Generalized Earliness

Norvin Richards, MIT

norvin@mit.edu

Current Minimalist approaches to syntax typically assume at least two kinds of conditions that regulate the order in which operations occur. One is some version of cyclicity, which could be captured by principles like those in (1):

(1) a. Trees are constructed from the bottom up (External Merge is to the root of the tree under construction)

b. Operations like Agree and Move must take place, if they are to take place at all, as soon as everything necessary for the operation is in the workspace (cf. Pesetsky 1989, Chomsky 1995).

Another widely assumed condition segregates operations by their linguistic subfield:

(2) Within a spellout domain, all syntactic operations must precede all phonological and semantic operations.

The condition in (2) is enforced, both in the classic 'T model' and in the more recent multiple spellout model, by positing a derivation in which the 'narrow syntax', which performs only syntactic computations, creates representations which are sent to the LF and PF interfaces, where semantic and phonological operations are performed. This architecture is meant to capture a commonly assumed principle about the nature of the interaction between syntax and the interfaces, namely that syntactic operations are blind to semantic and phonological information (Zwicky and Pullum 1986). This principle has been challenged in a number of works, for example in the literature on Contiguity Theory (Richards 2010, 2016, 2019, Branan 2018, Yamada 2019, Archibald and Croteau 2020, Stepanov and Moussaoui 2020).

I propose that we replace (1-2) with (3), in which (2) is discarded and (1b) has been generalized to hold of every kind of operation, syntactic or not:

(3) a. Trees are constructed from the bottom up (External Merge is to the root of the tree under construction)

b. Operations ~~like Agree and Move~~ must take place, if they are to take place at all, as soon as everything necessary for the operation is in the workspace (cf. Pesetsky 1989, Chomsky 1995).

I will refer to the principle in (3b) as *Generalized Earliness*. In Generalized Earliness, there is no separation of operations into the 'narrowly syntactic', the semantic, and the phonological; all kinds of operations are in principle interspersed, and the only condition determining their ordering is a general version of Pesetsky's (1989) Earliness, which requires every operation to occur as soon as the conditions for it to occur have been met.

A natural question, then, is: how we are to derive the effects of (2)? To the extent that syntax does tend to behave in a way which is oblivious to phonology and semantics, how can we capture this state of affairs? I will try to show that (3) can capture this tendency, while still leaving room for some exceptions of a kind which I believe are desirable.

1. Generalized Earliness and lexical insertion

Suppose we consider a moment in the derivation of an English clause, when the next element to be Merged is an instance of T:

(4) T *v*P
…

One operation which will take place at some point is Merge of T with *v*P. Another is Lexical Insertion of various kinds of information in T.

The special facts about a particular instance of English T that would distinguish it from other syntactic nodes, apart from its label, might include:

(5) a. T bears a ϕ-probe

b. T is pronounced /d/

c. T is a suffix

d. T means… (appropriate semantic representation)

Each of the facts in (5) must at some point be introduced into the representation. Conventionally, we posit an operation of Lexical Insertion which inserts at least the facts in (5b-d); in Distributed Morphology, for example, this is a single operation which occurs once the 'narrow syntax' is finished and the representation is about to be handed off to the interfaces.

Instead of positing a unified operation of lexical insertion, let us imagine that each of the facts in (5a-d) is subject to its own process of insertion. Moreover, each instance of insertion is governed by the Earliness condition in (3b); it must take place as soon as the workspace contains all of the structure to which insertion is sensitive. We can then ask: how soon can each of the facts in (5a-d) be inserted?

Consider first the fact in (5a): T bears a ϕ-probe. This is arguably a fact about every instance of T in English, if we follow Marantz (1991), and much subsequent work, in assuming that the special properties of infinitival clauses do not have to do with failure of Case-assignment by T. As soon as T is drawn from the Numeration or the lexicon, then, the computation has as much structure as it needs to insert the fact in (5a); T can be Merged with its ϕ-probe already inserted.

Next let us turn to the fact in (5b): T is associated with certain phonological segments (in English, these include /d/ and /z/, for example). How soon can a property like this be inserted into the structure?

In some kinds of cases, it is at least clear that insertion of the segments associated with T must take place after T's ϕ-probe is inserted and Agrees with a DP. Consider the sentences in (6):

(6) a. She like**-s** Syntax.

b. I like-**Ø** Syntax.

The segments associated with T cannot be determined until after T and the subject have Agreed. At the very earliest, then, T must be given its phonological segments after the Agree operation.

In fact, I think we have reasons to believe that T's phonological segments must be inserted even later than this, at least in some cases. Consider pairs of sentences like the one in (7):

(7) a. It broke**-Ø**.

b. Di**-d** it break?

The facts in (7) demonstrate that the segmental content of T is not simply sensitive to T itself, or even to T along with the verb to which T is attached; in order to know the pronunciation of T, we must first Merge C in order to find out whether C is triggering *do*-support.

In other words, while the only structure necessary for insertion of the fact in (5a) ("T bears a ϕ-probe") might be (8a), the structure necessary for insertion of the fact in (5b) ("T is pronounced /d/") must be at least (8b):

(8) a. T

b. $[_{CP}$ C… …T… …V…]

We have arrived at a desirable result. The fact that T bears a ϕ-probe can be established immediately, at or even before the point in the derivation at which T is Merged into the larger syntactic structure. The particular segments that T will contribute to the pronunciation of the verb, on the other hand, cannot be introduced until later, certainly after the ϕ-probe on T has Agreed with a DP, and in fact, after TP has been completed and C has been Merged.

Because of Earliness, as soon as T and its ϕ-probe have been Merged into the larger syntactic structure, the ϕ-probe will be compelled to Probe for a Goal, and trigger whatever operations take place as a result of the Probe operation, possibly including movement of a DP to the specifier of TP. All of this would have to happen, in this kind of derivation, before T has any phonological segments attached to it—not because we have stipulated that lexical insertion is late, but because every operation must take place as early as it can, and we can determine on independent grounds that insertion of T's phonological segments must be late. On this view, there is no "narrow syntax" per se: "syntax" is simply the study of the operations which can be triggered by the earliest features to appear on a head.

Many new possibilities now arise. One of these is that there could be properties which we think of as belonging to phonology or morphology, but which can actually be inserted early, and we would then hope to find evidence that syntactic operations can be sensitive just to these properties. Consider, for example, the fact in (5c) ("T is a suffix"). The suffixal property of T, unlike the phonological segments associated with it, seems to be a reliable property of any head with the label T in English (assuming, following Abusch 1985 and Copley 2002, that expressions like *will* are modals rather than Tense morphemes). We would then expect the fact that T is a suffix to be established early, along with the fact that T bears a ϕ-probe. And in fact, Richards

(2016) argues that suffixhood is one of the phonological properties to which syntax can be sensitive, while segmental information is not.

The next sections will briefly discuss some other issues connected with the proposal. Section 1 will be about phonological features, and the timing of their introduction, and section 2 will consider semantic features. In section 3, I will consider another kind of syntactic probe, the one responsible for wh-movement. Section 4 will suggest a further simplification of the theory, proposing a way of connecting the condition in (3a) above (requiring that External Merge be to the root of the current tree) to the condition that requirements be satisfied as quickly as possible. Section 5 will conclude.

1. When are phonological features introduced?

The conditions on insertion of T's phonological segments can be thought of as an instance of outward-sensitive allomorphy (Bobaljik 2000, Embick 2010, Merchant 2015, Moskal 2015, Deal and Wolf 2017, Choi and Harley 2019, and much other work); the form of T is determined, in part, by material which is Merged higher than T. Stem suppletion represents one of the most widely attested forms of outward-sensitive allomorphy:

(9) a. broke-**Ø**

b. break-**s**

The English verb *break* shows allomorphy conditioned by the Tense affixes attached to it; it has one form if it combines with the past tense morpheme, and another if it does not.

I think it is fair to say that there is no general agreement, in the lively literature on outward sensitive allomorphy, on the precise size of the domain to which outward sensitive allomorphy can be sensitive[1]. What is clear, however, is that the phenomenon exists; there are

[1] One possibility is that this domain is what is traditionally known as the 'phase'. Whether that idea can help us account for the various phenomena that are standardly attributed to phases is a question I will leave for future work.

cases in which the segments appearing under a particular syntactic node cannot be determined until more syntactic structure has been built[2].

In the previous section, I proposed an account of why syntax seems to be 'blind' to phonological features, using the allomorphy of English T as an example; the idea was that because the segments making up T cannot be determined at least until C has been Merged, they are introduced too late to affect the behavior of the ϕ-probe on T. In other words, the environment in which phonological features are inserted is not merely T, but some larger syntactic domain.

In some cases, of course, C will actually have no effect on T; if C is declarative, for example, then T will have the same segments in it that it would have had before C is Merged. The idea could be that lexical insertion is universally required to wait for the construction of the largest domain in which allomorphy could conceivably be conditioned. English T is required to wait until C has been Merged before its segments can be inserted; sometimes this wait will turn out to have been needless, but the domain for insertion of segments is at least as large as CP, as the allomorphy facts show. If we are willing to think of insertion of phonological features as *invariably* applying in syntactic domains that are larger than the single head to which features are

---

[2] David Pesetsky (p.c.) points out to me another striking case of a similar kind, discussed by Andrews (1982), Frampton and Gutmann (2000) and Chomsky (2001), involving Icelandic data like those in (i):

(i)	a.	Hún	var	handtekin.				
		she.NOM	was	arrested.NOM				
		'She was arrested'						
	b.	Þeir	segja	hana	hafa	verið	handtekna.	
		they	say	her.ACC	to.have	been	arrested.ACC	
		'They say that she has been arrested'						

In (i) the participle meaning 'arrested' agrees in Case with the subject of its clause—but the value of this Case is determined, in (ib), only after the introduction of the matrix ECM verb *segja* 'say' and whatever other structure of the main clause is relevant for making the embedded subject Accusative. The phonological form of *handtekna* 'arrested.ACC' therefore cannot be settled on until long after the maximal projection of the participle has been completed (in fact, the syntactic structure intervening between the participle and the nominal from which it receives its case value can be arbitrarily large, since raising infinitives can intervene between them).

being added, then the account proposed in the last section can become a general account of why syntax seems to ignore phonological segments. On this account, insertion of phonological segments under a given head X would very generally not take place until some larger domain α has been constructed:

(10) [$_{\alpha}$ X]

On this view, even in some language in which T never bears any allomorphy for anything in C, insertion of the segments of T invariably waits just as long as it does in English. Insertion of segments, on this account, simply has a domain of application which is larger than the head to which the segments are to be attached, and insertion never takes place before this domain has been constructed.

2. When are semantic features introduced?

The preceding discussion has focused exclusively on phonological features. When should semantic properties be introduced?

The semantic properties of a node, like its phonological properties, are sometimes determined partly by its syntactic environment (see, for example, Preminger 2022 for recent discussion). An N like *bucket*, for example, will have different semantic properties in the two sentences in (11):

(11) a. John kicked the bucket [="John died"].

b. There's a hole in the bucket.

Just as we cannot give English T its phonological segments until after C has been Merged, then, we cannot fully specify the semantics of *bucket* at least until we know whether it is the object of the V *kick*. Any semantic properties that distinguish *bucket* in (11a) from *bucket* in (11b) should be absent when the N is first introduced into the structure. As with allomorphy, the literature on idioms is large and lively, and I think it is again fair to say that there is no generally held consensus on how large an idiom may be (see Nunberg et al 1994, O'Grady 1998, Bruening 2020, Preminger 2022, and other work on the syntax of idioms). As long as we can follow the same reasoning outlined for phonology in the last section, however, the general result seems to be safe; insertion of semantic features, for the most part, applies to structures larger than a single head, and as a consequence, when a new head is Merged, any syntactic operations that it triggers will be unaffected by its semantics, just as they are unaffected, for the most part, by its phonology.

In the case of phonology, I claimed above that while most phonological information may not be present when a head is first Merged and begins triggering syntactic operations, there are some kinds of phonological information which may safely be inserted 'early'. The example given above was the fact that T is a suffix, and I have argued, in Richards 2016, that we do want syntax to be able to make reference to this kind of fact. A logical question, on this view, is whether there are any aspects of semantic representation which are similarly 'safe' to insert early; we would then hope to find evidence that syntactic operations can make reference to some limited set of semantic properties. On the view outlined here, we would be looking for properties which are reliably true of any head with a given syntactic label in a given language (just as, in English, any morpheme with the label 'T' will be a suffix). If there are semantic properties which idioms generally share with their non-idiomatic readings (see Marantz 1997, McGinnis 2002 for discussion), then these would be of interest as well, as candidates for properties which can be introduced into a lexical item before a larger syntactic structure has been constructed. Whether such properties exist is a question I will have to leave to future work.

3. Wh-features

In sections 1 and 2 of this paper we considered the behavior of T, and arrived at the conclusion that T's ϕ-probe must be active at a point in the derivation at which T has no segmental phonological features—not because we stipulate that phonological features are inserted late, but because all properties of a head are to be inserted as soon as all of the information necessary to insert them is present in the representation, and this point in the derivation comes earlier for T's ϕ-probe than it does for T's phonological segments. I claimed that this would be a general result about the features typically taken to drive syntactic operations; these are some of the earliest features to appear on a head, and their behavior is therefore typically blind to many of the other properties (for example, phonological properties) of the material around them. Suppose we try applying this reasoning to the case of wh-features.

Consider, for example, the pair in (12):

(12) a. I don't know [**what** they think [we should eat __]].

b. I don't know [**who** thinks [we should eat **what**]].

On certain kinds of assumptions, (12a-b) raise a puzzling question. The wh-phrase *what* moves overtly in (12a), and covertly in (12b), and the relevant difference between the examples has to do with the presence or absence of another wh-phrase higher in the structure (*who*). Moreover, the two wh-phrases can be arbitrarily structurally distant from each other; well-formed pairs like those in (12) can involve any number of embedded clauses dominating the lower but not the higher wh-phrase, as long as all the intervening material is transparent to wh-extraction. And we have excellent evidence that wh-movement is successive-cyclic; the first step of movement of *what* in (12a) is not to the landing site where it is pronounced, but to some lower position (perhaps the edge of the lowest *v*P dominating it). This first step will always have to take place, on conventional assumptions, before we know whether movement of *what* is to be overt or covert.

Related issues arise for the pair in (13):

(13) a. **What** do they think [we should eat __]?

b. I think [we should eat spaghetti].

Wh-movement in (13a) is triggered, ultimately, by an interrogative wh-feature on matrix C. But we think that wh-movement proceeds successive-cyclically. Are there features on embedded C, and on *v*, which trigger these successive-cyclic steps of movement in (13a)? And if there are, what are these features doing in (13b)?

There are proposals already in the literature which offer answers to these kinds of questions. In (13), for example, we might suppose that the features triggering successive-cyclic movement are always present, but (following Preminger 2014) that failure to check these features does not result in ungrammaticality. Or we might adopt an approach that overgenerates and filters; the successive-cyclic-movement-triggering features are optionally present, but their absence in (13a), or their presence in (13b), leads to ungrammaticality. Similarly, in (12), we could declare that overt wh-movement of *what* is always optional, with bad choices filtered out. Alternatively, we could deny that the choice between overt and covert movement must be made at an early stage of the derivation; on this view, we might say that *what* always undergoes the same kind of movement, that the distinction between overt and covert movement is a choice of which copy in the resulting chain to pronounce, and that this choice can be made quite late in the derivation.

Fox (2022) proposes a new answer to these questions, which (as Fox points out) also solves a problem raised by Davis (2020) having to do with the licensing of parasitic gaps in structures involving intersecting A-bar paths. He develops his proposal assuming a version of cyclicity based on the Extension Condition, but I think his ideas are especially congenial to the Generalized Earliness approach. Fox' proposal incorporates the claim, defended in Rackowski and Richards (2005) and van Urk and Richards (2015), that for a wh-phrase to be extracted out of a CP into a higher clause, the *v* of the higher clause must Agree with the CP from which extraction takes place.

For Rackowski and Richards (2005) and van Urk and Richards (2015), this requirement was to follow from conditions on locality; *v* seeks a phrase dominating a wh-feature, and if the wh-phrase is contained in an embedded CP, the CP is a closer target than the wh-phrase itself is (either because, by virtue of dominating the wh-phrase, the CP dominates a wh-feature, or because the C head bears a wh-feature of its own, responsible for triggering successive-cyclic movement). Their idea was that *v* must therefore Agree first with the closest relevant phrase—namely, the CP—after which it can Agree second with the wh-phrase to be extracted from the CP. They offer evidence from the verbal morphology associated with wh-movement in Tagalog and in Dinka, respectively, in support of this claim.

Fox (2022) suggests that this Agree relation has another consequence: it is the Agree relation with a higher *v* which gives declarative C the features necessary to trigger wh-movement at all. We might imagine, for example, that *v* and declarative C invariably enter the derivation with an unvalued wh-feature, but that what values a wh-feature, and gives it the ability to Agree and trigger wh-movement, is the interaction with interrogative C[3].

We can illustrate this idea with an example like the one in (14):

[3] Pursuing this idea further, we could entertain the possibility that C becomes interrogative because of its interaction with other heads (with a verb that selects for an interrogative declarative clause, for example). I will put this possible elaboration aside in what follows.

(14) **What** C_1 do they v_1 think [C_2 we should v_2 eat]?

The derivation of (14) begins by building the entire clause, without yet performing any wh-movement at all:

(15) $C_{1\,[wh]}$ they v_1 think [C_2 we should v_2 eat **what**]

The matrix complementizer, marked C_1, is interrogative, and bears a [+wh] feature. Following Fox' (2022) idea, the wh-feature on C_1 Agrees with its closest possible target, which is the inert wh-feature on v_1. As a result of this Agree relation, v_1 acquires a syntactically active wh-feature of its own. This feature on v_1 then Agrees with the embedded C_2, giving C_2 a feature which can Agree with v_2:

(16) $C_{1\,[wh]}$ they $v_{1\,[wh]}$ think [$C_{2\,[wh]}$ we should $v_{2\,[wh]}$ eat **what**]

At this point in the derivation, there is only one wh-feature which has *what* as its closest possible target for Agree, which is the feature on v_2. This feature therefore Agrees with *what*, triggering movement to a specifier of v_2:

(17) $C_{1\,[wh]}$ they $v_{1\,[wh]}$ think [$C_{2\,[wh]}$ we should [**what**] $v_{2\,[wh]}$ eat ____]

The move in (17) makes *what* accessible to C_2, which triggers another step of successive-cyclic movement:

(18) $C_{1\,[wh]}$ they $v_{1\,[wh]}$ think [[**what**] $C_{2\,[wh]}$ we should ____ $v_{2\,[wh]}$ eat ____]

Successive-cyclic movement continues in this way until the wh-phrase reaches the specifier of the matrix C:

(19) [**what**] $C_{1\,[wh]}$ they ____ $v_{1\,[wh]}$ think [____ $C_{2\,[wh]}$ we should ____ $v_{2\,[wh]}$ eat ____]

All of these operations are legitimate under the approach to cyclicity being developed in this paper, which requires only that operations take place as early in the derivation as possible. In our version of Fox' hypothesis, the only wh-features capable of probing are those on interrogative C, and those which have been the goal of probing by an active wh-feature. In the derivation just discussed, the matrix C introduces the first interrogative wh-feature, when C is Merged, and this feature Probes immediately, finding the closest Goal, which is located on the matrix *v*. The wh-feature on the matrix *v*, in turn, only becomes capable of probing after having been a Goal of the active wh-feature on the matrix complementizer, and *v*'s wh-feature then Probes, as early in the derivation as it can. The movement operations in (17-19) are also triggered as early in the derivation as possible; their landing sites are, in some cases, quite deeply embedded in the tree, but all of them are landing in positions dictated by the properties of the features which triggered them (that is, they are performing the shortest possible movement operations that will allow them to c-command the feature which is triggering movement).

I think the behavior of these features is also compatible with another general goal of this paper, which is to guarantee that syntactic features will, for the most part, be blind to phonological properties. English complementizers can have different forms depending on what kinds of Agree relations they have participated in. Interrogative complementizers, for example, can be pronounced if they have no overt specifier, but must be null if their specifier is overt:

(20) a. I wonder [**if** it's raining]

b. I wonder [**what** (***if**) she's reading]

In other words, the segments that appear in English C can only be settled on once C has performed all of the Agree operations that it will participate in. Consequently, by the logic being followed in this paper, the Agree operations will have to take place before the segments are introduced, and will thus be blind to any properties of segments.

As Fox points out, his approach has the potential to answer a question raised by Davis (2020). Davis notes (echoing an observation made by Pesetsky (1982)) that when wh-paths intersect, parasitic gaps may be licensed by the wh-phrase which begins in the lower position (and lands in the higher position), but not by the wh-phrase that begins in the higher position (and lands in the lower position):

(21) a. ? This Volvo is one car **OP**$_2$ that I know **who**$_1$ to persuade [owners of __ $_2$]

to talk to __$_1$ about __$_2$

b. * This Volvo is one car **OP**$_2$ that I know **who**$_1$ to persuade [friends of __ $_1$]

to talk to __$_1$ about __$_2$

The examples in (21) are adapted from Pesetsky (1982, 426-7), and make use of the technique (used both by Pesetsky and by Davis) of making examples with intersecting paths easier to parse by making the movement paths of different kinds; here, the intersecting paths involve relativization and wh-movement. If we forgo this help to our parsing ability by making both of the paths wh-paths, we arrive at something like (22):

(22) a. ? **What**$_2$ do you know **who**$_1$ to persuade [owners of __ $_2$]

to talk to __$_1$ about __$_2$

b. * **What**$_2$ do you know **who**$_1$ to persuade [friends of __ $_1$]

to talk to __$_1$ about __$_2$

Here I think the basic judgment is still the same, though the contrast is certainly more subtle. Strikingly, however, if we convert the example to one involving only a single landing site for the multiple wh-phrases, the facts change; it is only the wh-phrase which begins the derivation higher (the one which moves overtly, in English) which can license the parasitic gap:

(23) a. ***Who**$_1$ do you want to persuade [owners of __ $_2$]

to talk to __$_1$ about **what**$_2$

b. ?**Who**$_1$ do you want to persuade [friends of __ $_1$]

to talk to __$_1$ about **what**$_2$

As Davis (2020) points out, the contrast in (22-23) raises problems for a number of standard assumptions about how the derivation proceeds. At the point in the derivation at which the phrase containing the parasitic gap is introduced, neither of the two eventual landing sites for the wh-phrases is present in the tree. Nevertheless, the choice of licenser of the parasitic gap is determined by the future behavior of the two wh-phrases; whichever wh-phrase will be pronounced in the specifier of the matrix CP must be the licenser of the parasitic gap. Davis (2020) points out that on Nissenbaum's (2000) approach to parasitic gaps, the data in (22-23) must be understood as meaning that in the *v*P in which the parasitic gap is introduced (that is, before either landing site for the wh-phrases is Merged), *what* is in a higher specifier than *who* in (22), but the reverse is true in (23). How are we to guarantee this result?

Fox' approach gives us an answer. In (23), only a single interrogative C is ever Merged, in matrix position, and once this C has been Merged, it triggers successive-cyclic movement of both wh-phrases, in a way that obeys "tucking in"; consequently, the wh-phrase which begins higher in the structure, *who*, remains higher than *what* throughout the derivation, and is correctly chosen as the parasitic gap licenser.

In (22), on the other hand, there are two instances of interrogative C, each contributing its own wh-feature, and each of these features will make its way down to the most deeply embedded *v*. As long as movement by distinct features triggers nested paths (and see McGinnis 1998 and Rackowski 2002 for arguments to this effect), the parasitic gap facts ought to follow. The information about how many final landing sites of wh-movement there are will be transferred down to the lowest point in the clause, which is what Davis' observations seem to demand[4].

Just to see how the mechanics of the account could work, let us consider the derivations of (22) and (23) in more detail. It will be useful in what follows to assume that covert wh-movement happens quite late in the derivation, after overt movements have been triggered. We can start by deriving (23a), and we will begin, as before, by building up the entire tree, without performing any wh-movement at all until matrix C is reached:

(24) $C_{1\ [wh]}$ you v_1 want C_2 to v_2 persuade [owners of PG]

[C_3 to v_3 talk to **who** about **what**]

The lone active wh-feature, contributed by matrix C, will Agree with the closest inactive wh-feature, which is on matrix *v*, and this feature will propagate down through the tree, as before, until it reaches the embedded clause:

(25) $C_{1\ [wh]}$ you $v_{1\,[wh]}$ want $C_{2\,[wh]}$ to $v_{2\,[wh]}$ persuade [owners of PG]

[$C_{3\,[wh]}$ to $v_{3\,[wh]}$ talk to **who** about **what**]

All of these wh-features are now active, and the lowest one has as its closest target the higher of the two wh-phrases. This wh-phrase therefore moves, successive-cyclically, up to the matrix clause, acting as a Goal of each wh-feature in turn:

[4] It may be possible to bring this approach to bear on the Williams Cycle (Williams 1974, 2003, Abels 2008, Müller 2014, Keine 2019, and much other work), which is intended to prevent, for example, movement into the specifier of CP followed by movement into the specifier of a (higher) TP. In the account given here, the features of T which trigger movement are often present in the derivation earlier than the corresponding features in C, and we might therefore expect certain kinds of Williams Cycle effects to follow from the general cyclicity requirement being defended here. I will leave further development of this idea for future work.

(26) **who** $C_{1\,[wh]}$ you __ $v_{1[wh]}$ want __ $C_{2[wh]}$ to __ $v_{2[wh]}$ persuade [owners of PG]

[__ $C_{3[wh]}$ to __ $v_{3[wh]}$ talk to __ about **what**]

In English, active wh-features only trigger overt movement of a single wh-phrase. Contiguity Theory holds that overt wh-movement is triggered by phonological conditions imposed on Probe-Goal relations; it presumably also satisfies needs imposed by the semantic nature of questions. Covert wh-movement is triggered only by the latter, semantic factors. Since the driving principle determining the order of operations, in this model, is a condition requiring the computation to satisfy as many conditions as possible with each operation, and since overt wh-movement satisfies a superset of the conditions satisfied by covert wh-movement (not only semantic conditions, but also phonological ones), we expect overt wh-movement to precede covert movement in the derivation. Covert movement of *what* will therefore begin after overt movement of *who* has reached its landing site, with each covert movement tucking in under the existing specifier created by overt movement. In the specifiers of the *v*P hosting the parasitic gap, *who* is higher than *what*, and has the privilege of licensing the parasitic gap, following Nissenbaum (2000)[5].

---

[5] At this point in the account, one might wonder whether the conditions on parasitic gaps could be made to follow from the central idea of this paper; the parasitic gap must be licensed as soon as possible, and because the overt movement precedes the covert movement in the derivation, it is the overt movement which must license the parasitic gap. This idea will not survive the next case to be considered, involving nested wh-paths, both overt, landing in specifiers of different clauses; as reviewed above, in this kind of case, it is the wh-phrase which lands in the higher clause which must license the parasitic gap, and this will be the second wh-movement to be performed, not the first. We are apparently driven to the conclusion that "licensing parasitic gaps" is not among the requirements which the computation is driven to satisfy as quickly as possible; if it were, we would derive the wrong result for the nested wh-movement case. The issue seems to me to be connected to the larger question, largely unexplored in this paper, of when semantic representations are to be constructed.

Next we can turn to the case of nested overt wh-movement paths, repeated in (27):

(27) a. ? **What**$_2$ do you know **who**$_1$ to persuade [owners of __ $_2$]

to talk to __$_1$ about __$_2$

b. * **What**$_2$ do you know **who**$_1$ to persuade [friends of __ $_1$]

to talk to __$_1$ about __$_2$

As before, the derivation of (27a) will begin by constructing the tree up to the C which contributes the first semantically contentful wh-feature, which in this case will be the complement of *know*:

(28) $C_{2\ [wh]}$ to v_2 persuade [owners of PG]

[C_3 to v_3 talk to **who** about **what**]

As before, this feature will percolate down the tree, and trigger movement of the higher wh-phrase to itself:

(29) **who** $C_{2\ [wh]}$ to __ $v_{2\ [wh]}$ persuade [owners of PG]

[__ $C_{3\ [wh]}$ to __ $v_{3\ [wh]}$ talk to __ about **what**]

Now that the wh-features have all triggered all of the overt operations that they need to trigger (satisfying all of the phonological and syntactic needs imposed by these features), building of the tree can resume, eventually reaching the matrix C, which has an interrogative wh-feature of its own:

(30) $C_{1\ [wh]}$ you v_2 know **who** $C_{2\ [wh]}$ to __ $v_{2\ [wh]}$ persuade [owners of PG]

[__ $C_{3\ [wh]}$ to __ $v_{3\ [wh]}$ talk to __ about **what**]

The wh-feature on the matrix C must Probe, and finds its closest Goal on matrix *v*. The closest wh-phrase to matrix *v*, in turn, is the embedded CP (C_2P), and then the wh-phrase *who* which this CP has in its specifier. If C_2 had another overt specifier (as in Bulgarian), extraction of *who* would be possible; as it is, the wh-feature from C_1 must percolate further down the tree[6], eventually encountering the second wh-phrase, *what*. And what we are learning is that this second wh-movement, triggered by a distinct wh-feature, moves to specifiers outside the ones created by the first wh-feature (that is, that McGinnis (1998) and Rackowski (2002) are right when they claim that while a single Probe triggering multiple movements enforces 'tucking in', multiple Probes on a single head do not).

The above account has left several important questions unexplored. Why is 'tucking in' enforced by a single Probe, but not by multiple Probes? Why can a wh-phrase sitting in a 'criterial position' not undergo further movement in English? I think Contiguity Theory has answers to these questions, but will defer them for now to more specialized discussion. The only point I am trying to make in this section is that the approach to cyclicity being developed here offers a new way of answering the problem Davis (2020) uncovered with parasitic gaps, thanks to Fox' (2022) insight. We can say that the behavior of intersecting wh-paths is determined by the eventual landing sites of the wh-phrases, because in fact wh-movement does not begin until the landing sites have been established. What would be countercyclic on other approaches to cyclicity is simply another case, in this model, of an overarching generalization; operations take place as soon as the structures responsible for triggering them are in place, but crucially no sooner.

[6] This need of the wh-feature to forego triggering movement of the first actual wh-phrase that it finds is presumably the cause of the wh-island effect.

4. A further simplification?

In this paper I have been arguing for an approach to cyclicity built on the following requirements:

(31) a. Trees are constructed from the bottom up (External Merge is to the root of the tree under construction)

b. Operations must take place, if they are to take place at all, as soon as everything necessary for the operation is in the workspace (cf. Pesetsky 1989, Chomsky 1995).

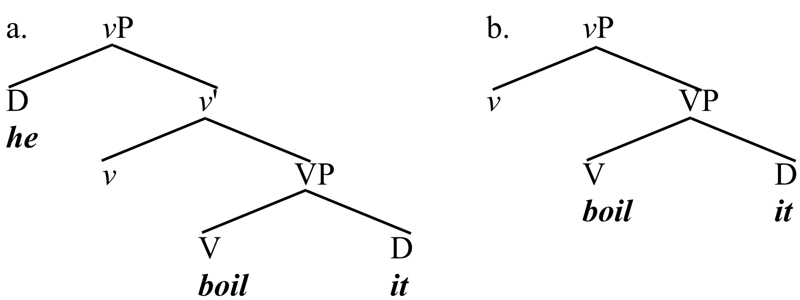
Is it possible to simplify (31) further? In particular, can we eliminate the residue of the Extension Condition in (31a)?

It seems to me that we might be able to derive the effects of (31a) from a particular understanding of (31b). Consider the derivation for the sentences in (32), which would start by assembling something like the *v*Ps in (33):

(32) a. He boiled it.

b. It boiled.

(33) a. b.



(32a) requires the derivation of the structures in (33) to begin by Merging the V *boil* with the D *it*—and not, for example, by Merging *v* with V, or the D *he* with *v*. How can this requirement be derived?

One difference between V and *v*, which I think we can make use of, is that once V and D have been Merged, there are no further requirements imposed by either of the Merged heads. This particular V selects for a nominal complement, and once it is provided with one, all its requirements are satisfied; it has no other operations to trigger. Similarly, the D sister of V will not itself trigger any syntactic operations; it has no Probes to trigger movement operations, nor any selectional requirements.

If we were to begin instead by Merging *v* either with V or with its DP specifier, the resulting tree would still have some requirements to satisfy: the V would not yet have its D complement, and if *v* Merged first with its specifier, it would still have the unsatisfied selectional properties which will lead to Merge of its complement VP. In other words, Merge of the V with its DP complement, uniquely among the various options, creates a tree in which there are no unsatisfied Probes or selectional requirements imposed by any material in the current tree.

We might be able to derive the preference for building the tree from the bottom up, then, from a particular understanding of the requirement of Generalized Earliness in (31b) above, which we could rephrase as in (34):

(34) Whenever possible, prefer operations which tend to create trees in which all selectional and Probe-Goal requirements have been satisfied.

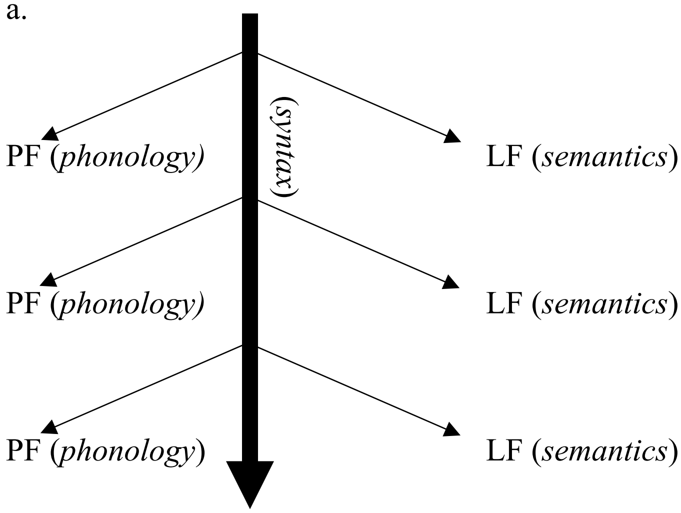
The condition in (34) would force requirement-satisfying operations to take place as soon as possible, as desired, and it would also prefer Merge of V with D over other possible Merge operations, since the resulting tree would be one in which (34) is maximally satisfied. And once the derivation has begun at the bottom of the tree, we can again rely on (34) to guarantee that Merge will proceed from the bottom up; once V and D have been Merged, for example, the next head to be Merged should be *v* (which selects for VP, and will therefore have a requirement satisfied by Merge), and not some higher head (which would not select VP). On this view, the preference for building trees from the bottom up is something like a preference for beginning jigsaw puzzles at their edge; the computation is seeking an easy place to start.

Numerous[7] questions[8] remain. I will leave these for future work.

5. Conclusion

I have proposed in this paper that we replace the model in (35a) with the one in (35b):

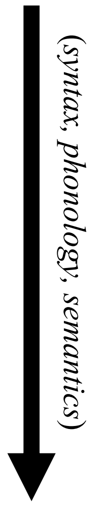
(35) a.



[7] If we consider, for example, a tree in which the lowest head in the clausal spine has both a specifier and a complement (the kind of tree we might want to consider for double object constructions, for example), there will be no choice of first Merge operations which will yield a tree in which all requirements are satisfied. Does this mean Merge may apply in any order in such cases? or that such trees are impossible? or should the requirement be refined to make a prediction in such cases?

[8] Another question has to do with the Merge of adjuncts. If adjuncts are not in Probe-Goal or selection relations with anything in the tree, then when should they be Merged? This may depend on how (34) is interpreted. Since Merging an adverb to V will not contribute to the satisfaction of any requirements, it should take place after Merge of V's complement, which is arguably desirable. But once V and its object have been Merged, should the adverb be Merged to VP before or after the VP is Merged to *v*? On the one hand, Merging the adverb prolongs the situation in which there are no Probes to satisfy or selectional relations to set up, which we might take to be an argument for preferring Merge of the adverb over Merge of *v*. On the other hand, Merge of the adverb will not contribute to the satisfaction of any of the requirements imposed by the other heads which will participate in Merge, and perhaps the computation is sophisticated enough to recognize this and prefer Merge of *v*, since that will allow for satisfaction of *v*'s needs. Again, I will leave the question here for now.

b.



The claim, in other words, has been that there is only a single generative engine, responsible for all linguistic phenomena (see Pesetsky (1985), Halle and Marantz (1993), and further work in Distributed Morphology for similar claims). The operations that we traditionally think of as syntactic, semantic, and phonological are all interspersed in the derivation. One condition regulating the ordering of these operations is a generalized version of Pesetsky's (1989) Earliness; every operation must take place as early in the derivation as it can. The hope is that by stating the conditions on operations with sufficient precision, we will also be stating the order in which they apply.

Bibliography

Abels, Klaus. 2008. Towards a restrictive theory of (remnant) movement. *Linguistic Variation Yearbook* 7.53-120.

Abusch, Dorit. 1985. *On verbs and time*. Doctoral dissertation, University of Massachusetts, Amherst.

Andrews, Avery. 1982. The representation of case in Modern Icelandic. In Joan Bresnan (ed.), *The mental representation of grammatical relations,* pp. 427-503. MIT Press, Cambridge, Mass.

Archibald, John, and Nicole Croteau. 2020. Acquisition of L2 Japanese WH questions: evidence of phonological contiguity and non-shallow structures. *Second language research* 37.649-679.

Bobaljik, Jonathan. 2000. The ins and outs of contextual allomorphy. In Kleanthes Grohmann and Caro Struijke (eds.), *University of Maryland Working Papers in Linguistics*, pp. 35-71.

Branan, Kenyon. 2018. *Relationship preservation*. Doctoral dissertation, MIT.

Bruening, Benjamin. 2020. Idioms, collocations, and structure: syntactic constraints on conventionalized expressions. *Natural Language and Linguistic Theory* 38.365-424,

Choi, Jaehoon, and Heidi Harley. 2019. Locality domains and morphological rules: phases, heads, node-sprouting and suppletion in Korean honorification. *Natural Language and Linguistic Theory* 37.1319-1365.

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

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: a life in language*, pp. 1-52. Cambridge, Mass.: MIT Press.

Copley, Bridget. 2002. *The semantics of the future*. Doctoral dissertation, MIT.

Davis, Colin. 2020. The nature of overlapping A-bar chains as revealed by parasitic gaps. In Mariam Asatryan, Yixiao Song, and Ayana Whitmai (eds.), *Proceedings of NELS 50*, pp. 145-158. GLSA, University of Massachusetts, Amherst, MA.

Deal, Amy Rose, and Matthew Wolf. 2017. Outward-sensitive phonologically conditioned allomorphy in Nez Perce. In Vera Gribanova and Stephanie Shih (eds.), *The morphosyntax-phonology connection: locality and directionality at the interface*, pp. 29-60. Oxford University Press.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. MIT Press, Cambrdge, Mass.

Fox, Danny. 2022. Crossing vs. Nesting, and Cyclic Linearization--Davis 2020. Class handout, MIT.

Fox, Danny, and David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31.1-45.

Frampton, John, and Sam Gutmann. 2000. Agreement is feature sharing. ms., Northeastern University.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Ken Hale and Michael Kenstowicz (eds.), *The view from Building 20*, pp. 111-176. MIT Press, Cambridge, Mass.

Keine, Stefan. 2018. Selective opacity. *Linguistic Inquiry* 50.13-61.

Marantz, Alec. 1991. Case and licensing. In German Westphal, Benjamin Ao, and Hee-Rahk Chae (eds.), *Proceedings of the 8th Eastern States Conference on Linguistics*, pp. 234-253. Cornell Linguistic Circle Publications, Ithaca, NY.

Marantz, Alec. 1997. No escape from syntax: don't try morphological analysis in the privacy of your own Lexicon. In *University of Pennsylvania Working Papers in Linguistics 4:2: Proceedings of the 21st Annual Penn Linguistics Colloquium*, pp. 201-225.
McGinnis, Martha. 1998. *Locality in A-movement*. Doctoral dissertation, MIT.
McGinnis, Martha. 2002. On the systematic aspect of idioms. *Linguistic Inquiry* 33.665-672.
Merchant, Jason. 2015. How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* 46.273-303.
Moskal, Beata. 2015. *Domains on the border: between morphology and phonology*. Doctoral dissertation, University of Connecticut.
Müller, Gereon. 2014. A local approach to the Williams Cycle. *Lingua* 140.117-136.
Nissenbaum, Jon. 2000. *Investigations of covert movement*. Doctoral dissertation, MIT.
Nunberg, Geoffrey, Ivan Sag, and Thomas Wasow. 1994. Idioms. *Language* 70.491-538.
O'Grady, William. 1998. The syntax of idioms. *Natural Language and Linguistic Theory* 16.279-312.
Pesetsky, David. 1982. *Paths and categories*. Doctoral dissertation, MIT.
Pesetsky, David. 1985. Morphology and Logical Form. *Linguistic Inquiry* 16.193-246.
Pesetsky, David. 1989. The Earliness principle. ms., MIT.
Preminger, Omer. 2014. *Agreement and its failures*. MIT Press, Cambridge, MA.
Preminger, Omer. 2022. Natural language without semiosis. Talk handout, University of Maryland.
Rackowski, Andrea. 2002. *The structure of Tagalog: specificity, voice, and the distribution of arguments*. Doctoral dissertation, MIT.
Rackowski, Andrea, and Norvin Richards. 2005. Phase edge and extraction: a Tagalog case study. *Linguistic Inquiry* 36.565-599.
Richards, Norvin. 2010. *Uttering trees*. MIT Press, Cambridge, MA.
Richards, Norvin. 2016. *Contiguity theory*. MIT Press, Cambridge, MA.
Richards, Norvin. 2019. Detecting Contiguity-prominence. ms., MIT.
Stepanov, Artur, and Ali Al Moussaoui. 2020. When a wh-word refuses to stay in-situ. *Linguistic Inquiry* 51.410-423.
van Urk, Coppe, and Norvin Richards. 2015. Two components of long-distance extraction: successive cyclicity in Dinka. *Linguistic Inquiry* 46.113-155.
Williams, Edwin. 1974. *Rule ordering in syntax*. Doctoral dissertation, MIT.
Williams, Edwin. 2003. *Representation theory*. MIT Press, Cambridge, MA.
Yamada, Akita. 2019. Contiguity Theory and Crystallization: *wh*-phrases and concord adverbs in Japanese. In Hisao Tokizaki (ed.), *Phonological externalization* vol. 4, pp. 1-22. Sapporo University.
Zwicky, Arnold, and Geoffrey Pullum. 1986. The principle of phonology-free syntax: introductory remarks. *Ohio State University Working Papers in Linguistics (Interfaces)* 32.63-91.