

# Arguments for Top-down Derivations in Syntax

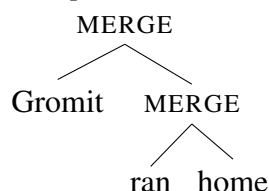
Robert Frank & Hadas Kotek\*

**Abstract.** This paper argues for a top-down approach to syntactic structure building, instead of the more popularly assume bottom-up approach. We present phenomena where properties of a higher structure condition elements or operations in a lower domain. The opposite pattern is, we claim, much rarer and more restricted. This follows from a top-down approach, given the relative time at which elements are integrated into the structure. Bottom-up derivations, in contrast, would naturally lead to opposite predictions. Our goal is not to show that bottom-up theories cannot explain the facts, but rather that a top-down account is more natural and less stipulative.

**Keywords.** top-down vs. bottom-up, derivational timing, superiority, negative concord, sequence of tense, resumptive pronouns

**1. Derivational Directionality and its Genesis.** Derivational theories provide an account of the properties of natural language syntax in terms of the operations that are employed to construct the representation of a sentence. Typically, such derivations proceed in a “bottom-up” fashion: morphemes and words are composed to form larger constituents and these can in turn be further combined with other constituents. Thus, a simple sentence like *Gromit ran home* is derived by first combining *ran* and *home*, and the result is then combined with *Gromit*. This sequence of derivational steps can be represented in the form of a *derivation structure*: each (non-leaf) node corresponds to a derivational operation, MERGE, where the daughters serve as the input to the operation.

(1) *A simple derivation structure:*



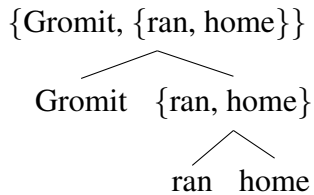
This derivation structure encodes derivational ordering via its dominance relation: an operation  $O_1$  precedes another operation  $O_2$  if the node corresponding to  $O_1$  is dominated by  $O_2$ . Thus the derivation structure given here represents the fact that the MERGE of *ran* and *home* precedes the merge of *Gromit*.

Because derivational operations produce output, each node in the derivation structure can also be annotated with the object that results from the associated operation. Taking the combinatory operation to be the minimalist Merge operation, this result will be a set containing the elements that were merged, giving rise to a representation that closely corresponds to the usual constituent structure.

---

\* We are grateful to comments from Milena Šereikaitė, Jim Wood, and audiences at the Berkeley Syntax Circle and the Yale Syntax Group. Authors: Robert Frank, Yale University (bob.frank@yale.edu) & Hadas Kotek, MIT (hkotek@alum.mit.edu).

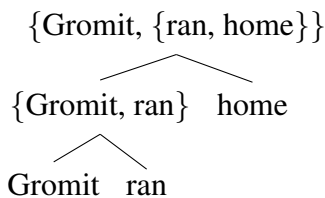
(2) *A simple constituent structure:*



One can entertain different hypotheses about the set of derivational operations that the grammar permits, and different choices will give rise to different sequences of operations and correspondingly different derivation structures. However, the structure that results from these differing derivational paths can nonetheless be preserved. This is seen famously in Combinatory Categorical Grammar (Steedman 2000), where the addition of type shifting and function composition to the operation of function application allows for both left and right branching derivations of a transitive sentence, both of which produce the same LF.

Phillips (2003) explores the consequences of a derivational system that incorporates an operation that builds structure in a top-down/left-to-right fashion. Under this view, the derivation structure for the sentence *Gromit ran home* is left branching: *Gromit* first combines with *ran* and the result is combined with *home*. Phillips assumes a combinatory operation, Merge Right, in which the newly merged element enters into the structure as part of a constituent formed with some element that is already part of that structure at its right edge. Doing this, the object that is produced by the derivation is identical:

(3) *A top-down derivation structure:*



We will call this a *top-down derivation*, because the constituent structure that is produced by the derivation is expanded from the root downward. We note, though, that there is a sense in which this derivation remains bottom-up, as is usually assumed. If we attend to the derivation tree, more deeply embedded nodes are constructed prior to nodes that embed them. As a result, the construction of the derivation tree will always proceed in a bottom-up fashion. In our discussion, we will use the top-down/bottom-up terminology to reference the order in which the constituent structure is built, as opposed to the order of the derivation.

The bottom-up approach to derivations has been by far the dominant approach since the earliest work in generative grammar (Chomsky 1955, 1957). This work develops a system of grammatical derivations in which clausal “kernel” structures are generated by a phrase structure base and then modified using *singular transformations* (to form questions, passives and negated clauses) and combined using *generalized transformations* (creating instances of clausal embedding and conjunction). Fillmore (1963) notes, however, that such a system permits interactions between generalized and singular transformations that are unattested: while some singular transformations must apply to a kernel structure before it is embedded in another with a generalized transformation, there are no cases in which a singular transformations must apply to a structure before something is embedded into it. Chomsky (1965) proposes a solution to this

overgeneration conundrum: by allowing the base rules to apply recursively, Deep Structures of unbounded size can be created without the use of generalized transformations. Because there are no more generalized transformations, there is no longer any need to regulate the interaction between them and singular transformations. Further, Chomsky proposes that singular transformations (now simply transformations) apply in a bottom-up fashion, working their way from lower “cyclic” domains to higher ones. Though this proposal for bottom-up derivations was not extensively explored in Chomsky (1965), it is elaborated in Chomsky (1973; pp.243–244), where a bottom-up derivational direction is used to constrain *wh*-movement. In particular, Chomsky’s goal was to rule out *wh*-island violations like the following:

- (4) *A wh-island*:
- a. D-Str: [<sub>S<sub>1</sub></sub> COMP he wondered [<sub>S<sub>2</sub></sub> COMP John put what where]]
  - b. S-Str: \*What did he wonder where John put?

The widespread assumption behind the impossibility of such examples is that the presence of the *wh*-word *where* in the lower COMP position prevents a derivational step in which the word *what* is moved to the higher (matrix) COMP. However, one could imagine an alternative derivation for (4) in which the *wh*-phrase *what* moves first to the matrix COMP position (5-a), unimpeded by the intervening *where*. Then, in a subsequent derivational step, *where* is moved to the lower COMP (5-b):

- (5) *A potential derivation for (4), to be ruled out*:
- a. Step 1: *what* moves directly into the matrix clause  
[What did he wonder [COMP John put where]]?
  - b. Step 2: *where* moves into the specifier of the embedded clause  
[What did he wonder [where John put]]?

This kind of derivation is ruled out by the following condition:

- (6) *The Strict Cycle Condition (SCC)* (Chomsky 1973; p.243):  
No rule can apply to a domain dominated by a cyclic node *A* in such a way as to affect solely a proper subdomain of *A* dominated by a node *B* which is also a cyclic node

The SCC blocks (5) precisely because the second movement step impacts only the domain determined by a lower cyclic node (the embedded clause) that is a proper subdomain of the higher cyclic domain of the main clause.

Chomsky’s paraphrase of the SCC makes clear its connection to derivational order: “In other words, rules cannot in effect return to earlier stages of the cycle after the derivation has moved to larger, more inclusive domains.” On that reading, it is important to observe that the SCC does not per se require that derivations proceed in a bottom-up manner, only that it proceed in way so as to apply to ever larger substructures. This is compatible with top-down derivations if the relevant substructures are derivation, as opposed to derived, trees. Thus the SCC would be satisfied if an operation cannot apply so as to affect only a higher domain once a lower domain had been integrated, and indeed such a treatment of *wh*-islands is not difficult to formulate.

The analogously problematic top-down derivation would involve the matrix *wh*-phrase *what* being moved (lowered) unimpeded into its argument position in the embedded clause, followed by the insertion and movement of the intermediate *wh*-phrase *where*. Such a derivation requires the return to a higher domain (the embedded CP) at a point in the derivation after which a lower

one (the embedded VP) has been targeted. It would be ruled out by the SCC for the same reason that the SCC rules out (5).

Because bottom-up and top-down derivational regimens can be devised to produce similar results, it has proven difficult to provide arguments in favor of one over the other.<sup>1</sup> However, these two modes of derivation will inevitably differ in terms of the informational content present in a structure at the time that an element is merged into it. Under a bottom-up derivation, an element is introduced into a structure prior to its embedding context, but after the structure it c-commands. In contrast, under a top-down derivation, an element is introduced into a structure that already includes its embedding context, but not the context that it c-commands. Our goal in this paper is to use this temporal asymmetry in the availability of information to argue for top-down derivations. Specifically, we present phenomena where properties of a *higher* structure condition the viability of an element or an operation that can apply in a lower domain. This is precisely the pattern expected under a top-down derivational scheme, but not a bottom-up one. Further, we argue that patterns where properties of a higher domain are conditioned by lower ones are considerably rarer and more restricted.

Our arguments will consist of two sorts of case studies. The first, which we will call syntactic, involve cases in which the possibility of a syntactic operation, specifically movement, depends on information about the higher context. The second kind of argument will be semantic in character. Here, the dependence on higher context is seen in interpretation. Under the assumption that semantic interpretation takes place in some, possibly coarse-grained, incremental fashion, we expect that the interpretation of an element can be conditioned by what has already been introduced into the derivation. On a top-down approach, this is information in higher, embedding domains, while a bottom-up approach this would be information in lower, embedded domains.

We note that our goal is not to show that bottom-up theories cannot be constructed which explain the facts we review. Nor will we present a full detailed top-down analysis of each of our case studies. Instead, we will show that a top-down view naturally contains the ingredients needed to build an analysis of the facts, whereas bottom-up accounts require additional stipulations. Our main goal is to stimulate discussion and encourage other linguists to view problems from this perspective, which we believe could lead to productive new research avenues.

## 2. Syntactic Arguments.

2.1. SUPERIORITY. As Kuno & Robinson (1972) and Chomsky (1973) observe, English exhibits a restriction on *wh*-movement in multiple *wh*-questions: the *wh*-phrase that is associated with the syntactically higher base position must raise.<sup>2</sup>

---

<sup>1</sup> One line of argument that has sometimes been adduced in favor of top-down derivations comes from their potential connections to a model of incremental sentence processing: the cognitive operations attributed to a hearer or speaker during sentence comprehension and production could plausibly be identified with the operations that underly grammatical theory Steedman (1989); Phillips (1996). This consideration is, however, not decisive. Stabler (1991) and Shieber & Johnson (1993) argue that even right-branching structures are fully compatible with incremental comprehension, so long as we avoid the assumption that interpretation cannot begin on a constituent before it is syntactically complete. However, such a view requires a processor that makes use of computational steps that go beyond the grammatical operations themselves. In any case, we put aside this consideration in our discussion below and instead consider only distributional arguments about derivational directionality.

<sup>2</sup> We include a matrix clause in (8) to ensure that the subject *wh*-phrase has moved out of its base position.

(7) *A superiority effect in matrix questions:*

- a. Who *t* bought what?
- b. \*What did who buy *t*?

(8) a. Who do you think [*t* bought what]?

- b. \*What do you think [who bought *t*]?

It is unsurprising to note that we observe an identical Superiority effect in embedded questions.

(9) *A superiority effect in embedded questions:*

- a. I wonder [what you gave *t* to whom].
- b. \*I wonder [who you gave what to *t*]

(10) a. I wonder who (you think) [*t* bought what].

- b. \*I wonder what (you think) [who bought *t*].

Contrary to expectations, however, such embedded Superiority violations are ameliorated just in case that there is a *wh*-phrase in the matrix clause (Huang 1982; Lasnik & Saito 1992):

(11) *Superiority effects are ameliorated in the presence of a matrix wh:*

- a. Who wonders [who you gave what to *t*]?
- b. Who wonders [what (you think) who bought *t*]?

As Lasnik & Saito note, these examples are acceptable only when the embedded in-situ *wh*-phrase is construed with matrix scope, such that the answer provides a pair or list of pairs. That is, an answer to (11-a) requires a pair, as in (12-a), corresponding to a multiple question construal of (11-a). In contrast, the answer cannot simply fill in the value for the matrix *wh*-phrase as in (12-b) corresponding to a simplex matrix question construal of (11-a).<sup>3</sup>

(12) *Exceptional superiority violations as in (12) must have a list-of-pair answer:*

- a. ✓ Alice wonders who you gave the book to. *multiple question interpretation*
- b. \*Alice wonders who you gave what to. *simplex matrix question interpretation*

Crucially, matrix scope for the in-situ *wh*-phrase is possible in English only when there is another *wh*-phrase from a higher clause that has itself moved to the specifier of the matrix CP. In the absence of such a *wh*-phrase the example is ungrammatical on any interpretation, as already seen in (9)–(10).

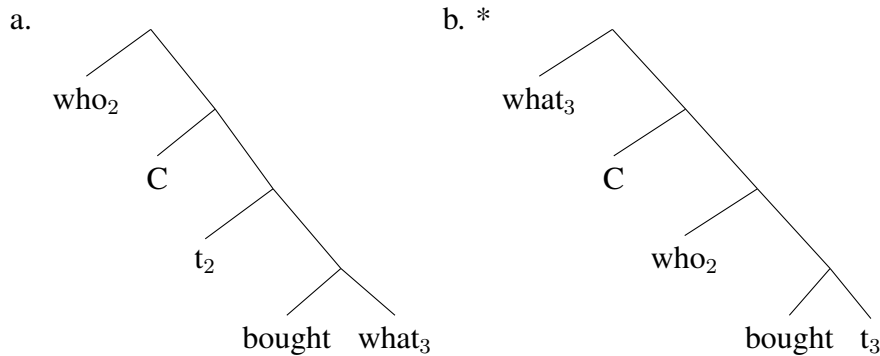
Current Minimalist analyses (e.g. Rizzi 1990; Pesetsky 2000; Kotek 2019) take the Superiority phenomenon to derive from a locality condition on the relation between the head that triggers movement, which we will take to be C, and the phrase it is attracting. At the relevant point in a bottom-up derivation, C probes for the closest element with appropriate [+WH] features. The usual assumption is that probing initiates a search process, which proceeds from closer to more distal elements that are c-commanded by the probe.

Taking (7) as an example, the tree structures in (13)a–b illustrate the two logically possible derivations of a question with two *wh*-words. In both structures, the first element with relevant features to be identified by C is the subject *wh*-phrase (*who*<sub>2</sub>): it is c-commanded by C and asym-

<sup>3</sup> On the other hand, the superiority-obeying variants of (11) can be construed with either matrix or embedded scope for the in-situ *wh*-phrase. For example, *who wonders what you gave to whom?* can be answered either with (a) Alice wonders what you gave to Bob or with (b) Alice wonders what you gave to whom.

metrically c-commands the object *wh*-phrase (*what*<sub>3</sub>). In (13)a, the subject is moved to Spec,CP (internally merged with the entire structure), yielding the superiority-obeying word-order.

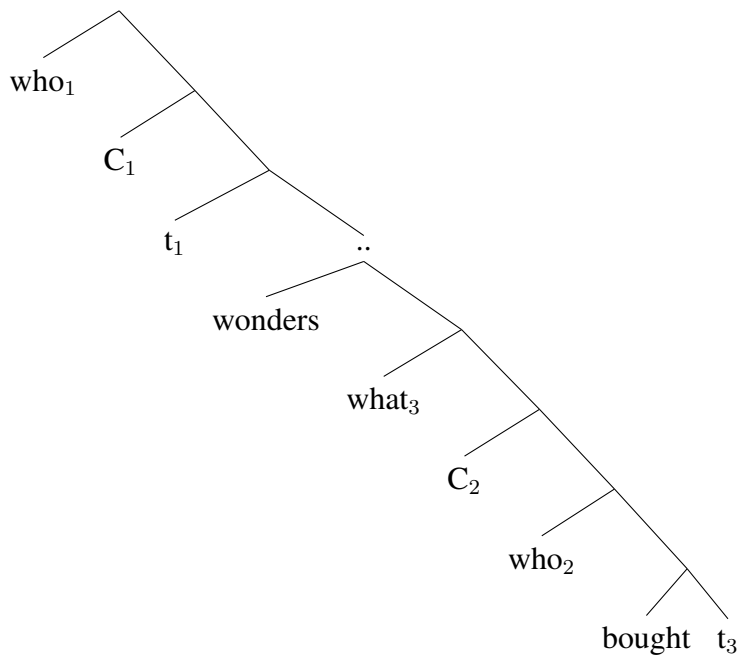
(13) *A simplified derivational structures for (7-a) and (7-b):*



In (13)b, the lower (object) *wh*-phrase is moved over the structurally higher subject *wh*-phrase, yielding a superiority-violating word-order. Under current Minimalist views, this is simply unavailable as a derivational step, as it involves a more distant goal than the closest one to the probe, violating locality constraints. We thus correctly predict that *who bought what?* is grammatical, but *what did who buy?* is ruled out.<sup>4</sup>

The examples in (11) tell us, however, that this cannot be the end of the story, as the raising of a lower *wh*-phrase in a superiority-violating configuration is exceptionally possible just in case there is a higher scopal position at which the in-situ *wh*-phrase can be interpreted. The superiority-violating movement we must assume is precisely the movement argued to be unavailable in (13)b. The necessary structure is illustrated in (14).

(14) *A simplified derivational structure for (11-a); the embedded clause is identical to (13)b:*



<sup>4</sup> Of course, we'll need to say more about German, which does allow superiority violations in sentences akin to (9) and (10).

Notice, crucially, that the possibility of a higher interpretive position for the embedded in-situ *wh*-element only becomes apparent once the higher clause has been constructed. However, allowing later structure building to govern movement in the lower clause would require either a violation of the extension condition or some amount of derivational lookahead.

In the only available account of the phenomenon at hand we are aware of in the literature, Lasnik and Saito (1992) adopt a structure similar to (14). To derive the exceptional superiority-violating reading of examples such as (11-a), Lasnik and Saito introduce the notion of *operator disjointness*, which characterizes the configuration resulting from a superiority-violating movement of X past a c-commanding element Y.

(15) *Operator disjointness* (Lasnik and Saito 1992, p. 120):

A *wh*-phrase X in Spec of CP is operator disjoint from a *wh*-phrase Y if the assignment of the index of X to Y would result in the local  $\bar{A}$ -binding of Y by X (at S-structure).

Lasnik and Saito stipulate that operator disjointness is incompatible with the *absorption* operation that is necessary for the interpretation of a multiple *wh*-question. It can then be seen as a feature indicating non-identity of scope domains.<sup>5</sup>

In a Minimalist top-down account, we must assume that at the point at which the embedded C probes for a *wh*-element, there must be some property of either the C head or the embedded subject *wh*-phrase that allows C to see past this subject *wh*-phrase to the object. One way of accomplishing this is to assume that the higher *wh*-phrase does not bear a relevant [+WH] feature, rendering it invisible to the C probe and allowing the object *wh*-phrase to be the goal of C's probing and thereby undergo movement.

If we pursue this path, though, it is crucial that we carefully regulate the non-appearance of [+WH]. If *wh*-phrases could freely lack this feature, superiority effects would always be avoidable, even in cases like (7). Specifically, the pattern illustrated in (11) indicates that we must connect the possibility of a *wh*-phrase lacking [+WH] to the possibility that this phrase will take scope in a higher domain. This would mean, then, that the scopal positioning of a *wh*-phrase must be reflected as a property of the *wh*-element itself, either at the point of first merge of the *wh*-element or as a result of the derivation prior to the C's probing.<sup>6</sup> The incorporation of such a scopal feature strikes us as antithetical to the idea that properties of scope are determined configurationally as well as to the prohibition on non-lexically determined features (Chomsky's 2000 Inclusiveness Condition).<sup>7</sup>

On a top-down account, on the other hand, at a minimum no look-ahead problem arises. The higher clause, containing a matrix question and a matrix *wh*-element has already been constructed by the time the rogative verb and embedded C are merged into the derivation. The existence of an exceptionally high scope-taking position for the in-situ *wh*-element in the embedded clause is therefore clear before that *wh*-element is merged. We can therefore construct a structure such as (16), with the (phonologically) in-situ *wh*-element taking scope in the matrix (here, we

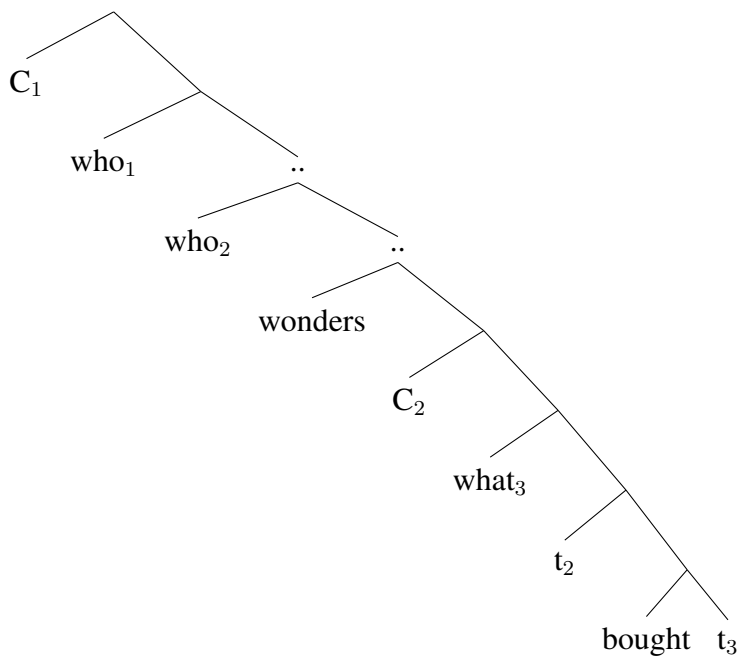
<sup>5</sup> Note that this proposal denies the derivational character of superiority violations, as it assumes that the superiority violating configuration in the embedded clause of (11-a) can be freely generated.

<sup>6</sup> This is in essence Lasnik and Saito's (1992) analysis, as described above.

<sup>7</sup> An alternative strategy that would avoid derivational lookahead could posit a denotation for *wh*-phrases lacking [+WH] that yields a coherent interpretation only in a context with a higher interrogative. Without such a licensing context, then, the derivation will result in semantic incoherence, and the effect of unacceptability. It is, however, not clear to us what such a denotation would look like and how it would differ from that of a [+WH] *wh*-phrase.

assume, through covert *wh*-movement) without the need for additional assumptions.

(16) *A simplified derivational structure for (11-a) with high scope for the in-situ wh:*



This structure is akin to (15) but incorporates a covert movement for the in-situ *wh*. On a top-down view, it can be constructed without lookahead. Notice, further, that on a conception of superiority (yet to be fully spelled out) where the relative ordering of *wh*-elements matters not globally but in relation to the operator that interprets them or assigns them scope, the structure in (16) doesn't violate superiority. Importantly, this is only the case when the (phonologically) in-situ *wh*-element is interpreted in the matrix, below the matrix subject, but not if it's interpreted in the embedded clause, below the fronted *wh*-element, as expected from the available readings of (11) presented in (12). For space reasons, we leave a fully spelled out account to future work.

2.2. RESUMPTIVE PRONOUNS. Another argument for top-down derivations, also discussed by Georgi & Salzmann (2017), stems from the realization of the element at the “tail” of  $\bar{A}$ -movement chains. In English this element (the “trace”) is typically unpronounced. Yet, exactly when the  $\bar{A}$  dependency spans an island boundary, it is realized as a resumptive pronoun. This is illustrated in the following contrast (examples from Ackerman et al. 2018):

(17) *A resumptive pronoun is available and required iff it is inside an island:*

Which woman did Carlos report that . . .

- a. [<sub>island</sub> the newscaster who exposed {<sup>✓</sup>her/\* $\emptyset$ }] threatened the detective's case?
- b. [<sub>island</sub> the newscaster who exposed the criminal] threatened {\*her/<sup>✓</sup> $\emptyset$ }?

Ackerman et al. (2018) shows, using both forced choice and fill-in-the-blank tasks, a reliable preference for resumptive pronouns across a range of island contexts, as well as a reliable preference for unpronounced gaps in non-island contexts. That is, the availability of a resumptive pronoun correlates with the presence of an island.

If we follow Sichel (2014) in assuming that this choice in realization is made during the course of the derivation, the question arises as to when during the derivation the decision can be



made. Under a bottom-up derivational scheme, this clearly cannot be done when the element to be moved is initially merged into its base position. However, neither can this choice be made at the point at which the element is moved out of this position. In the examples in (17), the initial movement of the object *wh*-phrase is typically assumed to target *v*P. At this point in the derivation it is not yet determined whether this movement will cross an island boundary.

One possible solution would be to adopt a more traditional view, under which *wh*-movement proceeds directly to specifier of CP. This would allow us to determine the form of the gap at the right point in the derivation, as the local CP containing the base position in the examples in (17) provides sufficient information to determine the presence or absence of an island. However, examples like the following, modelled after (17) but adding an additional CP embedding, show that even this flexibility is insufficient as a general solution:

(18) *The island edge can be separated from the gap position by multiple embeddings:*

Which woman was the detective surprised . . .

- a. [<sub>island</sub> after the newspaper claimed [<sub>CP<sub>1</sub></sub> that the police had questioned {<sup>✓</sup>her/\* $\emptyset$ }]]?
- b. [<sub>non-island</sub> that the newspaper claimed [<sub>CP<sub>1</sub></sub> that the police had questioned {\*her/<sup>✓</sup> $\emptyset$ }]]?

In both cases, the CP closest to the base position of the moved *wh*-element, namely the one labeled CP<sub>1</sub>, is structurally identical. It will only be on a subsequent movement, which can be delayed arbitrarily in the derivational future through the inclusion of more bridge-verb-headed CPs, that the determination of the gap's form can be made. Once again, then, we find a situation where a derivational operation, namely the choice in form of the tail of the chain of *wh*-movement, requires unbounded lookahead, something that is in conflict with a phase-based view of derivations.

If we instead adopt a top-down approach to derivations, the lookahead problem just sketched evaporates. Prior to the point at which the *wh*-phrase is lowered to its base position, the island context has already been integrated in the structure. Again, we leave open the technical details of how this contextual information is passed through the derivation to inform the realization of the base position. However, the important point is that top-down derivations open up the possibility of formulating such a theory.

**3. Semantic Arguments.** The next set of arguments we will discuss concern the assignment of a semantic interpretation to syntactic elements. Under a phase-based derivational model, the interpretation of an element should be determinable on the basis of structure that is available at the point of spell-out. In this section, we will review a number of cases in which relevant information is provided by material that is higher in the structure. Such information would therefore be unavailable in a bottom-up derivational model, but would be available on a top-down view.

3.1. SEQUENCE OF TENSE. Consider first the interpretation of the embedded past tense in the following English example:

(19) *Sequence-Of-Tense reading of the lower predicate:*

Mary said that John was sick.

- a. *Back-Shifted Reading:* Mary said “John was sick.”
- b. *Simultaneous Reading:* Mary said “John is sick.”

The example is ambiguous between two interpretations: the eventuality denoted by the embedded clause, the state of being sick, may be interpreted as taking place prior to the eventuality of the main clause, the saying, or simultaneously with it. The first of these, the back-shifted reading, is

the interpretation we would expect. With both the matrix and embedded predicates having morphological past tense, the interpretation of both predicates as past will give the embedded clause a “past of the past” interpretation:

- (20) *A back-shifted interpretation involves two independent past tense variables:*  
 $\lambda e.\exists t[t < \text{now} \wedge \text{say}(m, e, t, \lambda e'.\exists t'[t' < t \wedge \text{sick}(j, e', t')])]$

Under the latter interpretation, however, the matrix and embedded eventualities are interpreted with respect to the same time:

- (21) *A simultaneous interpretation involves a single past tense variable:*  
 $\lambda e.\exists t[t < \text{now} \wedge \text{say}(m, e, t, \lambda e'.\text{sick}(j, e', t))]$

The embedded clause, then, apparently does not introduce any tense semantics, and thus the embedded morphological past tense is assigned a semantically vacuous interpretation.

As has been widely discussed (Ogihara 1996; Kratzer 1998, *inter alia*), such a zero interpretation of past tense is possible only when the past tense of a stative predicate is embedded below another past tense clause. When the matrix predicate is either present or future, the embedded past tense morphology must be interpreted as semantically past.

- (22) *SOT readings are not possible with a present or future embedding predicate:*  
 a. Mary says that John was sick.  
 b. Mary will say that John was sick.

Thus, (22-a) cannot mean that Mary is saying that John is currently sick, and (22-b) cannot mean that Mary will say in the future that John is sick at that time.

This sequence of tense phenomenon thus shows an interpretive dependency between lower and higher tense morphology, such that the interpretation of the lower tense is dependent on the higher one. Such a dependency is naturally accommodated in a top-down model. At the point in the derivation at which the the embedded tense is interpreted, regardless of the phasal granularity, the matrix clause will have already been interpreted. Assuming that the outcome of that interpretive process is available to guide subsequent interpretation, the interpretation of the lower past tense morphology can be made sensitive to its embedding context.

In contrast, a bottom-up model struggles to accommodate this kind of interpretive dependency. We adopt the simplest assumption that it is one and the same syntactic object that can receive either the semantic past or zero tense interpretations. We know of no distributional evidence that distinguishes the structures associated with the two interpretations. Neither one, for instance, prohibits or favors the deletion of the embedded complementizer. And extraction out of these complements does not restrict the interpretation of embedded past tense.

- (23) *Extraction does not affect the availability of a SOT interpretation:*  
 a. Mary said that John was angry.  
 b. Why did Mary say that John was angry?  
 c. How angry did Mary say that John was?

If the embedded clause is syntactically identical across both of the inputs, then the interpretive process must be able to determine, at the appropriate derivation point, whether the context licenses a zero tense interpretation for the past tense morphology. In a phase-based derivation, this point will be the point of spell out of the phase including the embedded tense projection, either

the embedded CP or the next higher *v*P phase depending on the theory of phase interpretation. Regardless, neither of these syntactic objects will include the matrix tense, which will therefore be unable to condition the interpretation of the embedded tense.

We know of at least two other approaches to the interpretation of tense in SOT contexts, though neither fares better under bottom-up phase-based derivations. The first approach posits an operation of tense deletion (Ogihara 1996), which deletes the semantic content of the past tense morpheme when it is locally *c*-commanded by another past tense morpheme. It is not clear, however, how this operation is compatible with the assumptions of phase-based derivations. Assuming the embedded clause's tense is interpreted prior to the introduction of the higher T, there will be no way that the conditioning context and context of structural change can be present simultaneously in the derivation. A second approach to SOT posits a lexical ambiguity for the past tense morpheme. Alongside the semantically contentful past tense morpheme, Kratzer (1998) posits the existence of a null tense morpheme which can be realized with past tense morphology so long as it is locally *c*-commanded by a past tense morpheme with which it can agree. Again, the problem with such an approach stems from derivational locality, as the pronunciation of the lower tense head will need to be determined prior to the introduction of the matrix tense into the derivation.

3.2. NEGATIVE QUANTIFIERS. A second argument for top-down derivation comes from the interpretation of neg-words in negative concord (NC) languages. In a subclass of NC languages, called non-strict NC languages, neg-words can exhibit negative concord with another negative element, but need not. In Italian and Spanish, negative words that occur post-verbally must co-occur with a preverbal negative element, such as sentential negation.

(24) *Post-verbal neg-words must be licensed by another negative element:*

- a. \*(Non) ho visto niente.  
not have.1sg seen nothing  
'I didn't see anything.' Italian
- b. \*(No) comió nadie.  
not ate.3sg nothing  
'He/she didn't eat anything.' Spanish

In contrast, pre-verbal neg-words do not require a co-occurring negative element and are typically interpreted with their own negative force.

(25) *Pre-verbal neg-words can carry independent negative force and do not require a licenser:*

- a. Nessuno ha chiamato.  
nobody has.3sg called  
'Nobody called.' Italian
- b. Nadie vino.  
Nobody came.3sg  
'Nobody came.' Spanish

As in the case of past tense, we will assume there is only a single lexical element corresponding to a neg-word, whether it occurs pre- or post-verbally, and whether it contributes negative force to the sentence or not.<sup>8</sup> We will also need to make one of two additional assumptions:

<sup>8</sup> One could alternatively assume that neg-words are lexically ambiguous between a negative quantifier and an (NPI)

either there is some interpretive process that directly interprets a neg-word as a negative quantifier or indefinite on the basis of its context (Rizzi 1982), or there is a syntactic process that inserts or deletes a feature that determines the interpretation (Longobardi 1987; Zanuttini 1991).

Unlike the SOT case, in this case the relevant contextual information appears to be locally available. For either bottom-up or top-down derivations, this can be done on the basis of the category of the element with which the neg-word is merged. When merged into a post-verbal position (i.e., directly with something of category V, say), a neg-word will be interpreted without inherent negative force.<sup>9</sup> In contrast, a neg-word that is merged pre-verbally, say via merger with T, will be interpreted with negative force.

While this approach captures much of the pattern, there exist a number of examples in which the structural position of the neg-word does not uniquely determine its interpretation. For instance, compare the Spanish examples in (26):

- (26) *Exceptional negative concord reading of embedded subject (Herburger 1999; p.102):*
- a. Es imposible [que lo sepa *nadie*].  
is impossible that it knows.3sg nobody  
'It is impossible that anyone knows it.' (unambiguous)
  - b. Es imposible [que *nadie* lo sepa].  
is impossible that nobody it knows.3sg  
'It is impossible that anyone/nobody knows it.' (ambiguous)

As Herburger (1999) notes, the first of these examples, with a post-verbal neg-word in an embedded subjunctive clause, must interpret the neg-word as an indefinite without negative force, with the licensing negation provided by the negative matrix predicate *imposible*. In contrast, when the embedded neg-word appears in pre-verbal position, the sentence is ambiguous: the neg-word may be interpreted either without negative force, synonymous with the sentence with the post-verbal neg-word, or with negative force, in which case the interpretation is essentially synonymous with the sentence *Someone must know it*. Importantly, without the higher negative predicate, the pre-verbal neg-word only admits a single interpretation, namely the one with negative force:

---

indefinite (Herburger 2001). Under such a view, one must posit some kind of syntactic filter or interpretive principle that serves to constrain the distribution of the two types of elements. Herburger posits a syntactic filter, which will be subject to the same argument in favor of top-down derivations discussed in the main text. As for an interpretive principle, it is imaginable that the distribution of the indefinite might be constrained in a manner similar to that used by Kadmon & Landman (1993) to rule out NPIs in unlicensed positions. However, it is not at all clear how the distribution of negative quantifiers could be constrained so that post-verbal neg-words in examples like (24) could not be interpreted as negative quantifiers.

<sup>9</sup> In fact, for bottom-up derivations this determination needs to be done in a somewhat more subtle fashion, as the base position in which a neg-word is merged does not completely determine interpretation. A neg-word that is fronted, whether via passivization or focalization, is interpreted as negative, and does not require a licensing negation, as shown in (i)–(ii) below from Italian.

- (i) Niente è stato detto.  
nothing is.3sg been said  
'Nothing was said.'
- (ii) Niente, ho detto.  
nothing have.1sg said  
'Nothing, I said.'

Nonetheless, since neg-words that undergo movement will bear features to trigger it, we can restrict the interpretation without negative force to only those neg-words merged with V without such movement-triggering features.

- (27) Es posible [que nadie lo sepa].  
 is possible that nobody it knows  
 It is possible that nobody knows it.’

According to the description in (Rizzi 1982; p.126), the situation in Italian is different: pre-verbal neg-words admit only the interpretation with negative force, while post-verbal neg-words must be interpreted as indefinites in the scope of licensing negation.

- (28) a. Piero non crede che mi possa spaventare niente.  
 Piero not thinks.3sg that me can.Subj3sg scare niente  
 ‘Piero doesn’t think that anything can scare me.’  
 b. Piero non crede che niente mi possa spaventare.  
 Piero not thinks.3sg that nothing me can.Subj3sg scare  
 ‘Piero doesn’t think that nothing can scare me.’

However, this appears to be a simplification of the pattern. While pre-verbal neg-words in such a context may prefer a negative interpretation, and such an interpretation is impossible with post-verbal neg-words, the indefinite interpretation appears to be available, judging from examples like the following and others found in the literature:

- (29) Non credo che nessuno lo sapesse  
 not think.1sg that nobody it knew.Sbj3sg  
 ‘I think that anybody knew it.’ (Maiden & Robustelli 2014; p.155)

From the perspective of top-down derivations these Spanish and Italian data patterns pose no problem: at the point in the derivation at which the neg-word is interpreted, any higher context will have been constructed. If it contains a negation that is sufficiently local, the neg-word, whether in pre- or post-verbal position will be interpretable as an indefinite. If it does not, pre-verbal neg-words will be interpreted as negative quantifiers and post-verbal neg-words will not be assigned a coherent interpretation.

For bottom-up derivations on the other hand, the situation is more problematic. In the examples in (28) and (29), the embedded subject can be interpreted as an indefinite only in the presence of a higher negation. This negation is not present in either the embedded CP or the next higher  $\nu$ P phase, the candidates for the times at which the embedded TP would be interpreted. Consequently, to allow a derivational system incorporating phasal interpretation to work in such a case, we will need to allow for further delay. But if such delay were to be generalized from semantic interpretation to phono-/morphological spell-out, it would yield unfortunate predictions for patterns of allomorphy (Marantz 2007; Embick 2010).

**4. Discussion.** This short paper presented two syntactic arguments and two semantic arguments which favor top-down derivations. These arguments share a logical structure: information in a higher domain conditions operations in a lower domain. On a top-down approach, such information is readily available, but on a bottom-up approach, additional assumptions must be made.

Several similar semantic arguments support this conclusion, including the interpretation of *any* as an NPI vs. FCI, the interpretation of *wh*-elements as relative pronouns vs. questions, and the Predicate Abstraction rule.<sup>10</sup> Additionally, arguments in favor of *Upward Agree*, where mor-

<sup>10</sup> As Heim & Kratzer (1998; §5.2.4) put it, on a bottom-up construal, the Predicate Abstraction rule requires “a certain amount of foresight to write up” (emphasis in the source).

phological marking on a syntactic element is driven by properties of a higher element (Zeijlstra 2012; Wurmbrand 2012, 2014; Bjorkman & Zeijlstra 2019) also naturally fit within a top-down approach. Well-known examples of such cases include Arabic subject agreement (Aoun et al. 1994) and Romance past participle agreement (Kayne 1989) (see also Longenbaugh 2019).

Nonetheless, potentially problematic cases remain, where structurally lower syntactic context conditions the form of higher elements. Examples include overt case marking on a fronted *wh*-phrase, and long-distance agreement (Polinsky & Potsdam 2001; Bhatt 2005). We leave the analysis of such cases in a top-down framework for future work.

Our goal in this paper has been to show that the question of derivational order is not simply conceptual — it drives how we structure our research questions and define the space of possible solutions. Turning the logic of structure building on its head would undoubtedly lead to new research efforts. For example, what unites the few cases where a lower domain appears to modulate properties of a higher one? Can a top-down view shed light on the lack of SOT phenomena in head-final languages like Japanese, and on the lack of relative/question ambiguities in their use of *wh*-elements? Moreover, the vast majority of the arguments we have presented here could also be used to favor a Left-to-Right (linear) perspective: what kind of evidence could tease these two apart? Beyond these theoretical considerations, top-down approaches to syntax can more straightforwardly interface with theories of sentence processing — something that has eluded bottom-up approaches.

## References

- Ackerman, Lauren, Michael Frazier & Masaya Yoshida. 2018. Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry* 49(4). 847–859.
- Aoun, Joseph, Elabbas Benmamoun & Dominique Sportiche. 1994. Agreement, word order, and conjunction in some varieties of Arabic. *Linguistic Inquiry* 25(2). 195–220.
- Bhatt, Rajesh. 2005. Long distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory* 23(4). 757–807.
- Bjorkman, Bronwyn & Hedde Zeijlstra. 2019. Checking up on ( $\phi$ -)Agree. *Linguistic Inquiry* 50(3). 527–569.
- Chomsky, Noam. 1955. The logical structure of linguistic theory. Draft, Harvard University.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton publishers.
- Chomsky, Noam. 1965. *Aspects of the theory of grammar*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, New York: Holt, Reinhart, and Winston.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. MIT Press.
- Embick, David. 2010. *Localism versus globalism in morphology and phonology*. MIT Press.
- Fillmore, Charles J. 1963. The position of embedding transformations in a grammar. *Word* 19. 208–231.
- Georgi, Doreen & Martin Salzmann. 2017. The matching effect in resumption: A local analysis-based on case attraction and top-down derivation. *Natural Language & Linguistic Theory* 35. 61–98.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Blackwell.
- Herburger, Elena. 1999. On the interpretation of Spanish *n*-words. In *Semantic issues in romance syntax*, John Benjamins.

- Herburger, Elena. 2001. The negative concord puzzle revisited. *Natural Language Semantics* 9(3). 289–333.
- Huang, C. T. James. 1982. *Logical relations in Chinese and the theory of grammar*: Massachusetts Institute of Technology dissertation.
- Kadmon, Nirit & Fred Landman. 1993. Any. *Linguistics and Philosophy* 353–422.
- Kayne, Richard. 1989. Facets of Romance past participle agreement. In Paola Benincà (ed.), *Dialect variation and the theory of grammar*, Foris.
- Kotek, Hadas. 2019. *Composing questions*. Cambridge, MA: MIT Press.
- Kratzer, Angelika. 1998. More structural analogies between pronouns and tenses. In *Proceedings of SALT 8*, Ithaca, NY: CLC Publications, Cornell University.
- Kuno, Susumu & Jane Robinson. 1972. Multiple *wh* questions. *Linguistic Inquiry* 3.
- Lasnik, Howard & Mamoru Saito. 1992. *Move  $\alpha$ , conditions on its application and output*. Cambridge, MA: MIT Press.
- Longenbaugh, Nicholas. 2019. *On expletives and the agreement-movement correlation*: Massachusetts Institute of Technology dissertation.
- Longobardi, Giuseppe. 1987. Parameters of negation in Romance dialects. In *GLOW dialectology workshop, venice, italy*, .
- Maiden, Martin & Cecilia Robustelli. 2014. *A reference grammar of modern Italian*. Routledge.
- Marantz, Alec. 2007. Phases and words. In S.H. Choe (ed.), *Phases in the theory of grammar*, 191–222. Dong-In Publishing Co.
- Ogihara, Toshiyuki. 1996. *Tenses, attitudes, and scope*. Dordrecht: Kluwer.
- Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press.
- Phillips, Colin. 1996. *Order and structure*. Cambridge, MA: MIT dissertation.
- Phillips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34(1). 37–90.
- Polinsky, Maria & Eric Potsdam. 2001. Long-distance agreement and topic in Tsez. *Natural Language and Linguistic Theory* 19. 583–646.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.
- Shieber, Stuart M. & Mark Johnson. 1993. Variations on incremental interpretation. *Journal of Psycholinguistic Research* 22(2). 287–318.
- Sichel, Ivy. 2014. Resumptive pronouns and competition. *Linguistic Inquiry* 45(4). 655–693.
- Stabler, Edward P. 1991. Avoid the pedestrian’s paradox, 199–237. Dordrecht: Kluwer.
- Steedman, Mark. 1989. Grammar, interpretation and processing from the lexicon, 463–504. Cambridge, MA: MIT Press.
- Steedman, Mark. 2000. *The syntactic process*. Cambridge, MA: MIT Press.
- Wurmbrand, Susi. 2012. The syntax of valuation in auxiliary-participle constructions. In Jaehoon Choi, Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz & Alex Trueman (eds.), *Proceedings of WCCFL 29*, 154–162. Somerville, MA: Cascadilla Press.
- Wurmbrand, Susi. 2014. The Merge Condition: A syntactic approach to selection. In Peter Kosta, Steven Franks, Teodora Radeva-Bork & Lilla Schurcks (eds.), *Minimalism and beyond: Radicalizing the interfaces*, 130–166. Amsterdam: John Benjamins.
- Zanutini, Raffaella. 1991. *Syntactic properties of sentential negation, a comparative study of Romance languages*. Philadelphia, PA: University of Pennsylvania dissertation.
- Zeijlstra, Hedde. 2012. There is only one way to agree. *The Linguistic Review* 29. 491–539.