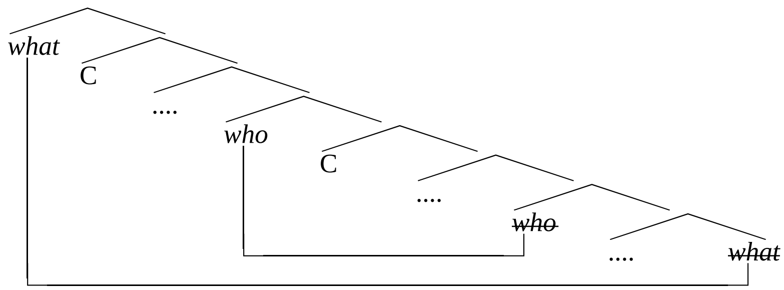
(87)



Finally, new material is inserted between the two wh-copies, until the copy of *what* reaches its trace position:

[13] Leaving a copy of *what* behind and beginning to Sink *who* would have the virtue of delaying lexical access further; in the account so far, at any point in the derivation at which the creation of a copy is possible, copying is preferred to other options. A careful definition of the circumstances under which copying is “possible” is clearly in order here; it will have to be impossible in this instance, in the relevant sense. I will have to leave this important issue for further work.

(88)



As desired, the result of this derivation is a representation with nested wh-movement paths.

There is a general and quite serious problem with the account developed thus far, which I will have to leave unresolved here. I have been claiming that lexical access is avoided whenever possible; this was the principle underlying the preference for Move over Merge, which has been useful to us so far both in the discussion of expletive-associate relations (section 2 above) and in deriving the Path Containment Condition. There are problems with the principle as stated, however. Consider the case of a movement which does not involve Sinking. Because lexical access is avoided whenever possible, a non-Sinking movement should be required to land in the closest possible landing site. Once the part of the tree which could act as a landing site for movement is created, we must choose between creation of a copy of the moving object and accessing the lexicon to put some other lexical item in that part of the structure. The theory developed thus far would lead us to expect a preference for movement under such circumstances. For instance, I have claimed that adjuncts are unable to Sink; if this is correct, the theory apparently leads us to the prediction that (89b) should be ill-formed, blocked by (89a) (assuming that *when* is an adjunct in the relevant sense):

(89) a. When did Bill say __ [that Susan will resign in September]?

b. When did Bill say at 2:00 [that Susan will resign __]?

The question of how to distinguish the case in (89) from the other cases discussed above is a very difficult one, which I hope to address in future work.

Let us now turn our attention to languages in which the opposite of the Path Containment Condition holds. These languages include Bulgarian (90) and Chinese (91), as shown below:

(90) a.* Kakvo se opitva da razbere na kogo dade Ivan ___ ___?
what SELF try to find-out to whom gave Ivan
'What is he trying to find out to whom Ivan gave?'

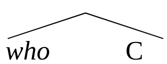
b. Na kogo se opitva da razbere kakvo dade Ivan ___ ___?
to whom SELF try to find-out what gave Ivan
'To whom is he trying to find out what Ivan gave?'

(91) jingcha xiang-zhidao [shei sha -le shei]
police want know who kill PERF who

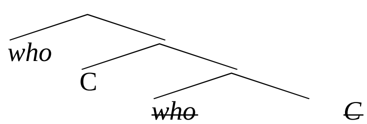
(i)*'Who are the police trying to find out who ___ killed ___?'

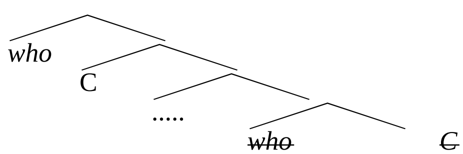
(ii) 'Who are the police trying to find out who ___ killed ___?'

Let us consider the derivation from the beginning, starting with the highest wh-phrase and the C of which it will be the specifier:

(92)


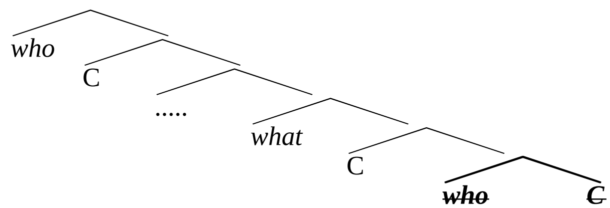
In considering the derivation involved in the Path Containment Condition case in English, the derivation began by Sinking the wh-word. We could in principle have begun by Sinking the entire constituent in (92), as we did in the derivation for movement to multiple specifiers of a single head; nothing crucial would have been affected by this move. Suppose we perform that kind of Sinking here:

(93) a.


b.


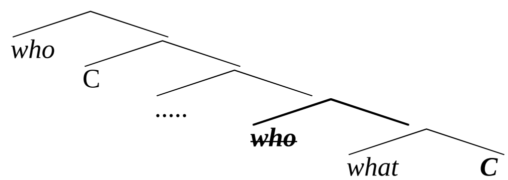
Eventually the Sinking constituent reaches the embedded clause, and a decision must be made. We could insert the second wh-word (*what* in the tree in (94)), and a new C, above the Sinking material, yielding the structure in (94):

(94)



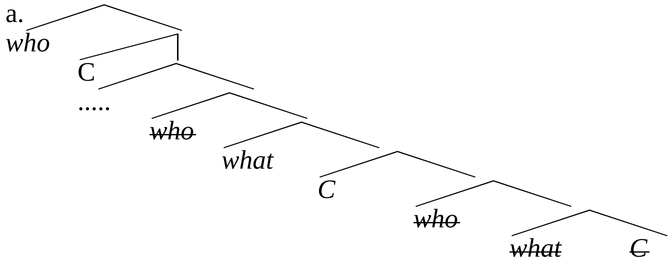
Alternatively, we could simply add the second wh-word, reusing the Sinking C as the C of the embedded clause:

(95)

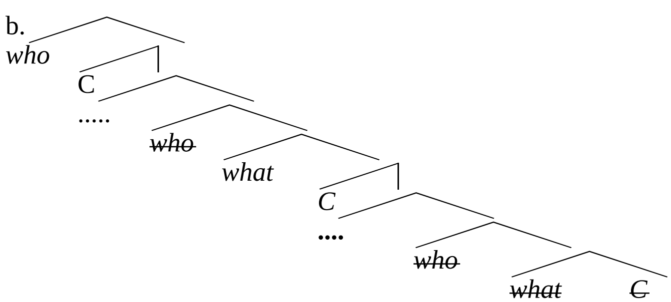


If reusing the Sinking C is possible, then the requirement that Merge be to a maximally low position in the tree will force *what* to Merge as the sister of C, as in (95). Moreover, the option in (95) involves less lexical access than the one in (94), so (95) is clearly to be preferred, all other things being equal. From (95), the derivation could proceed as in the derivation for movement to multiple specifiers given before (in (75-77)). This derivation yields crossing paths, as we have seen:

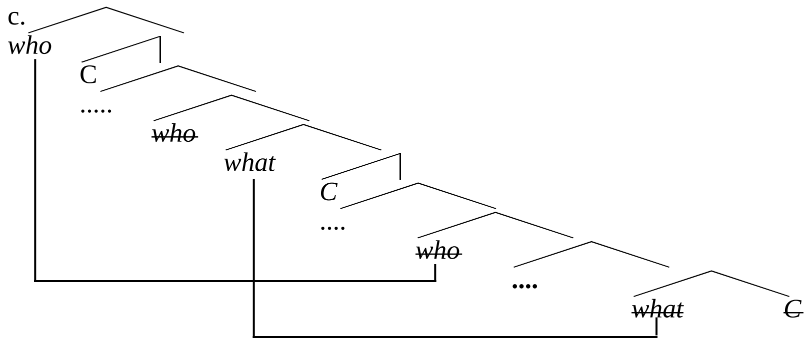
(96) a.



b.



c.



Here the paths of movement are predicted to preferentially cross, as desired.

All that remains is to determine why this kind of derivation is unavailable in English. According to the theory as developed so far, the derivation just sketched is preferable to the alternative derivation given to English; it involves less lexical access, "reusing" the Sinking complementizer as the complementizer for the embedded clause. If we could explain why English differs from languages like Bulgarian and Chinese in failing to allow reuse of its complementizers, we might be in a position to explain the distribution of Path Containment Condition effects and their opposite. In fact, it is not particularly surprising that English should differ from the other languages in this way; English complementizers, in Minimalist terms, have a single strong feature which must trigger one instance of overt wh-movement. Once this strong feature has been checked, the English complementizer has different properties than it did at the beginning of the derivation, lacking the feature that would motivate overt movement. Compare, on the other hand, the behavior of complementizers in languages like Bulgarian or Chinese; in such languages, a complementizer may be involved in arbitrarily many overt wh-movements (Bulgarian) or none (Chinese). Once the English complementizer has participated in one wh-movement, then, it cannot be reused, since it has had its strong feature checked off; the derivation in (92-96), which would yield crossing paths, is therefore unavailable in English. Chinese and Bulgarian complementizers, on the other hand, do not change regardless of how often they participate in wh-movement, and therefore may be reused arbitrarily many times.

There does appear to be one instance in which the Path Containment Condition resurfaces in languages like Bulgarian and Chinese. An additional wh-movement whose landing site is in the higher clause triggers a preference for nesting of the movement paths whose tails are both in the embedded clause:

(97) a. Koj kakvo ___ se opitva da razbere na kogo dade Ivan __ __?
who what SELF try to find-out to whom gave Ivan
'Who is trying to find out to whom Ivan gave what?'

b. * Koj na kogo __ se opitva da razbere kakvo dade Ivan __ __ ?
who to whom SELF try to find-out what gave Ivan
'Who is trying to find out what Ivan gave to whom?'

(98) shei xiang-zhidao [shei sha -le shei]?
who want know who kill PERF who

(i)'Who$_k$ who$_j$ t$_k$ is trying to find out who$_i$ t$_i$ killed t$_j$?'

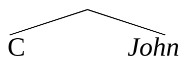
(ii)*'Who$_k$ who$_i$ t$_k$ is trying to find out who$_j$ t$_i$ killed t$_j$

In this derivation, the higher C participates in a movement which lands before the higher C Sinks into the embedded clause. We might conjecture that the preference for nested paths arises here, again, from some factor which prevents "reuse" of the complementizer. One candidate for the relevant factor is the PMC; the higher complementizer has participated in a well-formed dependency by the time it Sinks into the embedded clause, and perhaps this is enough to make it unusable. Alternatively, perhaps the second wh-phrase stops Sinking once the landing site for the first wh-phrase is reached, and undergoes ordinary movement (copying) to the lower trace position, as in English. Deciding between these theories will involve finding evidence bearing on the question of whether the wh-movement from the higher C into the embedded clause moves there by Sinking or not; I do not yet have evidence bearing on this question.

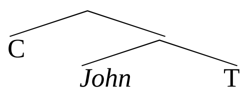
This appendix has reviewed a number of cases described in the literature in which movement paths are required to either cross or nest. I have tried to show that the relevant patterns can be derived moderately straightforwardly, though many unresolved issues still remain, some of which I have mentioned briefly. Notions playing a role in the account included avoiding lexical access wherever possible, the Sinking operation developed in section 3 above, and the notion of "reuse" of a Sinking complementizer, along with certain assumptions about when such reuse is possible.

Appendix 2: Top-Down Derivation and Checking Relations

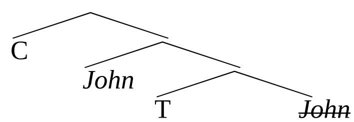
The derivations discussed in this paper have been, in large part, mirror images of the derivations generally postulated in Minimalism. This has the result that objects are sometimes introduced before there is any obvious "motivation" for introducing them. Consider a straightforward top-down derivation for a sentence like *John ate mackerel*. The sentence might begin by Merging C together with the NP *John*:

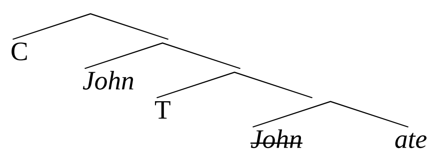
(99)


Next T would be introduced, destroying the constituent in (99):

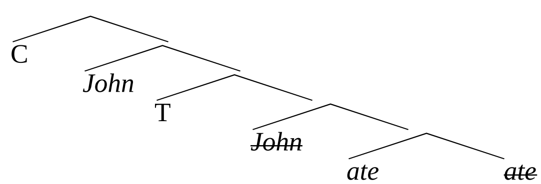
(100)


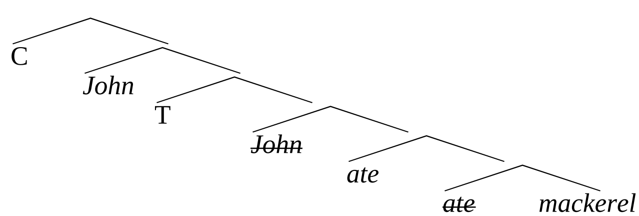
The next step would be the construction of vP, introducing a copy of *John* and then the v head (occupied, I assume, by the main verb):

(101) a.


b.


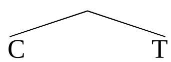
Finally, we would create the VP, copying the main verb into its lowest position and adding the direct object:

(102) a.


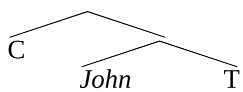
b.


The order in which syntactic objects are introduced in this top-down derivation is simply the opposite of that in a Minimalist derivation. One result of this is that new material is sometimes introduced before there is any motivation to do so. There is, for instance, no reason to Merge C and *John* at the beginning of the derivation. On standard assumptions, these two objects bear no relation to each other; there is no selectional relation between them, and neither checks features of the other[14].

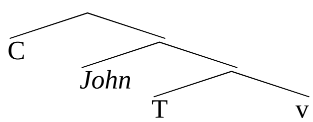
We could imagine another kind of derivation of this sentence which would not have this property, but which would obey the requirement that new material be introduced as low as possible in the tree[15]. Suppose we were to begin the derivation by Merging not C and the subject, but C and T, a Merge operation motivated by selectional restrictions of C:

(103)


Features on T that need to be checked would then permit Merge of *John*:

(104)


T's selectional requirements might then drive Merge of v:

(105)


The derivation could continue in this way, with each step being motivated by requirements of material already in the workspace, and with each new piece of syntactic material being introduced in a position c-commanding as little of the existing tree as possible. Introducing Sinking into such a derivation would not be trivial; I will not attempt it here.

Regardless of which of these versions of the top-down derivation is used, these derivations have the interesting property that heads enter into sisterhood relations, at some point in the derivation, with all of the objects that they are commonly considered to have checking or selectional relations with. In the derivation fragment in (103-105), for instance, T is the sister of *John* at (104) and of the head of vP at (105). These derivations thus might allow for a particularly simple restatement of checking and selectional relations; they take place under sisterhood[16].

---

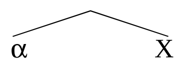
[14] One possible reaction to this state of affairs would be to abandon some standard assumptions. For instance, there sometimes do appear to be syntactic relations of various kinds between complementizers and subjects, including complementizer agreement (in languages like Bavarian German) and apparent Case assignment. It is not obvious that this kind of approach could account for all of the relevant facts, but it might be worth pursuing to some extent.

[15] Colin Phillips (p.c.) tells me that Schneider (1999) develops a theory similar to this one; I have not yet had a chance to read this work.

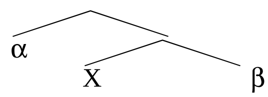
[16] Again, this ceases to be true once Sinking is introduced. This may be an argument against introducing Sinking, or it may be the beginning of an account of certain kinds of

The set of XPs that may be assigned theta-roles by a head, on this theory, is interestingly similar to the definition developed by Pesetsky (1995). Pesetsky's claim is essentially that a head may assign a theta-role to its specifier, to its sister, or to the specifier of its sister. In a top-down derivation, all of these XPs can be sisters of a theta-assigning head X at some point in the derivation:

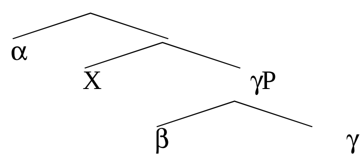
(106) a.



b.



c.



The theory developed here differs from Pesetsky's in that theta-assignment could in principle continue infinitely; not only the specifier of the sister of X, but also the specifier of the specifier of the sister of X, and so forth, can be X's sister at some point in the derivation. This is not an obviously desirable result.

Bibliography

Agbayani, Brian (1997). Category raising, adjunction, and minimality. *UCI Working Papers in Linguistics* 3, 1–26.

Agbayani, Brian (1998). *Feature Attraction and Category Movement*. Doctoral dissertation, University of California, Irvine.

Bobaljik, Jonathan (1995). *Morphosyntax: the Syntax of Verbal Inflection*. Doctoral dissertation, MIT. Distributed by MITWPL.

Bobaljik, Jonathan (1999). A-chains at the interfaces: copies, agreement and "covert" movement. Ms., McGill University.

Borsley, Robert (1996). *Modern Phrase Structure Grammar*. Oxford: Blackwell.

Bošković, Zeljko (1995). Superiority effects with multiple wh-fronting in Serbo-Croatian. Ms., University of Connecticut.

Chomsky, Noam (1986). *Barriers*. Cambridge, Mass.: MIT Press.

Chomsky, Noam (1993). A minimalist program for linguistic theory. In Kenneth Hale and Samuel Jay Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, Mass.: MIT Press, pp. 1–52.

Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, Noam (1998). Minimalist inquiries: the framework. Ms., MIT.

---

islandhood; Sinking past certain things is impossible because it would block checking relations. See, for instance, Uriagereka's (1999) claim that wh-movement past certain kinds of agreement relations is disallowed. Much work would have to be done to make this plausible, however.

Cole, Peter, and Gabriella Hermon. (1998) The typology of WH-movement: WH questions in Malay. *Syntax* 1.2.

Epstein, Sam, Erich Groat, Ruriko Kawashima, and Hisatsugu Kitahara (1998). *A Derivational Approach to Syntactic Relations*. Oxford: Oxford University Press.

Fox, Danny (1999). Reconstruction, binding theory, and the interpretation of chains. *Linguistic Inquiry* 30:157–196.

Fox, Danny, and Jon Nissenbaum (1999). Extraposition and scope: a case for overt QR. Ms., MIT.

Groat, Erich, and John O'Neil (1996). Spell-out at the LF interface. In Werner Abraham, Samuel Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart (eds.), *Minimal Ideas*, Amsterdam: John Benjamins, pp. 113–139.

Hagstrom, Paul (1998). *Decomposing Questions*. Doctoral dissertation, MIT. Distributed by MITWPL.

Huang, James (1982). *Logical Relations in Chinese and the Theory of Grammar*. Doctoral dissertation, MIT.

Kayne, Richard (1983). Connectedness. *Linguistic Inquiry* 14:223–249.

Kitahara, Hisatsugu (1997). *Elementary Operations and Optimal Derivations*. Cambridge, Mass.: MIT Press.

Kroch, Anthony (1989). Asymmetries in long-distance extraction in a Tree-Adjoining Grammar. In Mark Baltin and Anthony Kroch (eds.), *Alternative Conceptions of Phrase Structure*. Chicago: The University of Chicago Press, pp. 66-98.

Lasnik, Howard, and Mamoru Saito (1992). *Move α: Conditions on its Application and Output*. Cambridge, Mass.: MIT Press.

Lebeaux, David (1991). "Relative clauses, licensing, and the nature of the derivation," *Syntax and Semantics* 25.

Manzini, Maria (1992). *Locality*. Cambridge, Mass.: MIT Press.

Marantz, Alec (1994). The two modes of A-bar binding are not tied to the argument/adjunct distinction. Conference abstract.

Mulders, Iris (1997). Mirrored specifiers. *Linguistics in the Netherlands 1997*.

Nissenbaum, Jon (1998). Movement and derived predicates: evidence from parasitic gaps. In Uli Sauerland and Orin Percus (eds.), *MITWPL 25: The Interpretive Tract*. Cambridge, Mass.: MIT Working Papers in Linguistics.

Pesetsky, David (1982). *Paths and Categories*. Doctoral dissertation, MIT. Distributed by MITWPL.

Pesetsky, David (1995). *Zero Syntax*. Cambridge, Mass.: MIT Press.

Phillips, Colin (1996). *Order and Structure.* Doctoral dissertation, MIT. Distributed by MITWPL.

Phillips, Colin (to appear). Linear order and constituency. *Linguistic Inquiry*.

Reinhart, Tanya (1995). Interface strategies. OTS Working Papers, Utrecht.

Richards, Norvin (1997). *What Moves Where When in Which Language?* Doctoral dissertation, MIT. Distributed by MITWPL.

Richards, Norvin (1998). The Principle of Minimal Compliance. *Linguistic Inquiry* 29:599–629.

Richards, Norvin (to appear). Featural cyclicity and the ordering of multiple specifiers. To appear in Sam Epstein and Norbert Hornstein (eds.), *Working Minimalism*. Cambridge, Mass.: MIT Press.

Rudin, Catherine (1988). On multiple questions and multiple wh-fronting. *Natural Language and Linguistic Theory* 6:445–501.

Saito, Mamoru, and Naoki Fukui (1998). Order in phrase structure and movement. *Linguistic Inquiry* 29:439–474.

Sauerland, Uli (to appear). Scope reconstruction without reconstruction. To appear in *Proceedings of WCCFL 17*.

Schneider, Dave (1999). *Parsing and Incrementality*. Doctoral dissertation, University of Delaware.
Sugahara, Mariko (1996). Shuri Okinawan *kakari-musubi* and movement. In Masatoshi Koizumi, Masayuki Oishi, and Uli Sauerland (eds.), *MITWPL 29: Formal Approaches to Japanese Linguistics 2*. Cambridge: MIT Working Papers in Linguistics.
Tada, Hiroaki (1993). *A/A-bar Partition in Derivation*. Doctoral dissertation, MIT.
Takahashi, Daiko (1994). *Minimality of Movement*. Doctoral dissertation, University of Connecticut, Storrs. Distributed by MITWPL.
Uriagereka, Juan (1999). Minimal restrictions on Basque movements. *Natural Language and Linguistic Theory* 17:403–444.
Uriagereka, Juan (to appear). Multiple spell-out. To appear in Sam Epstein and Norbert Hornstein (eds.), *Working Minimalism*. Cambridge, Mass.: MIT Press.
Watanabe, Akira (1992). Subjacency and S-structure movement of wh-in-situ. *Journal of East Asian Linguistics* 1:255–291.

Norvin Richards
Department of Linguistics and Philosophy
E39–245 MIT
77 Massachusetts Avenue
Cambridge MA 02139
USA

norvin@mit.edu