MOVEMENT
Contents

Introduction 1

Constituent Questions 21
  Alternatives 31

Quantifier Raising 39

Successive Cyclic Movement 59

Argument Movement 79

Bibliography 83
Introduction

Movement has been used to model a variety of syntactic relations that, frankly, oftentimes look quite different. Here are some examples.

(1) a. Mary kaupir ekki skó? (Icelandic)
   Mary buys not shoes Head Movement
   ↑

b. I asked which book Mary had read Ä Movement
   ↑

c. A child seems to have left. A Movement
   ↑

d. every bank a different guard stood before every bank. QR
   ↑

Why are we tempted to see each of these cases as special instances of the same relation? Perhaps because they (sort of) share these three properties.

(2) **Semantics Displacement**
   Some part of the meaning of the moved expression is applied to a position different from where it is spoken.

(3) **Terseness**
   The moved item semantically occupies two positions, but is spoken in only one of them.

(4) **Locality**
   The two positions that a moved item is related to are subject to a locality condition.

These properties don't manifest themselves in exactly the same way across these various kinds of movement, though, and so that’s a challenge to seeing these as shared properties.

A. **Difference in Semantic Displacement**
   a. **Total Reconstruction:**
      Mary kaupir ekki skó. ≡ ¬ Mary kaupir skó
b. **Variable Binding:**

Which book Mary had read ≡
The set of propositions such that \( \exists x \) Mary had read \( x \), \( x \) a book.

A guard stands before every bank ≡
\( \forall x \text{ if } x \text{ is a bank then a guard stands before } x \)

An argument that the semantic relationship between a moved verb and its underlying position is different than a moved phrase and its underlying position can be made from their different behavior in ellipsis contexts. VP Ellipsis can leave behind an object DP or a verb in contexts where that DP or verb has moved out of the VP. In English this is possible with a DP, in what goes by the name "pseudogapping."

(5) I know what kind of rice she will eat but not what kind of starch she should.

The position vacated by the moved object must be semantically equivalent to the matching position in the antecedent VP for the ellipsis to be licensed. That position must be occupied by a variable.

(6) What kind of rice \( x_1 \) she will [VP eat \( x_1 \)] but not what kind of starch \( x_2 \) she should [VP eat \( x_2 \)]

Because [VP eat \( x_1 \)] and [VP eat \( x_2 \)] are semantically equivalent for the purposes of ellipsis resolution, VPE is possible. The same is not the case, however, for (some) situations in which a verb has moved out of a VP that elides. See Goldberg (2005). The following are her examples, inspired by Doron (1999).

(7) Q: Ha'im Miryam hevi'a et Dvora la-xanut?

Q  Miryam bring ACC Dvora to.the-store
'Did Miryam bring-past Gvora to the store?'

Ai: Ken, hi hevi'a.
yes, she bring-past
'Yes, she brought.'

Aii: * Ken, hi lakxa.
yes, she take-past
'Yes, she took.'

This is derived if the movement in these cases doesn't leave a variable, but is instead, semantically vacuous.

(8) Q: Ha'im Miryam hevi'a [VP hevi'a Dvora la-xanut ]

A: Ken, hi lakxa [VP lakxa Dvora la-xanut]

Because the VPs in these examples aren't semantically equivalent, VPE isn't allowed.
B. Differences in Locality

a. Head Movement Constraint:

* Have Mary should read a book.

↑

b. Ross's Islands

Which book has Mary shown \[CP \text{ would change your life}]?

↑

* Which book has Mary shown \[DP \text{ the proof } [CP \text{ that would change your life}]\]?

c. Tensed S Condition

* A child seems \[CP (that) \text{ has left}].

↑

* every bank a different guard showed \[CP \text{ that the road stood before every bank}].

↑

C. Differences in how Terseness is violated

(9) 崦ṅṅō ń wà ṇā ń kā 崦ṅṅō ā

sleep you want NA you FUT-A sleep Q

(Vata)

'Do you want to sleep?'

(Koopman, 1984, (2a): 154)

Here "崦ṅṅō" ('sleep') has been clefted and is pronounced in both the cleft position and the position inside its VP. This is probably movement since locality conditions are satisfied.

(10) * tākā ń wà fōtō` mūmō` ń tākā bō ābā

show you like picture ITIT you showed REL Aba

'It's show that you like the picture you showed Aba.'

(Koopman, 1984, (15): 159)

When a verb clefts in Vata, both copies must be pronounced. (9) is ungrammatical if either verb isn't pronounced. When nominal material clefts, by contrast, only the higher copy may be pronounced. (11) illustrates.

(11) 崦ṅṅōlī ń mī ń wà ā

sleep-NOM IT you want Q

'Is it sleeping you want?'

(Koopman, 1984, (2b): 154)

What we have in these examples is a kind of \(\overline{A}\) movement of a verb. When verbs, or predicates, move, we sometimes get violations of Terseness that involve two pronunciations of the moved predicate, and that's what Vata illustrates.

I don't know of anything similar, though, when a DP has \(\overline{A}\) moved. In those cases, violations of Terseness take a different shape. They produce resumptive pronouns. For instance, in Lebanese Arabic there are resumptive
pronouns that show Semantic Displacement effects when islands aren't violated. (See Sichel forthcoming and references therein.)

(12)  
\[ \text{talmiiz-}a_1 \text{ ik}sleen ma baddna ny\text{abb}i \{wala m\text{'allme}\}_1 \text{ ?onno student-}h_1 \text{ the-}b\text{ad NEG want-}1P \text{ tel}l-1P \{\text{no teacher}\}_1 \text{ that} \]
\[ \text{h}u\text{ww}e za'\text{bar} \quad b-l-faf\text{S} \]
\[ \text{he cheated.3}S\text{M in-the-exam} \]
\[ \text{her bad student, we don't want to tell any teacher that he cheated on the exam.} \]

(13)  
*\[ \text{talmiiz-}a_1 \text{ ik}sleen ma badkun t\text{abb}ro \{wala m\text{'allme}\}_\text{?an student-}h_1 \text{ the-}b\text{ad NEG want-}2P \text{ tel}l.2P \{\text{no teacher}\}_1 \text{ about} \]
\[ \text{l-bont yalli h}u\text{ww}e za'\text{bar} \quad \text{ma'\text{-a}} \quad b-l-faf\text{S} \]
\[ \text{the-girl that he cheated.3}S\text{M with-her in-the-exam} \]
\[ \text{Her bad student, you don't want to tell any teacher about the girl with whom he cheated on the exam.} \]

(Aoun et al., 2001, (25b) & (29b): 381–2)

I think, then, that there is a difference in how Terseness is violated depending on the category of the thing being moved. There is a potential problem for this belief in the wh-copying construction that colloquial German (and other languages) display.

(14) German

a. \[ \text{Wen glaubt John wen Mary getroffen hat?} \]
\[ \text{who thinks John who Mary met has} \]
\[ \text{Who does John think Mary has met?} \]

b. *\[ \text{Wieviel Geld meint sie wieviel Geld das kostet?} \]
\[ \text{how-much money thinks she how-much money that costs} \]
\[ \text{How much does she think that costs?} \]


It’s hard to say in these situations whether the intermediate copy is a resumptive pronoun or a copy. But there are certain properties of the copy construction which suggest that it is different from a “normal” movement construction with just a simple violation of Terseness. There are interesting semantic differences between the two constructions. A striking one is described in Felser (2004). She notes that in cases where a phrase has moved in across-the-board fashion out of two coördinated clauses, as in (15), the question seems to assume that the answer will provide individuals that meet the descriptions provided by both of the clauses.

(15) \[ \text{Wen glaubst du, dass sie getroffen hat und dass sie liebt?} \]
\[ \text{who think you that she met has and that she loves} \]
\[ \text{Who do you think that she met and that she loves?} \]

(Felser, 2004, (37a): 560)
By contrast, a parallel across-the-board movement but with the wh-phrase pronounced in the lower positions as well, as in (16), seems to assume that the answer will provide the identity of individuals that meet the descriptions provided in each of the clauses separately.

(16) Wen glaubst du, wen sie getroffen hat und wen sie liebt?

who think you who she met has and who she loves

‘Who do you think that she met and that she loves?’

(Felser, 2004, (37b): 560)

This difference in meaning suggests that there is a separate quantification, one for each of the lower wh phrases, in the copy construction that is absent in the non-copy construction version. If that is correct, it will require that the copy construction include more quantificational expressions than are found in the simpler, single pronunciation, movement structure.

I’ll assume that there is a difference between moved DPs and moved other things that is responsible for how Terseness is relaxed in them.

What we want, then, is a theory of movement that explains these three properties: Semantic Displacement, Terseness and Locality. But that theory should also be flexible enough that it gives us a handle on why these three properties manifest themselves differently depending on the particulars of the movement operation. I’m going to take a few, very small, steps in that direction, building on an idea about what movement is that was in an early unpublished manuscript by Stanley Peters and Robert Richie, carried forward by Engdahl (1980) and has now found many proponents, including Gärtner (1997), Starke (2001), Nunes (2001), Frampton (2004), Citko (2005), Kobele (2006) and de Vries (2007). That idea is that movement gives an expression two positions by re-merging it.

(17) \[ \text{merge}(\alpha, \beta) = \text{def.} \gamma, \text{where the linear order of } \alpha \text{ and } \beta \text{ is not determined.} \]

A derivation that involves movement:

(18) (She asked) which book he knows.
You can see how (18e) provides a way of capturing Semantic Displacement. The moved term — here *which book* — is syntactically in two positions and so its denotation has two positions where it can be applied. The differences in how Semantic Displacement arise are going to come about, I will claim, from the particular ways in which the expressions that are "moved" get broken up into two different positions. That is going to be the focus of most of my lectures, but we won't start that process until later.

It's not obvious that these representations provide any new handle on the question why movement is subject to locality conditions. I've taken a very vague stab at trying to make that connection in a paper that was delivered in the 2009 meeting of the Chicago Linguistics Society. The idea in that paper uses a constraint on derivations that Chomsky proposed which, I think, he called the "extension condition."

(19) Extension Condition
\[ \text{merge}'s \text{ arguments must be root nodes.} \]

The Extension Condition is one attempt to derive the fact that moved items always go "up" in a phrase marker. It would block the derivation in (18), and there is no way of getting to the representation at the end of that derivation without violating the Extension Condition. So if the Extension Condition is correct, then the closest we can come to (18) is with something like (20).

(20) a. VP
\[ \text{knows which book} \]

b. TP
\[ \text{knows which book} \]
See Citko (2005). Suppose islands are those phrases at which a phonological or semantic evaluation must take place. This might be what "phases" are. If islands are phases that arise at the point in a derivation where there remain two roots, as in any but the last of the steps in (20), then arguably a semantic and phonological evaluation will not be possible, as our semantics and phonology are (perhaps) defined only for representations with a single root. This, at any rate, is one way of thinking about how to seek an answer to the question of why islands emerge uniquely with movement on an account that uses multidominant representations. I'll have nothing more to say about islands in these classes. I'll focus on the remaining two properties: terseness and semantic displacement.

We'll start by tackling how multidominant representations derive Terseness. I'll adapt a popular explanation for Terseness due to Jairo Nunes. Nunes worked with a (possibly) different account of movement: the "copy" theory of movement. This theory does not countenance multidominant representations, like that in (18e), but instead involves a "copy" operation.

(21) \textit{copy}(\alpha) = \alpha', an exact syntactic and semantic replica of \alpha.

This gives us derivations like (22).

(22) (She asked) which book he knows.

a. VP  
V  
|  
knows which book  
b. TP  
T  
|  
VP  
|  
V  
|  
DP  
|  
knows which book  
c. TP  
DP  
|  
\Delta  
he  
|  
T  
|  
VP  
|  
V  
|  
DP  
|  
knows which book
This theory too is able to account for Semantic Displacement, and it does so in a way rather like that of the remerge account I will argue for. So, for instance, it gives an account of "reconstruction," a special case of Semantic Displacement that (23) illustrates.

(23) Which story about her should none of the women forget?
There is a copy of *her* spoken in a place different from where it is (apparently) interpreted.

These representations make Terseness arise from an operation that “deletes” one of the two phrases in the copy relation. I’ll sketch the way this is done in Nunes (2004, chapter 1), which is a reworked version of his 1999 University of Maryland dissertation, and an improved version of Nunes (1995), and then I’ll modify it so it works with phrase markers with multidominance in them.

Nunes speculates that there is a deletion process that can be invoked to remove (portions of) one of the copies. He points out that adopting the simple assumption that a term and its copy cannot be distinguished by the constraints that define a well-formed linearization will cause movement structures to be unlinearizable. The deletion process he proposes could be invoked to “fix” these representations, making them linearizable and also deriving Terseness.

To see this, we’ll need to spell out what those constraints are and what linearizations are. I will assume that syntactic representations are converted into phonological representations (PFs) by matching vocabulary items to terminals in the syntactic representations and linearizing those vocabulary items. I shall adopt the formalism, made popular by Kayne (1994), of expressing a linearization as a set of ordered pairs. A linearization results from an algorithm which evaluates a syntactic structure and computes from the information in that structure how each vocabulary item in the structure is ordered relative to every other vocabulary item in the structure. So, for instance, the structure in (24) would map onto the ordered pairs in (25).
Note that I’ve represented the stem, or root, of the verb kaupir here with kaup. The set of ordered pairs in (25) involves the ordering relation “<”, precedence. (25) corresponds to the string in (26).

(26) Mary kaupir kaup skó

This isn’t quite right. We’ll come back to its correction a bit later.

One thing to highlight about (25) is that the elements in the ordered pairs are words, or vocabulary items. They are not the terminals that make up those words. This differs from Kayne’s work, but it is what Nunes, and I, need. Thus, for instance, kaup and ir are not arranged according to the linearization algorithm.

In Kayne’s work, the linearization procedure produced linearizations which were then subjected to well-formedness conditions. These conditions require that everything in the sentence be linearized and that the linearization be consistent. I will formulate those conditions as (27).

(27) a. All vocabulary items in the phrase marker \( p \) must be in the linearization of \( p \).

(Totality)

b. For all vocabulary items, \( a \) and \( b \) in \( p \), the linearization of \( p \) cannot include both \( a < b \) and \( b < a \).

(Antisymmetry)

c. For all vocabulary items, \( a, b, c \) in \( p \), if the linearization of \( p \) includes \( a < b \) and \( b < c \) then it must include \( a < c \).

(Transitivity)

He then builds a linearization algorithm that has a variety of interesting consequences for the shapes that phrase markers may have.

In addition to Totality, Kayne’s linearization procedure includes another statement that forces all the vocabulary items it applies to to be linearized. This has the consequence that multidominant phrase markers are blocked.

(28) The linearization algorithm necessarily applies to every thing it can.
I won't adopt this condition, but we will want to put in place something that
generally has the effect of mapping phrases onto contiguous strings. I will
therefore add to (27) a violable constraint, I'll call it "Contiguity," following
Fanselow and Ćavar (2001, 56: 130), that has this consequence.

(29) Contiguity

Let \( A \) be the set of vocabulary items dominated by \( A \) and \( b \) be a vocab-
ulary item not in \( A \). For every \( A \) in a phrase marker, if a linearization
includes \( b < a \) then it cannot include \( a^' < b \), for \( a, a^' \in A \), and if it
includes \( a < b \), then it cannot also include \( b < a^' \), for \( a, a^' \in A \).

These constraints – Contiguity, Totality, Antisymmetry and Transitivity
– are sufficiently draconian that they manage to constrain the structure-to-
string mapping almost enough to ensure reasonably accurate outcomes.
Imagine that the linearization algorithm did nothing more than generate
all possible orderings of vocabulary items and submit them to the constraints.
The strings produced would include the correct one and a small number of
alternatives. For instance, a linearization algorithm of this sort would produce
from (30) a collection of sets that, once filtered through the constraints, would
result in those listed in (31). (I will indicate the linearizations with the (more
compact) strings they correspond to, rather than with the full sets of ordered
pairs.)

(30) \[
\begin{array}{c}
\text{TP} \\
\text{DP} & \text{TP} \\
\text{this sentence} & \text{T} & \text{VP} \\
\text{V} & \text{DP} \\
\text{illustrates} & \text{agreement}
\end{array}
\]

(31) a. this sentence T illustrates agreement
b. sentence this T illustrates agreement
c. this sentence illustrates agreement T
d. sentence this illustrates agreement T
e. this sentence T agreement illustrates
f. this sentence agreement illustrates T
g. sentence this T agreement illustrates
h. sentence this agreement illustrates T
i. T illustrates agreement this sentence
j. T illustrates agreement sentence this
k. illustrates agreement T this sentence
The ill-formed linearizations in (31) are, many of them, well formed in other languages. For instance, (31f) corresponds roughly to how German would linearize this structure, and (31i) corresponds roughly to how Nuiean would. While not all of these outcomes are ones that we might want to permit cross-linguistically, I will nonetheless treat them all as language-particular possibilities. The step from this range of linearizations to the one that is correct for English, then, engages that component of the theory which models word order variation. There are a variety of proposals in the literature on how to model word order variation. One of those is built into Kayne’s linearization scheme. We don’t need to choose among them, though, and it will be convenient (and harmless) to avoid engaging the details. In what follows, therefore, I will leave open how the choice from the possibilities allowed by the constraints to the one appropriate for English is made. I will call that portion of the linearization procedure that makes the language particular choice, the “language particular component.”

The linearization algorithm will have four parts, then: a function that produces orderings among the vocabulary items in a sentence, a set of constraints, a procedure that steers how those constraints choose the linguistically viable ones from that set, and then a final component — the language particular component — that picks the language particular best ordering. The function that produces all the possible sets of orderings among vocabulary items in a phrase I will call lin.

\[(\text{lin}(P) = \text{def. the set consisting of every } L(P)).\]

Nunes’s method of deriving terseness hinges on the proposal that Anti-symmetry, and the other constraints on a linearization, cannot distinguish a copy from the thing it is copied from. Moreover, the way Nunes executes his idea relies on Kayne’s requirement that lin necessarily applies to every vocabulary item in the structure being linearized. Nunes system, then, invokes (33) and (34).

\[(\text{lin applies to every } X^0 \text{ in a phrase marker).}\]

\[(\text{For Anti-symmetry, } \alpha \text{ and } \text{Copy}(\alpha) \text{ are the same thing).}\]

When these are coupled with the copy theory of movement, they will produce unlinearizable results.
Consider, for instance, what *lin* and the language particular component will together produce for a phrase marker created by movement.

\[
\begin{array}{c}
\text{(35)}\\
\text{CP}\\
\text{DP'} \\
\text{who'} \\
\text{C} \\
\text{TP} \\
\text{did} \\
\text{DP} \\
\text{TP} \\
\text{she} \\
\text{T} \\
\text{VP} \\
\text{V} \\
\text{DP} \\
\text{visit} \\
\text{who}
\end{array}
\]

The linearization of (35) that satisfies the language particular component as well as Totality, Contiguity, and Transitivity is (36).

\[
\text{(36)} = \text{who'} \text{ did she visit who}\\
\begin{cases}
\text{who'} < \text{did} & \text{did < she} & \text{she < T} & \text{T < visit} & \text{visit < who} \\
\text{who'} < \text{she} & \text{did < T} & \text{she < visit} & \text{T < who} \\
\text{who'} < \text{T} & \text{did < visit} & \text{she < who} \\
\text{who'} < \text{visit} & \text{did < who} \\
\text{who'} < \text{who}
\end{cases}
\]

This linearization has pairs like *who*’* < visit* and *visit < who* in it, and under Nunes’s proposal, these will be violations of Antisymmetry: *who*’* and who* are indistinguishable for Antisymmetry, and these pairs amount to saying, then, that *who* both precedes and follows *visit*.

To produce the correct outputs, Nunes suggests that there is a deletion process which removes the items introduced by movement that cause the violation of Antisymmetry. That deletion process is called “chain reduction.” It can be formulated as (37).

\[
\text{(37) Chain Reduction}\\
\text{Delete } \alpha, \alpha \text{ a syntactic constituent that has been put into the Copy relation.}
\]

Chain Reduction should be seen as an operation that is part of PF; the component of the grammar that converts syntactic representations to phonological ones. We want Chain Reduction to remove material from the phonological representation of the sentence, but leave unaffected how that material is semantically interpreted. In the case of (35), it could remove DP or DP’ and thereby produce representations from which the linearization algorithm produces the strings in (38).
Both these representations satisfy Antisymmetry, and they also satisfy Totality, if Chain Reduction is seen as removing the terminals that Totality requires be in the linearization. Of these, only (38a) is the correct one. We can credit this to the language particular component.

This is how Terseness is derived.

Moreover, this method provides an interesting account for those examples where Terseness seems to be relaxed. For instance, Vata predicate cleft constructions involve a structure like (39), in which the verbal root has moved and joined with a functional head that encodes focus.

\[(39) \quad \bar{l} \ d\bar{a} \ s\acute{a}k\acute{a} \ \bar{l}i \]
\[\text{eat she/he perf rice eat} \]
\['\text{she/he has eaten rice.}'\]

\[(40) \quad \text{FocP} \]
\[\quad \text{Foc}^0 \quad \text{TP} \]
\[\quad \bar{l}i \quad \text{Foc} \quad \text{DP} \quad \text{TP} \]
\[\quad \Delta \quad \text{T} \quad \text{VP} \]
\[\quad \framebox{\text{d\acute{a}}} \quad \text{DP} \quad \text{V} \]
\[\quad \framebox{\text{\acute{s}\acute{a}k\acute{a}}} \quad \bar{l}i \]

Nunes argues that because the result of clefting a verb in Vata puts it within an \(X^0\), Antisymmetry is able to be satisfied without invoking Chain Reduction. If constraints like Antisymmetry make reference to vocabulary items, and not the terminals from which those vocabulary items are composed, then putting a copy into a vocabulary item will effectively "hide" it from Antisymmetry. The representation in (40), for instance, can be assigned the linearization in (41).
This linearization satisfies not only Totality, but Antisymmetry as well. When \( \text{Foc}^0 \) is matched against the vocabulary item that spells out "lì+Foc," the correct string associated with (40) is produced.

A virtue of the remerge theory of movement is that it derives Nunes’ stipulation that Antisymmetry treats something and its copy as the same thing. On the remerge theory of movement, there are no copies, and a moved item really is one thing. It’s one thing in two positions. I suggest, then, that we take Terseness to provide an argument for the remerge theory of movement.

Because the details of Nunes’ method of deriving Terseness rely on the copy theory of movement, we’ll have to translate it into something that fits the remerge theory. We can’t adopt Chain Reduction, for instance. Deleting the vocabulary items that have been put into two positions by \texttt{MERGE} will not create a representation that allows those items to be pronounced in just one of the two positions they occupy, as it did on Nunes’s scheme. Indeed, invoking an operation that is tied to the existence of Chains, in the manner that Nunes’s deletion operation is, also no longer has traction. There is nothing in a multidominant representation that corresponds to a chain. We will have to look elsewhere for the mechanism that brings these representations into compliance with Antisymmetry, and thereby delivers Terseness.

If we jettison the redundant stipulation that \texttt{lin} apply to every vocabulary item, and rely instead on just \texttt{Totality} (=(27a)), then the linearization algorithm I presented above already has a way of avoiding the violations of Antisymmetry that movement creates.

(27a) \texttt{Totality}

All vocabulary items in the phrase marker \( p \) must be in the linearization of \( p \).

Because \texttt{lin} is defined so that it generates every possible set of orderings, including those that are incomplete, it need not produce linearizations that will violate Antisymmetry to begin with. It’s the job of the constraints, and the language particular component, to determine which of the sets of orderings offered by \texttt{lin} survive. Because the version of Totality in (27a) only requires that every vocabulary item within a structure be mapped onto a position in the resulting string, it will allow orderings that do not take into account all of the positions a vocabulary item might occupy. For these reasons, then, \texttt{lin} need not produce an ordering that makes a vocabulary item that has two (or more) positions fall into more than one spot in the string. Since this is what Antisymmetry requires, this is how Terseness arises.

To see this, consider some of the orderings that \texttt{lin} will produce for (42). These include those in (43).
(42) Which child did she visit?

(43) a. = which child did she visit

\[
\begin{align*}
\text{which} & < \text{child} \\
\text{child} & < \text{did} \\
\text{did} & < \text{she} \\
\text{she} & < \text{T} \\
\text{T} & < \text{visit}
\end{align*}
\]

b. = did she visit which child

\[
\begin{align*}
\text{did} & < \text{she} \\
\text{she} & < \text{T} \\
\text{T} & < \text{visit} \\
\text{visit} & < \text{which} \\
\text{which} & < \text{child}
\end{align*}
\]

Both of these sets of orderings obey Totality, since every vocabulary item shows up in the resulting strings. They also obey Antisymmetry and Transitivity. However, they differ with respect to Contiguity, which is repeated in (44). They also differ with respect to that part of the language particular of English that requires wh-phrases to be spoken in their higher position. I will call that the Wh Criterion; we can formulate it with (45).

(44) Contiguity

Let \( A \) be the set of vocabulary items dominated by \( A \) and \( b \) be a vocabulary item not in \( A \). For every \( A \) in a phrase marker, if a linearization includes \( b < a \) then it cannot include \( a' < b \), for \( a, a' \in A \), and if it includes \( a < b \), then it cannot also include \( b < a' \), for \( a, a' \in A \).

(45) The Wh Criterion

If just one wh-phrase is merged to CP, then the grammatical Ls in \( \text{lin}(CP) \) must position that wh-phrase so that it precedes everything else in that CP.
(43a) obeys the Wh Criterion, but violates Contiguity. (It violates Contiguity because *she* (for example) precedes *visit* but follows the other vocabulary items in the VP) (43b), by contrast, obeys Contiguity but violates the Wh Criterion. The language particular component is, by definition, inviolable and therefore of these two orderings, (43a) is the better. Further, there is no way of avoiding a violation of Contiguity if the Wh Criterion is to be satisfied. In particular, there are no elements of \( \text{lin}(42) \) that by virtue of violating Totality, Antisymmetry or Transitivity manage to satisfy the Wh Criterion and also avoid violating Contiguity. There are no candidates that beat out (43a) by relying on violations of one of the other of our constraints on linearization because these will all involve either additional violations of the Wh Criterion or violations of Contiguity. We must understand Contiguity to be violable.

(46) Of the constraints on \( \text{lin} \), only Contiguity is violable.

(47) Every element of \( \text{lin}(P) \) which incurs \( n \) violations of Contiguity is ungrammatical if there is an element of \( \text{lin}(P) \) that incurs fewer than \( n \) violations Contiguity.

Consider next (48), which is also a member of the set of orderings that \( \text{lin} \) produces when applied to (42).

\[
\begin{align*}
\{ & \text{which} < \text{did} \quad \text{did} < \text{she} \quad \text{she} < \text{T} \quad \text{T} < \text{visit} \quad \text{visit} < \text{child} \\
& \text{which} < \text{she} \quad \text{did} < \text{T} \quad \text{she} < \text{visit} \quad \text{T} < \text{child} \\
& \text{which} < \text{T} \quad \text{did} < \text{visit} \quad \text{she} < \text{child} \\
& \text{which} < \text{visit} \quad \text{she} < \text{child} \\
& \text{which} < \text{child} 
\}
\]

This satisfies Totality, Antisymmetry and Transitivity, and so threatens to be a successful linearization. Under the system proposed here, there are two potential reasons that (48) is blocked. One is that it may not satisfy the Wh Criterion. This depends on how the underlying causes of the Wh Criterion pan out. It could be that it requires the entire DP headed by a wh-determiner to be spelled out in the merged-to-CP position. Irrespective of the Wh Criterion, however, Contiguity will disfavor (48) relative to (43a). Contiguity is violated in both (48) and (43a), but it is violated less in (43a). In (48), CP, both TPs, VP and DP violate Contiguity; in (43a), CP, both TPs and VP violate Contiguity, but DP doesn't. There is therefore one fewer violation of Contiguity in the case of (43a). Here, then, is the reason for letting the evaluation procedure be sensitive to the numbers of times that Contiguity is violated.

The Wh Criterion makes unavoidable a violation of Contiguity, and of those candidates that satisfy the Wh Criterion (and violate Contiguity), (43a) is the best: it violates none of the other constraints on linearizations and it violates Contiguity the fewest number of times required by the Wh Criterion. This correctly derives, then, that (42) maps onto the string in (43a). Without the Wh Criterion, (43b) becomes the winning linearization, because it
uniquely violates none of the constraints on linearizations. What we see, then, is that our linearization algorithm, in concert with a multidominant model of movement, allows overt movement only if there is a language particular constraint that forces a phrase to be spoken in the higher of its two positions. Further, when there is a language particular constraint forcing a phrase to be spelled out in its higher position, then Contiguity will favor spelling out all of that phrase in the higher position.

The account Nunes gives of those cases where Terseness is lifted is preserved on my recasting of his system. Consider how my system will apply to the case of the Vata clefted predicates, for instance.

(39)  /í  sá  lí
     eat she/he perf rice eat
     'she/he has eaten rice.'

On the remerge definition of movement, this will now have the representation in (49).

(49)

Because lin cannot see inside Foc\(^0\), the orderings it will produce do not make reference to the verbal root that is part of Foc\(^0\). Just as in Nunes’s scheme, the best output lin will produce is (41).

(41)  =  /í  sá  lí
     \begin{align*}
     \text{Foc}^0 < \dá & \ dá < \text{Foc}^0 \ sá  \ sá < \lí \\
     \text{Foc}^0 < \dá & \ sá < \dá \ lí \\
     \text{Foc}^0 < \sá  & \ sá < \lí \\
     \text{Foc}^0 < \lí &
     \end{align*}

Managing to avoid a violation of Antisymmetry works the same in both Nunes’s and my account. Notice as well that (41) also manages to avoid violating Contiguity.

This gives us an explanation for why the clefted verb in Vata can be spelled out in both positions. But recall that the clefted verb in Vata must be spelled out in both positions. Nunes’s ensured this by making the process that prevents spelling out the verb in both spots costly: Chain Reduction is employed
only to the point necessary. This explanation doesn’t have a correlate under the scheme I am proposing, however. Under both systems, that the verb is spelled out in the higher position is probably independently ensured by Totality. Totality requires that the vocabulary item made up of $\text{Foc}^0$ and the verb be part of the resulting string, and to the extent that this requires that the verb be part of that vocabulary item it will force the verb to be put in the higher position. So, what we need to figure out is why the verb must be spelled out in its lower position as well.

It’s useful to compare this situation with the one involving Icelandic verb movement that we started with.

(1a) Mary kaupir ekki skó? *(Icelandic)*

Mary buys not shoes

On the remerge theory, this will get a representation like (50).

![Diagram](image)

A winning output from *lin* — one that satisfies Totality, Transitivity, Anti-symmetry, Contiguity and, to the extent known, the language particular component too — is (51)

(51) = Mary kaupir ekki kaup skó

\[
\begin{align*}
&\{ Mary < T^0, T^0 < \text{ekki}, \text{ekki} < V, V < \text{skó} \} \\
&\{ Mary < \text{ekki}, T^0 < V, \text{ekki} < \text{skó} \} \\
&\{ Mary < V, T^0 < \text{skó} \} \\
&\{ Mary < \text{skó} \}
\end{align*}
\]

This isn’t the outcome we want. Here, we want to force the verb to not be pronounced in its lower position.

Why are the Icelandic and Vata outcomes opposite? I don’t know, but I speculate that it has to do with morpho-phonological requirements. Perhaps Icelandic has no vocabulary item that corresponds to a verbal root. Indeed, the citation forms of Icelandic verbs are not roots but, like English, inflected forms. If there is no vocabulary item that can be matched to the V position in (50), then this will explain why nothing is pronounced in this position.
For Vata, it may be that the lower verb is required to be spelled out so that the tones associated with the verb can be pronounced. Koopman (1984) shows that clefted verbs in Vata always appear with midtones in their cleft position, and display the tones associated with the verb in their lower copy only. This was illustrated in (9).

(9) ɲgənɔ̃ n  wà  nà  ń  ká  ɲgənɔ̃  á

sleep  you want NA you FUT-A sleep  Q
‘Do you want to sleep?’

(Koopman, 1984, (2a): 154)

Perhaps the tones associated with the verb are also subject to Totality in Vata. If so, whatever forces the clefted verb to be expressed with midtones will consequently force the verb to also be spoken in its lower position, where it can support the expression of those tones.
Constituent Questions

Today we'll start to take a closer look at how movement structures are semantically interpreted. Our focus will be Wh Movement.

Simple wh-questions have the shape in (52), on the remerge model of movement.

\[(\text{five}\)/\text{two}] (I know) which child she kissed.

A standard, simple, view of the meaning of questions is that they denote a set of propositions, each proposition offering a kind of answer in those cases where the question is answer-seeking. This is the view introduced by Hamblin and modified by Karttunen. One way of representing a set is with the \(\lambda\)-operator, which can be used to represent a function.

\[(\text{five}\)/\text{three}] \(\lambda x P(x) = \) that function which, when applied to \(a\), gives \(P(a)\).

A function can be equated with the set of things that that function holds of. For (52), for instance, we can give the question a denotation like:

\[(\text{five}\)/\text{four}] \(\lambda p \exists x x \text{ is a child} \& p = \text{she kissed } x\)

So, the challenge is to get this kind of meaning out of (52).

The central problem a remerge definition poses is that it baldly predicts that the single meaning that is associated with the moved item should be found in both of its positions. That isn't what we want from questions. Instead,
we must associate the moved wh-phrase with both a binder meaning and a variable meaning. The first person to appreciate, and try to solve, this problem is Elisabet Engdahl.

In Engdahl (1980), what she proposes is that the moved wh-phrase has two meanings, and they are introduced at their two positions. The meaning introduced in the lower position must be a variable. But it should also provide a way of explaining “reconstruction,” one of the manifestations of Semantic Displacement that is found in wh-movement.

(55) Which picture of himself should no one put on his website?

Note that it is the position from which movement has occurred that matters.

(56) a. Which picture of himself, does this indicate that no one should bring?
    b. * Which picture of himself, does the thing no one heard indicate I should bring?
    c. * Which picture of himself indicates that no one should bring it?

Engdahl (1980) suggested doing that by letting the NP part of the moved wh-phrase be interpreted in its lower position. We can speculate that the NP part is not interpreted in its higher (spoken) position however, since in this position it is neither c-commanded by its binder, nor meets the locality condition that anaphors typically impose on their antecedents.

There is another, somewhat less obvious, difficulty involved in capturing these reconstruction cases. This problem is easier to appreciate in cases involving universal quantification, like that in (57).

(57) 

\[
\begin{array}{c}
\text{DP}_1 \quad \text{CP} \\
\text{D} \quad \text{NP} \\
\text{which} \quad \text{picture of himself} \\
\text{CP} \quad \text{TP} \\
\text{C} \quad \text{DP}_2 \quad \text{TP} \\
\text{shouldn't} \\
\text{anyone} \quad \text{T} \quad \text{VP} \\
\text{VP} \quad \text{PP} \\
\text{V} \quad \text{t}_1 \quad \text{on his website} \\
\text{put} \\
\end{array}
\]

\[ [t_1] = \text{picture-of-himself}_2(x_1) \]

If we adopt a Hamblin/Karttunen style analysis of questions, and make the NP part of the moved phrase interpreted in its unmoved position, then for (57) we might get an interpretation along the lines of (58).
(58) \( \lambda p \exists x p = \text{not anyone}_2 \) should put picture-of-himself\(_2(x)\) on his website.

(58) characterizes the question as seeking the identity of a single picture with the expansive property of being of a bunch of guys, none of whom should put it on their website. That’s not what we want. We want something that allows the pictures to vary with the variable it contains. The anaphoric connection between a moved phrase and its trace must be capable of carrying this duty. Elisabet Engdahl gave us a way of doing that.

What she suggests is that the wh-phrase in the lower position gets an interpretation like the definite description in examples of “donkey anaphora,” like that in (59).

(59) Everyone who owns a donkey loves the/that donkey.

These definite descriptions also act like restricted variables. Following Cooper (1979), she adopted the view that they have buried within them a function that picks out individuals which the restrictor donkey tells us are donkeys. In (59), that function is something like “owned by \( y \).”

(60) Everyone, who owns a donkey loves the \( x \), such that \( f(x) \), & donkey\(_1\)

\[ f = \text{owned by } y_1 \]

Her idea, then, is that in the lower position, a wh-phrase is interpreted as a function that can contain a variable within it. The values this function gives can depend on the value given to the variable it contains. That’s how in (59) the donkeys are made to vary with the values given to everyone. Indeed, it is this function that we can see questions as asking for the identity of.

(61) a. Which picture did you say you’d show every girl?

\[ \lambda p \exists f p = \text{you said you’d show every girl}_1 f \]

\( f \) might be:

Sally \( \rightarrow \) the picture of the Eiffel tower
Mary \( \rightarrow \) the picture of the Milkmaid
Myrtle \( \rightarrow \) the picture of hot-rods

\( f \) might be:

her favorite picture

\( f \) might be:

the picture of George Clooney

Some questions describe functions that can only depend on the values given to nearby quantifiers.

(62) Which whole number does every whole number precede?

The only function in the actual worlds that fit this question are ones like its successor, which vary as according to the value give every whole number.
We get from this model an account of why the trace can only get a value that varies with respect to expressions that c-command it.

Which picture did you show the guy every girl knows?

$$\lambda p \exists f p = \text{you showed the guy every girl knows } f$$

$$f$$ can’t be:
- Sally → the picture of the Eiffel tower
- Mary → the picture of the Milkmaid
- Myrtle → the picture of hot-rods

$$f$$ can’t be:
- her favorite picture

$$f$$ can be:
- the picture of George Clooney

What we need to do now is put into the meaning of the phrase in the lower position the contribution that the restrictor — the NP — makes.

The first step we can take, then, is to reinvoke Engdahl’s idea that the trace left by movement is semantically like a donkey-type DP. Let’s start by considering the syntax, and semantics, of these expressions.

If we start with a model of donkey-type DPs like that offered in Cooper (1979), we will want to build in a function whose arguments can be bound. We should notice that it is not just definite descriptions that can have this interpretation in donkey anaphora sentences, but personal pronouns can as well:

$$\lambda p \exists f p = \text{you showed the guy every girl knows } f$$

$$f$$ can’t be:
- Sally → the picture of the Eiffel tower
- Mary → the picture of the Milkmaid
- Myrtle → the picture of hot-rods

$$f$$ can’t be:
- her favorite picture

$$f$$ can be:
- the picture of George Clooney

We should build into pronouns, traces and definite descriptions a relational meaning, then, and to the extent that this relation is the same in all these cases, we will want an explanation for why it travels in this particular pack. A commonplace idea about explaining the similarity between pronouns and definite descriptions is to adopt Postal (1969)’s proposal that pronouns are “intransitive” definite determiners. One way this can be expressed is by building the relational meaning into the determiner/pronoun. Let’s look at such a way that builds on Heim (2013). Its ingredients can be found in Fox (1999), Elbourne (2005), Chierchia (1995) and Rullmann and Beck (1998). The first innovation will be to let definite determiners/pronouns take two arguments: one is an index, and the second the NP you see in the case of (most) definite descriptions.²

² This comes from Elbourne 2005, chapter 3.
The NP will express a presupposition. “n” represents an index, which we will take to be capable of being complex. It can bundle together a function and its arguments. Let $f^n$ be a variable ranging over functions of any adicity. We need to let $f$ be able to have more than one argument because of examples like (66).

\[(66)\]
\begin{align*}
\text{a. } & \text{ Which of his letters to her does every woman return to [her lover]?} \\
\text{b. } & \text{ The ones where he asks her for money.}
\end{align*}

(Engdahl, 1986, (77): 186)

I’ll let $f$, therefore have any number of arguments, including none.

I’ll follow Engdahl’s notation in representing the valency of $f^n$ as follows.

\[(67)\]
\begin{align*}
\text{a. } & f^0 = \text{an } f \text{ with no variable in it (the constant function)} \\
\text{b. } & f^1 = \text{an } f \text{ with one variable in it} \\
\text{c. } & f^2 = \text{an } f \text{ with two variables in it}
\end{align*}

An index can have different shapes, depending on the valency of the $f$ it contains. I’ll represent the arguments of $f$ with $j$.

\[(68)\]
\begin{align*}
\text{a. } & n \\
\text{b. } & n \\
\text{c. } & n
\end{align*}

Let’s express how the index and the NP are put together with the denotation for the given in (69).²

\[(69)\]
\[\llbracket \text{the} \rrbracket = \lambda n. \lambda P : P(x) = 1. n\]

where the business between “,” and “” gives the conditions under which the function is well-defined, and so expresses the presupposition. When $f$ is a constant function, as it might be in definite descriptions in out of the blue statements, we’ll get something like (70).

² This is inspired by Heim (2013).
(70) The donkey needs a kiss.

\[
\text{donkey}(x) = 1.f^0
\]

\[
\text{DP} \quad \lambda P : P(x) = 1.f^0
\]

\[
\text{D} \quad \lambda y.\text{donkey}(y)
\]

\[
\text{NP} \quad \lambda n.\lambda P : P(x) = 1.n
\]

\[
\text{D} \quad f^0
\]

\[
\text{donkey}
\]

When \( f \) has an adicity greater than 0, it will come with silent arguments, which are represented with “j.” This is what we have in (64b).\(^4\)

(64b) Every man who owns a donkey kisses the donkey.

\[
\text{donkey}(x) = 1.f^j
\]

\[
\text{DP} \quad \lambda P : P(x) = 1.f^j
\]

\[
\text{D} \quad \lambda y.\text{donkey}(y)
\]

\[
\text{NP} \quad \lambda n.\lambda P : P(x) = 1.n
\]

\[
\text{D} \quad j
\]

\[
\text{donkey}
\]

Because \( j \) is c-commanded by every man in (64b), it, and the result of applying \( f^j \) to it, can vary in a way that depends on the values given to every man.

In this case, we must understand that the events which the uniqueness requirement of \(! \) are satisfied by are part of the quantification. That is, the denotation of every must make the events vary as well. (See Heim 1990 and Elbourne 2005.)

The case in (64a) could now be given an identical treatment; the chief difference between the two being that the presupposition is expressed as an NP in (64b), but as \( \phi \) features in (64a).

(64a) Every man who owns a donkey kisses it.

\(^4\) Giving determiners relational indices like this is in Chierchia (1995).
\[ 3^{rd}(f^j(j)) = \text{sing}(f^j(j)) = 1.f^j(j) \]

\[ \lambda P : P(x) = 1.f^j \]

\[ \lambda y.3^{rd}(y) = \text{sing}(y) = 1 \]

\[ \lambda n.\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]

\[ \lambda n,\lambda P : P(x) = 1.n \]
Note that the NP *book about her* in the higher copy is not semantically interpreted there. That's something we saw to be necessary for such cases. I've put the presupposition introduced by *[the] and *[book about her] in “< >.”

What's missing from (71) is the set of propositions part. All we've got here is a sentence that existentially quantifies over functions — functions that pick out books about x that no x, x a linguist, should forget. What we want is the set of such propositions. One way of doing that would be to give the C that heads questions a meaning that introduces the proposition part and then enrich the denotation of *which* so that it formed a set from that proposition. That is a standard method, and what I did in Johnson (2012), but I'll go a different direction in a moment, so let's hold off on completing this picture for a little while.

The standard account of the semantics of movement does precisely what we see in (71). That account is due to Danny Fox,⁵ who adopts a copy theory of movement and assumes that there is a rule which converts the lower copy into something that matches what we have in the lower position of (71). He calls that rule “Trace Conversion,” and the way he formulates it is in (72), which is slightly different from how I’ve built the meaning of the lower copy, but close enough to be roughly equivalent.

(72) **Trace Conversion**

In

\[
\begin{array}{c}
\phi' \\
\text{DP}_n \\
\ldots \text{DP}_n \ldots \\
\end{array}
\]

interpret \( \phi \) as a function that maps, \( x \), to the meaning of \( \phi[x/n] \).

\( \phi[x/n] \) is the result of replacing the head of every constituent bearing the index \( n \) in \( \phi \) with the head \( \text{the}_x \), whose interpretation, *[the]_, is:

\[ \lambda P : P(x).x. \]

(adapted from Fox, 2003, (52): 111)

Fox thinks of this rule as a kind of generalized binding rule. That is, it is meant to be the rule that relates binders to their variables. It might replace what I’ll call the Heim \( \varepsilon \)-Kratzer rule, for instance, which is designed to do just that. (Note that I have used the Heim \( \varepsilon \)-Kratzer rule in interpreting (71).)

(73) **The Heim \( \varepsilon \)-Kratzer rule. (see Heim and Kratzer 1998)**

In

\[
\begin{array}{c}
\phi' \\
\text{YP}_n \\
\ldots \text{x}_n \ldots \\
\end{array}
\]

change the denotation of \( \phi \) to \( \lambda n. \phi \).

To mesh with the copy theory of movement correctly, it's been imbued with the ability to change the meaning assigned to a determiner. But I think we must see his Trace Conversion rule, really, as something that is specific to

⁵ See Fox (1999). There are others, all closely related to Fox’s. Two of them are found in Sauerland (1998, 2004) and Romero (1998), as well as Engdahl (1980, 1986).
movement, and not general to all variable-binding relations. If we don't restrict it to just movement contexts, we wouldn't expect the inequality in (74).

(74) Every problem; challenges us to find no problem's; solution.

≠ Every problem; challenges us to find the/that problem's; solution.

I think we should be skeptical of Trace Conversion. It says that the syntax-to-semantics mapping allows for rules that change what a determiner means. I think we can legitimately ask why that should be so, and why we don't see things like it elsewhere.

Engdahl's approach is also troubling, I feel. It says that there are expressions with two denotations, and that they are tailor made for movement relations. But this feels ad hoc to me as well. Why should there be certain expressions whose two denotations just happen to be ones that provide a binder for the other denotation? That proposal doesn't put us in a position to try to answer the questions I laid out at the outset: why does movement cluster the properties it seems to?

The picture I'd like to replace these with is one that says that wh-movement involves putting together a definite description of the sort that we see in donkey anaphora, with a Q morpheme that produces the question and binds the variable in the definite description. I'll add one twist to that view, a twist that makes the definite description somewhat different from that found in donkey anaphors.

As a first approximation, I suggest something like (75).
(75) \( \exists f \text{ no linguist}_i \text{ should forget } f(j_i) \)  
\( < \text{book about her}_i(f(j_i)) = 1 > \)

\( \lambda P \exists f. P(f) \)  
\( \text{QP}_2 \)

\( \lambda P \exists f. P(f) \)  
\( \text{Q} \)

\( \lambda. \text{ no linguist}_i \text{ should forget } 2(j_i) \)  
\( < \text{book about her}_i(2(j_i)) > \)

\( \text{CP} \)

\( \text{C} \)  
\( \text{TP} \)

\( \text{should} \)  
\( \text{DP}_1 \)

\( \text{no linguist} \)  
\( \text{T} \)

\( \text{VP} \)

\( 2(j_i) \)  
\( < \text{book about her}_i(j_i)=1 > \)

\( \text{forget} \)

\( \text{DP} \)

\( \text{D} \)  
\( \text{NP} \)

\( \text{the} \)  
\( 2 \)  
\( j_i \)

\( \text{book about her}_i \)

On this view, technically what has moved is just the DP portion that is interpreted as a variable. This denotation it supplies to the object position of *forget*. That DP has merged with the higher Q, which is the binder of the index within the DP in object position. It has merged with that Q, but its denotation is not computed there. I’ve indicated that with the dotted line. As a consequence, the QP in the higher position has the same meaning as the Q which heads it.

We need to determine where in (75) the question word *which* is inserted. I am going to assume that it is the D position of the DP that gets matched to the question word. We need this because it is possible for *which* to show up on unmoved DPs, as in (76).

(76) Which student asked which question?

But I want the form this D has to reflect the fact that there is the question morpheme, Q, in the sentence. We need to ensure that when Q is present, there is a *which*. So I suggest that we let *which* be the exponent of the Q morpheme. When the exponent of some morpheme is expressed at the position that another morpheme has, we say that “Agreement” holds between them. So I will assume that Agree holds between Q and the D containing *which*.

This is different than what I did in Johnson (2012). There, I made *which* always be a determiner that Agrees with Q, and that has the consequence that every *which* phrase must move to Q. Because *which* phrases can be linearized in their unmoved position, sometimes this movement will have to be covert, and this doesn’t mesh easily with my view of QR.
The DP that has moved in (77) has two different positions that lin can calculate its position from. The grammar of English requires that lin position this DP according to its higher position.

Okay, that’s a start. We’ve got two things left to do before we’ll have a complete picture.

1. Where does the set of propositions part of the meaning come from in questions?
2. What causes Q and DP to merge in the particular way indicated in (77)?

**Alternatives**

To get these remaining pieces, I’ll start by looking at how questions are formed in (some) wh-in-situ languages. In these languages, the D that is found in the lower DP and the Q that binds off the variable in these lower expressions map onto separate morphemes. In Japanese, for instance, a morpheme on the verb marks the scope of a question, and in the position of the variable is an interrogative phrase.

(78) (Kimi-wa) dono-gakusei-ga natto-o tabe-tagatte-iru-to (you-top) which-student-nom natto-acc eat-desirous-be-C omoimasu-ka? think-Q (Which student do you think wants to eat natto?)

We might think of these languages as having the same syntax that I’ve given to English questions, but with a small difference in how the syntax-to-morphology
works. In Japanese, the D and Q are mapped onto separate morphemes and, perhaps relatedly, the shared DP is spelled out in the lower of its two positions.⁶ Alternatively, we could see the Q and the DP as being completely independent, and there being no remerge/movement in these examples.

Interestingly, though, in these kinds of questions there is (sometimes) a kind of intervention effect that does not arise in overt movement cases. This shows up for some dialects of Korean, according to Beck and Kim (1997). According to them, the presence of man (‘only’) in (79) is responsible for destroying the relationship between nuku and ni, thereby causing this sentence to be ill-formed.

(79)  * Minsu-man nuku-lûl po-ss-ni?
       Minsu-only who-acc see-Pst-Q
       ‘Who did only Minsu see?’

Beck (2006) provides an explanation for these intervention effects — sometimes called “Beck Effects” — that gives to questions a slightly different semantics than I have adopted. Her semantics will, it turns out, provide the missing pieces to our picture so far. So I will modify what we have to bring it in line with her analysis.

Her leading idea follows Hamblin (1973) more closely than it does Kartunnen in the Hamblin/Kartunnen style account of questions. Hamblin suggested that the question word in questions introduces not a variable that gets bound off, but instead introduces “alternatives” that the set of propositions which makes a question vary on. These alternative generating terms have also been used by Rooth (1985) to model focus. What will go wrong in (79) is that the focus sensitive operator man will interfere with the question particle ni’s access to the alternatives generated by nuku. Let’s take a brief, sketchy, look at this.

The idea in Rooth (1985) is that focused items have, in addition to their “normal” denotation another denotation that certain operators like only interact with. That other denotation – its focus value – is a set made up of alternatives to the term. In something like (80), then, Sally has a normal semantic value that allows it to refer to Sally, and the focus semantic value in (80b).

(80)  a.  [Sally] = the individual named Sally
     b.  [Sally] = { Jerry, Max, Sam, Sean, Mary,…}

Phrases that contain terms with a focus semantic value inherit a focus semantic value by composing their normal denotation with the term in a point-wise fashion. In the case of a verb composing with its object, the verb will compose with each of the alternatives in the focus semantic value of the object by function application, and produce a set of alternative VP meanings.

(81)  [visited Sally] = { visited Jerry, visited Max, visited Sam, visited Sean, visited Mary,…}

⁶ See Hagstrom (1998, 2000) and Kishimoto (2005) for proposal that would translate into my system that way.
Rooth then gives only a meaning that, when combined with the VP, returns the same ordinary semantic value that the VP has, but adds that all the members of the focus semantic value of the VP are false.

The idea in Beck (2006) is to let the wh-words have the same focus semantic value that focussed items do. But she suggests that they have no regular semantic value. This will cause the phrases they are contained in to have only focus semantic values: they will be sets of alternatives that vary only with respect to the value given to the wh-word. We are letting the part of a wh-word that the question abstracts over be a function. If we leave all the rest of our system the same, but import Beck’s idea that the “variable” in the wh-word is an alternative generator, we’ll get representations like that indicated in (82).

Now, Beck’s proposal is that the Q morpheme, among perhaps other things, converts the focus semantic value of its sister into a normal semantic value. So, we’d get something like (83).
Here, then, is the set of propositions component to the meaning of questions that we were searching for earlier. Indeed, this is equivalent to the meaning for questions laid out at the beginning.

What goes wrong in (79), then, is that man (‘only’) manipulates the focus semantic value of the clause its in before the question morpheme can get its hands on it. Interestingly, English doesn’t have these kinds of intervention effects. Something parallel to (79) is perfectly grammatical.

We need to make it matter where the wh-phrase gets spelled out — that is what distinguishes the Korean example from the English one. We have to learn about QR before I can address that problem, so we have to leave this question unanswered for a while.
What we need now is an answer to the question why Q merges with the DP in English questions.

For this, I need to turn to work by Seth Cable. In Cable (2007), he studies questions of the sort that Korean and Japanese illustrate, but his object of study is Tlingit, a Na-Dené language spoken in Western Canada and South-eastern Alaska. Like Korean, Tlingit has a wh-determiner and another morpheme — I’ll call it Q — in its questions. Like English, the wh-phrase moves overtly to the left edge of the question sentence. But, interestingly, unlike Korean or Japanese, the Q morpheme does not show up in what we might associate with the C position. Instead, it is merged with some phrase that contains the wh-phrase. (85) illustrates.

(85) Aadóo yaagú sá ysíteen?
who boat Q you.saw.it?
’Whose boat did you see’

(Cable, 2007, (212)–(213): 155-6)

In (85), the Q particle, sá, has merged with a DP, inside of which lies the wh-word: Aadóo. The whole thing has moved to the left edge of the sentence.

Moreover, Cable argues that the Q morpheme is in an Agreement relationship with the wh-word, and that there are locality conditions on that agreement relation that determine where the Q particle can be merged. Very roughly, that locality condition can be described with (87).

(87) Q can agree with D only if there is no lexical item that c-commands D but not Q.

The Q morpheme in our system is also in an agreement relationship with the wh-determiner, and so we should expect that, like Tlingit, it will have to merge in a position that does not take it too far from the wh-determiner it is agreeing with. This is what forces Q to merge to a phrase that it does not semantically combine with.

And, indeed, as Cable emphasizes, the range of phrases that the Tlingit Q morpheme can merge with are very close to the range of phrases that Pied-Pipe in wh-questions in English. (We must understand adpositions to be a non-lexical item, as it can intervene between Q and the wh-word in (88b).)
He argues, therefore, that Pied-Piping in English arises because the phrase that moves in English has a Q morpheme merged with it in just the way that Tlingit sá does.\(^7\) (89) illustrates.

\(^7\) See also Cable (2008).

(89) Which philosopher’s book about her should no linguist forget?

Now, Tlingit sá has to have a different semantics than we need for our English/Japanese/Korean questions, because it can show up in non-questions as well.

(90) Tléil aadóo yaagú sá xwsateen.
not whose boat Q I saw it
‘I didn’t see anyone’s boat.’

(Cable, 2007, (187): 141)
Cable gives it a semantics in Tlingit that, like what we need, operates on the focus semantic values of its complement and converts them into regular semantic values. In Tlingit, though, its existential force is derived from a higher, silent, operator. My suggestion, then, is that English Q has the semantics of the Q found in Korean, but the syntax of that found in Tlingit.
Quantifier Raising

We left last time with a picture of wh-movement that involves representations like that in (91).

(91) Which book should no one forget?

There is an Agreement relationship between Q and D that is responsible for making Q and the DP headed by D merge. That Agreement relationship is also responsible for ensuring that Q merges to a phrase containing D that is not too large. Moreover, there is a semantic relationship between Q and D. D introduces alternatives and makes the phrases that contain it have only a focus semantic value. Q converts those focus semantic values into regular values, and creates thereby, the question. This semantic relationship requires that Q's denotation be introduced where the question is, and this, in turn, requires that it not semantically combine with the phrase that it's merged to. Resolving these requirements triggers the multidominant structure that is seen in (91).

We get a picture, then, of the syntax-semantics mapping of questions that makes wh-movement look like a natural member of the class of question formation constructions that includes the in situ strategies of Korean and Japanese. It also allows us to dispense with ad hoc rules like Trace Conversion,
while preserving the ability that Trace Conversion provided in capturing reconstruction effects.

At the same time, it gives us an explanation for why Wh Movement obeys Terseness. The syntax-phonology mapping involves, I suggested, a linearization scheme that includes an operation that generates a family of orderings \( \text{lin} \) and a set of constraints that choose from that family the best one.

(92) Let \( L(P) \) be a set of ordered pairs, \( x < y \), where \( x \) and \( y \) are vocabulary items dominated by \( P \), and "<" means "precedes."

\[ \text{lin}(P) = \{ L(P) \mid \text{ satisfies the Linearization Constraints} \} \]

(93) Linearization Constraints

a. **Totality**

Every \( X^0 \) involved with a vocabulary item in the phrase marker \( p \) must be in an ordered pair with every other \( X^0 \) involved with a distinct vocabulary item in \( p \).

b. **Antisymmetry**

For all vocabulary items, \( a \) and \( b \) in \( p \), the linearization of \( p \) cannot include both \( a < b \) and \( b < a \).

c. **Transitivity**

For all vocabulary items, \( a, b, c \) in \( p \), if the linearization of \( p \) includes \( a < b \) and \( b < c \) then it must include \( a < c \).

(d. **Contiguity**

Let \( A \) be the set of vocabulary items dominated by \( A \) and \( b \) be a vocabulary item not in \( A \). For every \( A \) in a phrase marker, if \( b < a \), then there can be no \( a' < b \), for \( a, a' \in A \), and if \( a < b \) then there can be no \( b < a' \), for \( a, a' \in A \).

e. **Language Particular Component**

A set of constraints that correspond to the language particular word order choices.

I’ve introduced a new relation in Totality: “involved with.” For QR I need to make an informal part of the account I gave of head movement more explicit. Let’s re-look at the output of Head Movement.
In the Icelandic case, illustrated here, we want Totality to be satisfied for \( V^0 \) by virtue of the pair \( kaupir < skó \). We want this same pair to satisfy Totality for the \( T^0 \) that dominates \( ir \). We want \( lin \) to be able to use the position that \( T^0 \) is in to compute \( kaupir < skó \), and by so doing satisfy Totality for all of the \( X^0 \) "involved" with that word. One way we could ensure this is by simply letting \( lin \) order not vocabulary items but \( X^0 \)'s, and then let it only see \( T^0 \). But we want to let \( lin \) see \( V^0 \) in geometries like this, because sometimes (e.g., Vata), the material in \( V^0 \) enters \( lin \) in a different position. So instead, I'll define "involves."

\( X^0 \) is involved with a vocabulary item if it dominates part of it.

This allows \( kaupir < skó \) to make both \( T^0 \)'s and \( V^0 \) satisfy Totality.

Okay, with this detail, let's put back in our heads how this scheme works for Wh Questions. For (91), this system will deliver, among others, the two linearizations in (96).

\[
\begin{align*}
\text{(96)} & \quad \text{a. } \text{should no one forget which book} \\
& \quad \begin{cases} 
Q < \text{should} & \text{should < no one} & \text{no one < T} & \text{T < forget} & \text{forget < which} & \text{which < book} \\
Q < \text{no one} & \text{should < T} & \text{no one < forget} & \text{T < which} & \text{forget < book} \\
Q < T & \text{should < forget} & \text{no one < which} & \text{T < book} \\
Q < \text{forget} & \text{should < which} & \text{no one < book} \\
Q < \text{which} & \text{should < book} \\
Q < \text{book} \\
\end{cases}
\end{align*}
\]

\[
\begin{align*}
& \quad \text{b. } \text{which book should no one forget} \\
& \quad \begin{cases} 
Q < \text{which} & \text{which < book} & \text{book < should} & \text{should < no one} & \text{no one < T} & \text{T < forget} \\
Q < \text{book} & \text{which < should} & \text{book < no one} & \text{should < T} & \text{no one < forget} \\
Q < \text{should} & \text{which < no one} & \text{book < T} & \text{should < forget} \\
Q < \text{no one} & \text{which < T} & \text{book < forget} \\
Q < T & \text{which < forget} \\
Q < \text{forget} \\
\end{cases}
\end{align*}
\]

I've taken the expediency here of (often) representing the vocabulary items that are linearized with the \( X^0 \) that they are involved with. Both of these linearizations obey Totality, Antisymmetry and Transitivity. (96a) incurs no violations of Contiguity, and (96b) violates Contiguity four times: once each for \( VP \), \( TP \), \( TP' \) and \( CP \). The language particular component of English,
however, is not obeyed in (96a); English requires that a wh-phrase merged to a constituent question CP be spoken in the merged-to-CP position. I called this the Wh Criterion. Because the Wh Criterion is inviolable, and Contiguity is violable, of these two linearizations, (96b) is better. Indeed, when all other linearizations produced by \textit{lin} are considered, there are no other that have fewer violations of Contiguity and still manage to obey the language particular component, Antisymmetry, Totality and Transitivity. (96b), therefore, is how (91) gets linearized.

A problem I posed in the first lecture is how to give an account of movement that both captured the family resemblance that different forms of movement have, and yet allowed for various forms of movement to differ slightly in how they express that resemblance. The two properties that all movement operations share, I declared, are Terseness and Semantic Displacement. We've now seen how those two properties are captured for Wh Movement. Like the copy theory of movement, the multidominant one correctly correlates these two properties. Today I want to look at how the rule that moves quantificational DPs, Quantifier Raising (QR), fits into this picture.

I choose QR because it differs in, perhaps, the most ways from Wh movement and so considering these two movement operations gives us a kind of view of the limits of variation we want movement to be allowed.

Unlike Wh movement, QR produces a representation that is interpreted as a variable-binder pair. In questions we want to allow the NP to be semantically interpreted in the position it moves from but not in the position it is spoken. That is, for instance, necessary to get the right interpretation out of (98).

(97) every hole a marble rolled into every hole.

I choose QR because it differs in, perhaps, the most ways from Wh movement and so considering these two movement operations gives us a kind of view of the limits of variation we want movement to be allowed.

Unlike Wh movement, QR produces a representation that is interpreted as a variable-binder pair. In questions we want to allow the NP to be semantically interpreted in the position it moves from but not in the position it is spoken. That is, for instance, necessary to get the right interpretation out of (98).

(98) Which book about her, should no linguist forget?

The system I've presented so far not only allows the NP part of a moved wh-phrase to be interpreted only in its lower position, it forces that. QR is different. The NP part of quantificational expression must be interpreted in the position it is spoken. One way that can be seen is by considering the disjoint reference effect in (99).

(99) every book about Julie she likes every book about Julie.

If book about Julie could be interpreted in just its higher position, then we should expect Julie and she to be able to corefer. But they cannot. So here is one difference in how semantic displacement works that we should explain.

Indeed, there is evidence that the NP part of a quantificational expression must also be interpreted in the unspoken position to which QR moves the phrase. That can be appreciated by considering (100).

(100) A different woman regrets each of her mistakes.
If the subject is interpreted within the scope of each, then it cannot bind her. If her is bound by the subject, then each must not out-scope the subject.

(101) If each > a different woman, then a different woman doesn't bind her

This follows if her mistakes must be semantically interpreted in the same position that each is.

(102) The NP part of QR'd material must be semantically interpreted in both of its positions, but Wh moved material is able to be semantically interpreted in just its unspoken position.

This is the syntax-semantics difference that I will focus on deriving.

A more obvious difference in Wh Movement and QR, though, is how they get spelled out. The syntax of wh movement allows the moved expression to be spelled out in its higher position, and the language particular constraints of English pick that possibility for “single” constituent questions. In the case of QR, however, the moved phrase is always spelled out in its lower position. So, that’s one difference we would like to capture.

(103) A wh-moved phrase can be spelled out in either of its two positions, but a QRd expression can only be spelled out in its lower position.

These are the tasks for today: explain (102) and (103).

Let me start by sketching out a theory of QR that captures some of the standard effects it has been designed for. I’ll do this by employing the copy theory of movement, since that is how the present literature on QR is written.

One of the situations in which QR can be seen arises when a quantificational object contains an elided VP whose antecedent is the VP the object sits in. This can happen when a quantificational object comes with a relative clause, as in (104).

(104) She read every book that I did △.

I’ll adopt the account of these so-called Antecedent Contained Deletions (ACD) in Fox (2002). On Fox’s analysis, ACD is licensed when the relative clause containing the ellipsis has extraposed from the antecedent for that ellipsis.⁸

(105) She [VP read [DP every book ] ] (yesterday) [CP that I did △]

Fox (and Fox and Nissenbaum 1999) argue that these sorts of extraposition operations are the result of “late merging” a clause into a QRd DP.

⁸ The original argument for using extraposition structures as the source for ACD is Baltin (1987).
On the copy theory of Movement, a copy of the object DP is made and then it is merged into a position outside the VP which serves as antecedent to the ellipsis. Unlike Wh Movement, this higher copy goes unpronounced and, instead, the lower copy is spoken. However, into the higher copy is merged the relative clause containing the elided VP, and this relative clause is pronounced in the position occupied by the higher copy of QR. A reason for using movement, and the copy theory of movement in particular, to model QR is that it provides a simple account of extraposition from NP and, with it, a good account of ACD. It also captures a fact about ACD that Sag (1976) established: the position where an ACD in a DP is resolved is the same position that the quantifier heading that DP is interpreted.

We can now see another difference between QR and Wh Movement: when material in the higher copy in Wh Movement is pronounced, that material gets linearized so that it precedes everything else in the clause it is dominated by. But when material in the higher copy of QR is pronounced, it follows everything else in the phrase it is dominated by. Wh Movement goes to the left, but QR goes to the right. This difference too should be derived.

When Wh-moved material is spelled out in its higher position, it shows up to the left of the phrase it is merged to. When the QR’d phrase is spelled out in its higher position, it shows up to the right of the phrase it is merged to.

One of the most interesting pieces of support for Fox’s analysis of ACD is the contrast in (108), from Tiedeman (1995).

(108)  a. * I said that everyone you did △ arrived.
  b. I said that everyone arrived that you did △.
  △ = said that x arrived

(Fox, 2002, (35b), (36b): 77)
The difference is credited to extraposition being able to generate the string in (108b) but not (108a). The representation in (109) is only available for (108b). Once the English particular constraints apply to lin, it will put the material in the embedded subject between the complementizer and the embedded VP, and linearize the extraposed relative clause so that it follows everything the higher copy has merged with.

(109)

```
(109)  
VP  
  /\  
  |  /\  
V  | CP  D'  
  |  |  |   
|  |  |  | 
said  C  TP  NP  
  |  |  | 
|  |  |    
that  DP  TP  
  |  |   
everyone  T  VP  
  |   
arrived
```

Hidden in this example, however, is yet another illustration of the difference between the semantics of QR and Wh Movement. To see this difference, consider a derivation in which the relative clause is part of the lower copy, and therefore maps onto the string in (108a). This derivation will QR the subject and its relative clause together into a higher position, one that puts the elided VP outside of its antecedent, as in (110).

(110)

```
(110)  
VP  
  /\  
  |  /\  
V  | CP  D'  
  |  |  |   
|  |  |  | 
said  C  TP  NP'  CP'  
  |  |  |  
|  |  |  | 
that  DP  TP  
  |  |  |   
everyone  T  VP  
  |   
that you did say arrived
```

This representation is ill-formed only if the relative clause in the lower copy must be semantically interpreted. Here, then, is another illustration of (102).

Let's now consider how QR can be modeled in our remerge theory of movement. As with Wh Movement, let's let there be a definite description in the lower position that is bound by an operator in the higher position. Unlike
the case with questions, however, the determiner in the lower position and the quantifier in the higher position will both combine semantically with the NP. That is because the denotation of quantifiers (I will assume) requires that they relate the meaning of the NP they combine with with the meaning of the clause they are in. Quantifiers like *every*, for instance, have a meaning something like that indicated for $\forall$ in (111).

\[(111) \quad [\forall] = \lambda p. \lambda q. \forall x. p(x) = 1 \rightarrow q(x) = 1\]

The “traces” of QR are a little different semantically than the “traces” of wh-movement. They range over individuals, not Skolem functions. So:

\[(112) \quad [D] = \lambda n_\langle \rangle \lambda P : P(x). n\]

QR will therefore produce a representation like that in (113).

\[(113) \quad \text{A student read every paper yesterday.}\]

\[
\forall x. \text{paper}(x) = 1 \rightarrow \text{a student read yesterday}(x) = 1
\]

As with the case of questions, we want the form of D to be determined by the quantifier that is in the higher position. When the quantifier is $\forall$ we’ll want D to be spelled-out as *every*, when it’s “$\neg$ any” we’ll want D to be spelled-out as *no*, and so on. We can’t do this with Agree, as Q does not c-command D and Agree only holds between things that are in a c-command relation. I suggest instead that D+Q are brought together by the morphology. Morphology is equipped with processes that allow two terminals to be mapped onto one vocabulary item. These processes show up in a variety of ways. They are responsible for mapping a preposition and determiner onto one lexical item in certain contexts in German (and other IndoEuropean languages), for instance.
These processes go by various names, and there seem to be slight differences in the conditions under which they may operate that depends on the case being modeled. But across all these cases, there is a similar locality condition on the two terminals that combine: they must be so close together that, under normal circumstances, they would show up adjacent in the string. Let us simply adopt this descriptive consequence as a well-formedness condition on “fusion,” as I will call the process that brings D and Q together into one word.

If $X^0$ and $Y^0$ are matched to a lexical item $w$ by fusion, then for all $\alpha$

- a. If $\alpha < w$, then $\alpha < X^0$ and $Y^0$, and
- b. if $w < \alpha$, then $X^0$ and $Y^0 < w$.

This requires a model that lets the lexical items be matched to their $X^0$ positions after linearization has applied. Let us adopt that view then.

The language particular component of English causes Q and D both to come before everything else in the phrases they head.

**lin**(QP) must put Q before everything else in QP
**lin**(DP) must put D before everything else in DP

If D and Q were to be linearized in non-QR contexts, then, they would show up in adjacent positions. They are, therefore, possible fusers.

However, if D and $\forall$ are to try to fuse in structures of QR, then the condition in (115) will stand in the way. In (113), for example, **lin** will put things between D and $\forall$ and prevent adjacency. There are two strings that **lin** could produce from (113). If the NP containing paper is linearized in the spot assigned to QP, **lin** will deliver (117a), and if this NP is linearized in the spot assigned to DP, **lin** will deliver (117b).

- a. a student read D yesterday $\forall$ paper
- b. a student read D paper yesterday $\forall$

Both of these linearizations violate **Contiguity**, but they only violate **Contiguity** to the extent required by the multidominant representation. These, then, are the minimal violators of **Contiguity** and therefore the candidate winners. In both of them, D and $\forall$ are separated by yesterday and so fusion is blocked. If we assume that the English lexicon does not provide vocabulary items for the D or $\forall$ in these structures, we will have a sentence that cannot be spelled out.

I suggest, then, that fusion is achieved before the entire QR structure is built. This will require a model of the syntactic-linearization relation that is cyclic. We adopt a view that allows decisions about the linearization of a sentence to be made in the derivation, before the entire sentence has been put together. It has been argued that there is support for this view that is
independent of our project. Fox and Pesetsky \((2005)\), and several others, for instance have argued that one can derive certain "structure preservation" effects in this way. In certain – it must be admitted, very specialized cases – movement seems to be constrained so that it does not disturb the order the moved terms would have if they had not moved. To capture this, cyclic linearization schemes have the form in (118).

(118) a. `lin` can apply in the derivation and its results submitted to the language particular component.
b. `lin` forms the union of its output with all other outputs of `lin`.

The effects of cyclic linearization schemes rest on finding points in the derivation when `lin` must be applied. Where those points might be won't matter for our local goals. I'll assume that `lin` is allowed to apply anywhere in the derivation.

Let's imagine, then, the stage in the derivation that leads to (113) immediately before the QP has merged with TP. This will look like (119).

(119) 

\[ 
\begin{array}{c}
TP \\
\text{DP} \\
\text{a student} \\
\text{TP} \\
\text{T} \\
\text{VP} \\
\text{yesterday} \\
\text{VP} \\
\text{V} \\
\text{DP} \\
\text{read} \\
\text{D} \\
\text{NP} \\
\text{the} \\
\text{2} \\
\text{paper} \\
\text{QP} \\
\text{Q} \\
\text{∀} \\
\end{array}
\]

This representation has two root nodes: TP and QP. If we define `lin` so that it runs on root nodes, then in this case it will apply to TP and QP independently, and produce the two sets of ordered pairs in (120).

(120) \(\text{lin}(\text{TP}) = \{\text{a student read D paper yesterday}\} = \)

\[
\begin{align*}
\text{a} &< \text{student} & \text{student} < \text{T} & \text{read} < \text{D} & \text{D} < \text{paper} & \text{paper} < \text{yesterday} \\
\text{a} &< \text{T} & \text{student} < \text{read} & \text{read} < \text{paper} & \text{D} < \text{yesterday} \\
\text{a} &< \text{read} & \text{student} < \text{D} & \text{read} < \text{yesterday} \\
\text{a} &< \text{D} & \text{student} < \text{paper} \\
\text{a} &< \text{paper} & \text{student} < \text{yesterday} \\
\text{a} &< \text{yesterday}
\end{align*}
\]

\(\text{lin}(\text{QP}) = \{\forall \text{ paper}\} = \forall < \text{ paper}\)

This linearization puts nothing between D and \(\forall\) and so they can fuse. At this stage, then, let's match the `X0`'s that `lin` has ordered with their morphological forms. In the case of D and \(\forall\), this will mean that they get spelled out as `every`.
After QP and TP have merged, *lin* applies again. It will form the union of the two sets in (120) and add the following:

\[
\begin{align*}
\{ & a < \text{every} & \text{student} < \text{every} \\
& T < \text{every} & \text{read} < \text{every} \\
& \text{every} < \text{paper} & \text{every} < \text{yesterday} \}
\end{align*}
\]

These orderings become possible because *every* has now been formed by fusing D and ∀. Once these are collected into one set along with the two collections in (120) we’ll have a set that satisfies the language particular constraints, Antisymmetry, Transitivity as well as Totality. Totality is satisfied because *every* involves Q⁹, and so now (121) determines how Q⁹ is ordered relative to every other vocabulary item in the sentence.

Note that this last run of *lin* could not have introduced statements like (122).

\[
\begin{align*}
\{ & \text{paper} < \text{every} \\
& \text{yesterday} < \text{every} \}
\end{align*}
\]

That’s because these would be put into the same set that includes D⁹ < *paper* and D⁹ < *yesterday*, and this violates (115), the condition on fusion.

(121) corresponds to the string in (123), which is precisely the right outcome.

(123) a student read every paper yesterday

We’ve just derived that QR is covert, i.e., the moved phrase is linearized according to its lower position.

Let’s consider next how this system derives (102).

(102) QR’d material must be semantically interpreted where it is spoken, but Wh moved material is able to be semantically interpreted in only its unspoken position.

We’ve already seen how the case of Wh Movement works. An interrogative DP can be semantically interpreted entirely in its lower position, and yet be part of a phrase that *lin* puts in a different position. Let’s now consider why something parallel is not possible for QR. One of the cases I used to demonstrate (102) is (110). Under a copy theory account, this gets the representation indicated.

(110) * I said that everyone you did △ arrived.
This representation cannot resolve the ellipsis it contains and this indicates that the relative clause in the lower copy must be semantically interpreted.

If QR is modeled with the multidominant representations proposed here, however, the string in (110) can only get the representation in (124). Like the representation in (110), The structural requirements for resolving the ellipsis are not met in (124), either, since the ellipsis is inside its antecedent: \[\text{VP}\]

(124)

To see why it's this structure that corresponds to the string in (110), consider how \textit{lin} will deliver the linearization of (124).

As we've seen, \textit{lin} must run before the QP is merged to \[\text{VP}\] in order to fuse D and \(\forall\) into \textit{every}. So, the string associated with (124) will be built upon the output \textit{lin} produces from applying to (125).
That linearization is (126).

(126)

a. \( \text{lin}(\text{VP}) \ [\text{said that D one that you did T arrived}] = \)

\[
\begin{align*}
\text{said < } & \text{that } < \text{D} \quad \text{D < one} \quad \text{one < that} \quad \text{that < you} \quad \text{you < did} \quad \text{did < T} \\
\text{said < D} & \quad \text{that < one} \quad \text{D < that} \quad \text{one < you} \quad \text{that < did} \quad \text{you < T} \quad \text{did < arrived} \\
\text{said < one} & \quad \text{that < that} \quad \text{D < you} \quad \text{one < did} \quad \text{that < T} \quad \text{you < arrived} \\
\text{said < that} & \quad \text{that < you} \quad \text{D < did} \quad \text{one < T} \quad \text{that < arrived} \\
\text{said < you} & \quad \text{that < did} \quad \text{D < T} \quad \text{one < arrived} \\
\text{said < did} & \quad \text{that < T} \quad \text{D < arrived} \\
\text{said < T} & \quad \text{that < arrived} \\
\text{said < arrived}
\end{align*}
\]

b. \( \text{lin}(\text{QP}) \ [\forall \text{ that you did}] = \)

\[
\begin{align*}
\forall & < \text{one} \quad \text{one < that} \quad \text{that < you} \quad \text{you < did} \\
\forall & < \text{that} \quad \text{one < you} \quad \text{that < did} \\
\forall & < \text{you} \quad \text{one < did} \\
\forall & < \text{did}
\end{align*}
\]

When \( \text{lin} \) runs again, after \( \text{QP} \) and \( \text{VP} \) have merged, it will be possible to add ordering statements for \( \text{every} \), the result of merging \( \forall \) and \( \text{D} \). Those statements must be able to form a set that includes the ordering statements in (126) and pass the constraints. The only statements that will do that are (127).

(127) \[
\begin{align*}
\text{said < every} & \quad \text{that < every} \quad \text{T < every} \quad \text{arrived < every} \\
\text{every < one} & \quad \text{every < that} \quad \text{every < you} \quad \text{every < did}
\end{align*}
\]

Putting these statements together those in (126) will make a linearization that corresponds to the string in (128).

(128) \( \text{said that everyone that you did arrived} \)

This string therefore corresponds to a structure that does not resolve the ellipsis: just what we want.
To resolve the ACD, we must have a representation that involves "late merge" of the relative clause. Under the present proposal, this will look like (129).

(129)  
```
(\[\text{VP} \quad \text{QP}_1\])
```

In this structure, the relative clause is not within the VP that serves as antecedent, and so the ellipsis can be resolved. To see that this structure does not correspond to the string in (110), consider how \textit{lin} will manufacture a linearization for it.

As always, \textit{lin} will be forced to apply before the QP has merged into the larger structure. In this case, that will look like (130).

(130)  
```
(\[\text{VP} \quad \text{QP}_1\])
```

From (130), \textit{lin} will produce (131).

(131)  
```
\text{\{said < that \quad that < D \quad D < one \quad one < T \quad T < arrived\}}
```

```
\text{\{said < D \quad that < one \quad D < T \quad one < arrived\}}
```

```
\text{\{said < one \quad that < T \quad D < arrived\}}
```

```
\text{\{said < T \quad that < arrived\}}
```

```
\text{\{said < arrived\}}
```

Notice that because the relative clause is not yet inside VP, it is not included in the string associated with VP. As a consequence, only $\text{lin}(\text{QP})$ has information about where the relative clause will be positioned: it will follow everything else in QP. It is only after QP has merged with VP — to form (129) — that $\text{lin}$ can order the material in the relative clause with the material in the VP. In order to satisfy Totality, $\text{lin}$ will therefore have to apply to this material after (129) is formed.

The ordered pairs that this second run of $\text{lin}$ will add to (131) must not only satisfy Totality, but they must also satisfy Antisymmetry and Contiguity. The best satisfaction of Contiguity will be linearizations that keep the material in the relative clause together and put it either all before the VP, or all after the VP. If $\text{lin}$ puts the relative clause before the VP, however, the ordered pairs it will generate will include (132).

(132) you < arrived and did < arrived

It will also have to include:

(133) every < you, every < did

That's necessary because every ordered pair that $\forall$ is in will have to be matched by a parallel ordered pair that $\text{every}$ is in. Transitivity requires that the ordering also have:

(134) every < arrived

But this will cause an Antisymmetry violation, since (131) contains:

(135) arrived < D

And every must be in an ordered pair that matches every ordered pair D is in. This derives:

(136) When material in the higher position of a QR'd phrase is linearized there, it cannot be to the left of the lower position.

For these reasons, the ordered pairs $\text{lin}$ will generate when applied to (129) will add to (131) those in (137): ones in which the relative clause follows the VP.

(137) The new outputs from $\text{lin}(\text{129}) =$

$$\begin{align*}
\forall < \text{one} & \quad \text{one} < \text{that} & \quad \text{that} < \text{you} & \quad \text{you} < \text{did} \\
\forall < \text{that} & \quad \text{one} < \text{you} & \quad \text{that} < \text{did} \\
\forall < \text{you} & \quad \text{one} < \text{did} \\
\forall < \text{did} & \quad \text{that} < \text{every} & \quad \text{every} < \text{did} & \quad \text{every} < \text{arrived} & \quad \text{arrived} < \text{every}
\end{align*}$$
When this is added to (131), it will satisfy all of the constraints. It corresponds to the string in (138).

(138) …said that everyone arrived that you did

For the relative clause to be positioned outside the VP that is serving as the antecedent for the ellipsis it contains, it will necessarily be positioned outside, and linearly to the right of, the string that corresponds to that VP. This result is perfectly general. We derive (102): QR cannot put spoken material in a position where it is not semantically interpreted. We’ve derived (107).

(107) When Wh-moved material is spelled out in its higher position, it shows up to the left of the phrase it is merged to. When QRd is spelled out in its higher position, it shows up to the right of the phrase it is merged to.

This way of deriving the rightwards direction of QR and, consequently, extraposition has another useful consequence. It correctly captures the fact that in English, the only material that can extrapose out of nominals is material that can be linearized at the right edge of those nominals. Extraposition cannot, for instance, form (/one/three/nine) from (/one/three/nine).

(139) a. * I met every student yesterday new.
   b. I met every new student yesterday.

(139a) would get the representation in (140).

As in every example of QR, lin will run before the QP has merged into the larger structure; in this case, then, lin will apply before QP and VP have merged. After fusion occurs, lin will produce from QP the ordered pairs that correspond to the string in (141), and from VP, it will produce ordered pairs that correspond to the string in (142).

(141) every new student

(142) met every student yesterday

After QP has merged with the VP to form (140), lin will have to run again in order to ensure that Totality is obeyed. As in the case of an extrapoosed relative
clause, this latter run of lin has two choices: it may put new after the VP or before the VP. Unlike in the case of a relative clause, however, both of these choices violate Antisymmetry. If new is positioned after the VP, then it will be positioned after student, and that is at odds with the previous run of lin, which positions new before student. If, instead, the choice is to put new before the VP, this will cause it to precede every, and again this will conflict with what the previous run of lin did. This effect arises for any material that gets linearized between the determiner and whatever shows up at the end of the DP. Only material at the end of a DP can be extraposed.

This scheme solves another problem with the standard, copy theory based, view of QR. Consider an example like (143).

(143) She bought every book about a different child yesterday that I had fathered.
≠ She bought every book about [a different child that I had fathered] yesterday.

Under a copy interpretation of movement, this could get a representation like (144).

This representation would give the sentence an interpretation it cannot have: one in which the relative clause modifies child. What has happened here is that the relative clause merges not to the NP headed by book, but instead to the NP headed by child.

This problem doesn't arise on the multidominant interpretation of movement and the linearization scheme I've proposed here, however. Consider the representation that would be delivered by our multidominant representations for QR applying in (143).
We’ll run the linearization algorithm at the step in the derivation that immediately precedes merging QP with TP. That is required in order to meet the requirements on fusion. When lin runs on that structure, however, it will put all of the material belonging to the QR’d DP in its lowest position, as we’ve seen. As a consequence, this structure will not produce a string in which the relative clause, associated here with the embedded DP, is extraposed.

The only way the relative clause associated with the embedded NP can be linearized to the right of *yesterday* is if that NP QRs. If that were possible, we’d get a representation that has features of those in (146).
But for independent reasons, the *a different* nominal cannot QR out of the scope of *every*, and so this representation is prevented.
Successive Cyclic Movement

In the last several lectures, we’ve taken a close look at two kinds of movement. One is Wh Movement, which I’ve argued involves putting a DP in two positions, as (147) reminds.

(147) Which philosopher’s book about her should every linguist forget?

And the second is QR, which causes an NP to be shared between a Determiner and a Quantifier, as (148) reminds.

(148) A student read every paper yesterday.
∀\(x\). paper(x) = 1 \rightarrow a\_student\_read\_yesterday(x) = 1

\[\lambda x. \text{paper}(x) = 1 \rightarrow a\_student\_read\_yesterday(x) = 1\]

\[\lambda q. \forall x. \text{paper}(x) = 1 \rightarrow q(x) = 1\]

Presupposition: 2 is a paper

In the Wh Movement example, there is a determiner that contains a function and causes the phrases containing it to have a meaning that ranges over different values for that function. The Q morpheme that marks the scope of the question converts those alternatives into the semantic value for the question. Because the Q is in an Agreement relationship with the D, it must merge to a reasonably small phrase containing the Determiner, and it’s this phrase that “moves.”

In the QR example, the Q and D must both semantically combine with the same NP. Their selectional requirements force them to combine with an NP, and because they must do so in such a way that allows them to fuse, they will merge to the same NP. They must fuse because English only has vocabulary items that correspond to the D+Q combination and not to Q alone. Because of the constraints on the fusion operation, the method by which these structures are linearized will force QR’d phrases to be spoken in their lowest position.

What I’ve tried to illustrate with these two cases is how “movement” can be seen as the result of demands made by the morphology (Agreement and fusion) in concert with the semantics and the linearization algorithm. These demands can only be met — I claim — by allowing one term to remerge, and that gives us movement.

What I’d like to look at now is how these two kinds of movement can be combined. They have an interesting interaction.

I’ll start by pointing to a difference in the lower DPs of questions and the lower DPs of QR. In the case of questions, the determiner of the lower DP introduces alternatives. These alternatives propagate up through the phrase
marker and are used, finally, by the Q operator to form a question. There is no binding, then, between the D and Q in questions. But there is in the DPs that QR. The quantifier in the higher position binds the variable introduced by the determiner in the lower position.

(149) The Q in a QRd phrase binds the variable in the lower DP. The Q in a Wh moved phrase does not bind the lower DP, but instead operates on the meaning of the clause it combines with.

This difference in QR and Wh Movement plays an important role in cases of successive cyclic derivations. The evidence for these derivations is strongest in the case of Wh Movement, and so my discussion will be confined to cases of Wh Movement. Successive cyclic derivations are ones in which two sequential applications of movement apply to reposition a term. The evidence for these derivations came first from a consideration of Wh Movement out of embedded clauses, which can occur with the two operations of movement indicated in (150).

(150) Which book did she say [CP that you should read]?

The proposal I will make is that the first movements in a successive cyclic derivation of Wh Movement are in fact QR. Only the last movement is Wh Movement.

There are two reasons for this. One is that it gives us an account for why overt Wh Movement in English does not show Beck effects. If the syntax for sentences like (151) is as indicated, then we should expect these sentences to have the same degraded status that they do in certain wh in situ languages.

(151) Which person should only Minsu see?
Beck’s account of Beck effects is that focus sensitive operators, like only, use the focus semantic values of the things they combine with, and this disturbs the use of these focus semantic values that a higher question morpheme makes. The syntax in (151) predicts that this problem should arise in English questions as well, but it doesn’t.

However if we let this sentence have a successive cyclic derivation, and allow the first step in such a derivation to be QR, we’ll have a representation like (152).

(152) Which person should only Minsu see?
In this example, there is a DP headed by a determiner that will get spelled out as *which*, under Agreement with Q. These are the DPs that we’ve seen before function as the “traces” of Wh Movement. In this example, however, this DP shares its NP with another DP, DP’, whose head also introduces variables over functions. These are the kinds of DPs that we’ve seen functioning as the traces of QR. So we must let the DPs that are headed by *which* be among those things that can QR.

Notice that because the higher determiner is getting matched to a vocabulary item by way of Agreement with Q, it need not fuse with the determiner of the lower DP. One of the consequences of fusion is that the QRd phrase must get spelled out in its lower position. When fusion is lifted, as it is in this example, so also is that consequence for where the expression gets spelled out. As a result, where the *which* phrase gets linearized will be determined by what the language particular component says about wh-phrases in general. For English, this means it will show up at the left edge of the question.

Adopting this view of successive cyclic movement, then, requires that whether fusion is forced or not hinges on the nature of Q. It also means that we must let the English lexicon include an unpronounced determiner that is functioning as the “trace” in this example. We can’t let fusion be forced, in other words, by the need to spell out the determiner in the lower copy of QR. So, we’ll need a system like that in (153).

(153) **Principle of Full Interpretation**

Every terminal in $\phi$ must have an exponent in Spell Out ($\phi$).

a. Spell Out $\left( \begin{array}{c} D \\ \left[ \text{the} \right] n \end{array} \right) = \text{silence}$
b. Spell Out $(\forall) = \emptyset$

c. Spell Out \[
\begin{align*}
\forall + \begin{array}{c}
\land
\end{array}
\end{align*}
\] = every

d. Spell Out $(D_{\text{wh}}) = \text{which}$

That QR must be covert — that is, invoke fusion and all its consequences — whereas Wh Movement can have an “overt” step devolves to how the language has decided to map its determiners and quantifiers onto lexical items.

This gives us a way of modeling the “exceptions” to terseness that arise when DPs move. Recall that in cases of verb movement, terseness is lifted in cases where the verb makes a different word in both its positions. This allows that verb to show up twice in the resulting string. That doesn’t happen with DP movement, however. When a DP moves, we don’t see the noun, say, showing up in the two positions in the string, each of which are the result of its two syntactic positions. We can see this as the result of how lin works. It only allows a moved word to be linearized in two different positions if it is inside a different word in those positions, and phrasal movement can’t achieve that. Phrases can’t be parts of words. What does happen, however, is that sometimes the lower position of a moved DP appears in the string as a pronoun. This happens most often in relative clauses. If we think of pronouns as intransitive determiners, this syntax allows for a way of understanding that. A relative clause might have the structure in (154).

(154)  the girl that you visited

This would be the syntax for a “head raising” analysis of relative clauses. In such examples, we have to let there be an operator associated with the complementizer that binds the variable left by movement. I’ve represented that with
an index on the relative operator. The key idea in the head raising analysis of
relative clauses is that the NP part of the moved expression is also what the
relative clause modifies. The representation shown here could achieve that.
This is another instance where the NP is semantically interpreted in both of
its positions, but the linearization scheme allows it to show up in the string by
virtue of only one of these positions. In English the determiner in the lower
position is silent. But it’s perfectly possible that in other languages, it would be
possible for that determiner to be overtly pronounced. If it were, no problem
for linearization would emerge. When determiners are pronounced alone,
they can be expressed as pronouns. This might be how we get resumptive
pronouns.

The structure in (152) does not lead us to expect Beck effects. The relation-
ship between DP3 and DP' is one of variable binding, and that can happen
over focus sensitive operators like only. The focus semantic values introduced
by which don’t commence until a larger portion of the phrase marker is en-
countered. As a consequence, there are no focus sensitive operators between
Q and the DP headed by which, and therefore no Beck effect is expected.

That is one reason for believing that the first step in a successive cyclic
derivation of Wh Movement is QR. Another reason has to do with the ability
of Wh Movement to bleed disjoint reference effects. (155) is an example of that
sort.

(155) Which picture behind Sam does he dislike?

It is possible to understand Sam and he to corefer in this example. Without
the representations made available by successive cyclic derivations, this is
unexpected. (155) should get the representation in (156), which places Sam
within the scope of he.

(156)
That should make (155) parallel to (157) with regard to coreference, and that is wrong.

(157) * He₁ dislikes a picture behind Sam₁.

The standard solution to examples like (155) is to let the PP containing Sam “late merge” into a higher NP, and therefore not be in the lower NP where it would incur a disjoint reference effect.¹⁰ If we are to adopt that solution, it means that there must be a higher position in which this PP can be semantically interpreted. That higher position cannot be the merged-to-Q position, however, since the question Q does not semantically combine with the phrase it has merged to. Without successive cyclic movement, our representations only offer one position for the material within a Wh-moved phrase to be interpreted, and that is the lowest position. But as these examples show, there needs to be some higher position that the material can be interpreted in.

If successive cyclic movement can involve QR steps, then those needed higher positions can be produced. Letting QR feed Wh movement in (155) will produce a representation like (158).

(158)

Presupposition: 3 is a picture

Presupposition: The output of f is a picture behind Sam

QR merges the DP which picture behind Sam to TP in this example, and this DP binds the picture. Because the higher DP is outside the scope of he, Sam and he can corefer.

This technique allows us a means of expressing the full gamut of reconstruction type effects. One that requires some thought, however, is what Barss (1986) calls the “multiple binding domain effect,” illustrated by (159).

¹⁰ The idea goes back to Lebeaux (1988).
(159)  a. * Sally₁ said that you bought a book about herself₁.

   b. Which book about herself₁ did Sally₁ say that you bought?

The representation we would give to (159b) is (160).

(160)

In this representation, book about herself is semantically interpreted in two positions. The higher of its two semantically interpreted positions puts it close enough to Sally for herself to be bound by Sally and still meet the locality conditions of Principle A. We need to formulate the locality condition that governs how far a reflexive can be from its binder in a way that is sensitive to the possibility of an anaphor having more than one position. So, for instance, something like (161) won’t work correctly in this case.

(161) A reflexive must be bound by something that is in its binding domain. A binding domain for α is the smallest TP that dominates α. X is smaller than Y if Y dominates X.

Successive Cyclic Movement 67
The multidominant trees make it hard to define what's needed here. One possibility is to leave everything of the traditional binding domain definition intact, and redefine “smallest TP that dominates α.”

(162) Let \( P(\alpha) = \{ n_1, n_2, \ldots, n_m \} \) be a series of nodes, such that each \( n_{i+1} \) immediately dominates \( n_i \), \( n_1 \) immediately dominates \( \alpha \), and \( n_m \) is a root. The smallest \( \beta \) dominating \( \alpha \) is that \( n_i \) with the smallest \( i \) which is in every \( P(\alpha) \).

It is thought that Principle C effect cannot always be overcome by Wh movement. There are certain circumstances where a name in a Wh moved item behaves as if it must be semantically interpreted within the scope of something that wh-phrase has moved past. The technique for controlling reconstruction effects here can capture these cases in a way that is parallel to the standard account.\(^1\)\(^1\) Consider, for instance, (163).

(163) * Whose looking up Sally, did she talk about?

\[ \text{(Assume that behind \textit{whose} is \textit{which of one's}.)} \]

The expression \textit{whose looking up Sally} is a nominal gerund in English that mixes nominal and verbal syntax. It's not found (I have discovered) in everyone's English, but it's described in Wasow and Roeper (1972) and Baker (1985) and it exists in my English. Roughly speaking, what happens is that a VP is nominalized and this allows that VP to exist within a DP.

The parse I've given in (163) would be one that avoids a Principle C effect by placing \textit{Sally} in a position where it is semantically interpreted outside of the scope of \textit{she}. We've “late merged” \textit{Sally}. The transitivity of \textit{looking up}, however, prevents the V in the lowest position from getting an interpretation, and so this representation is blocked. This is why I chose this particular kind
of gerund expression. Verbs, but not nouns, can be obligatorily transitive in English and that is what is responsible for blocking representations like (163).

Instead, this sentence must get the representation in (164), and this triggers a Principle C effect.

![Diagram](164)

So, if this is right, one component of the contrast between cases were Principle C is triggered by reconstruction and cases where it’s not is the transitivity of the head noun or verb involved. This, unfortunately, is a very slippery area, as most nouns are not obligatorily transitive. To trap them into their transitive meaning can be done (I think) by choosing deverbal nouns, supplying them with a subject argument and fixing on their process meaning. (Those ingredients are the ones recommended by Grimshaw (1990).) Let me briefly sketch out some of these factors.

Consider the following paradigm.

(165) a. The (many) examinations stunk.
    b. John’s (many) examinations stunk.
    c. The (many) examinations of Steve stunk.
    d. John’s (many) examinations of Steve stunk.

(166) a. The (lengthy) examination wasn’t well attended.
    b. John’s (lengthy) examination wasn’t well attended.
    c. The (lengthy) examination of Steve wasn’t well attended.
    d. John’s (lengthy) examination of Steve wasn’t well attended.

Consider first the b-examples. In (166b), John must be the object argument of examination, whereas in (165b), John can be understood as the subject (or possessor) argument. We learn from the pair (166c) and (166d), that subject
arguments of nouns are always optional. We can explain why (166b) does not allow John to have the subject argument role if examination, when it is argument taking, must take an object. (166b) would have the same status as (167).

(167) * John examined throughout the week.

If the same examination is present in (166a) that is present in (166b), then (166a) should also be ungrammatical. I think examination has taken on an eventive/process reading that is non-argument taking. I speculate that examination, then, has three meanings: a result meaning (= (165)), a meaning parallel to the verb evaluate (= (166b-d)), and a non-argument taking, eventive reading (= (166a)). I think parallel things could be said about promotion.

Consider now a question that involves movement of a DP with a noun inside, rather than a verb.

(168) * Whose examination of Sally, did she talk about?

We can see that the version of examination in the lower position must be the one that does not take arguments. But the one that is found in the higher position is the one that takes arguments. These are not the same noun, though, and so this representation fails.

It may well be the general fuzziness of the distinction between these two kinds of nouns that is responsible for the chaotic judgements reported in the literature, and their general fuzziness. I find contrasts of the sort that suggests the result vs. process distinction is playing a role in our reconstruction examples as well.
(169)  a. * Which examination of Jane₁ was she₁ a part of?
       b. * Whose examination of Jane₁ does she₁ dislike?
       c. Which examination of Jane₁ does she₁ dislike?
(170)  a. * Whose kissing (of) Jane₁ does she₁ recall?
       b. * Which kissing (of) Jane₁ did she₁ witness?
       c. * Whose painting (of) Jeremy₁ blue did he₁ see?
       d. ?? Which paintings of Jeremy₁ did he₁ want framed?

In those examples where a process reading is forced, either by the predicate that is selecting the DP or by the internal make-up of the DP, we get a Principle C effect.

That successive cyclic Wh movement can involve steps that are QR has independent support from cases where the wh-phrase contains a quantifier. A well-studied example of that kind is found in how many questions, like (171).

(171)  How many examples of Wh movement will we have to endure?

There is in this question both the variable introduced by how and the variable introduced by many. The variable introduced by how ranges over amounts, or numerical degrees. The variable introduced by many ranges over individuals that are examples. The variable introduced by how gets resolved at the CP that has the question meaning. But the meaning introduced by many can be evaluated in different positions. Where many is evaluated will control where the NP is semantically interpreted as well. That is what we learn from examples like those in (172) and (173), introduced by Heycock (1995) and discussed in Fox (1999).

(172)  a. * How many stories about Diana₁ is she₁ likely to invent?
       b. How many stories about Diana₁ is she₁ likely to reinvent?

       (fashioned after Fox, 1999, (19): 167)

(173)  a. * How many houses in John's₁ city does he₁ think you should build?
       b. How many houses in John's₁ city does he₁ think you should rebuild?

       (Fox, 1999, 20: 167)

The creation verbs invent and build favor an interpretation in which many stories is interpreted in the embedded clause. This forces Diana to be interpreted in this position, and a disjoint reference effect arises as a consequence.

Let's take a closer look at how the system here captures these facts.

We should begin by getting a fix on what the structure for a how many DP is. Let's start with the simpler, non-interrogative, cases of many DPs. Hackl (2000) has argued that these expressions involve degree phrases in which the degrees being measured are numerical amounts. That permits a unified analysis of the expressions in (174).
more is the comparative form of many: it means what manier would. We can see from (174b) that many can be part of a comparative construction then, and from (174c), we see that these comparatives can involve comparisons of amounts that aren't numerically expressible. A standard way of expressing the amounts that are involved in these expressions is with "degrees" on a scale that is invoked by the predicates involved. So our first step is, following Hackl, to take many to express an amount of a degree. In the case of simple expressions like (174a), those degrees can be thought of as simple numerical amounts. So the meaning we'll assign to (174a) will be something tantamount to "He ate an amount of kumquats that is many."

Here's a stab at that:

(175) He dodged that many questions.

(176) \[ \text{[that]} = \text{a contextually fixed degree} \]

(177) \[ \text{[many]} = \lambda d \lambda P \lambda x. \text{d-many}(x) \land P(x) = 1 \]

I've existentially closed off the x variable introduced by many at the DP level. I've represented that with "some." We can think of this as coming about by a existential closure operator that is inserted to fix free variables and shift the types, or we can think of this as being contributed by the "D" in this representation.

In how many questions, the how quantifies over the degree variable: d part. That is, the part that is occupied by that in (175). There is evidence that we should see how as being equivalent to which degree. That evidence is that how
many questions can be understood to seek functions as answers, just as we've seen which questions do.

(178) How many books should everyone read?
    A: 6
    A: more than her professor

The second of these two answers expresses a function, one that gives amounts of books that depend on the value given to everyone. What we're seeing here, then, is that the degree variable which the question determiner binds is parallel to the individual variable that which binds in simpler questions like (179).

(179) Which book should everyone forget?
    A: Movement in Language
    A: Her first

In the case of which-phrase, we built the variable up from a definite determiner and an “index” that contains the function.

(180) \[[\text{which book}]\] in trace position ≈
\[
\begin{array}{c}
\text{DP} \\
\text{D} \quad \text{NP} \\
\text{the} \quad \text{ } \quad \text{book}
\end{array}
\]

The meaning we require for (178) is something equivalent to “which degree many books should no one read?” So we're going to want to equate how with the “the degree” part.

(181) \[[\text{how}]\] in trace position ≈
\[
\begin{array}{c}
\text{DP} \\
\text{D} \quad \text{degree} \\
\text{the} \quad \text{ } \quad \text{degree}
\end{array}
\]

So, this means that the DP which moves in how many questions looks like (182).
Consider next what the highest “D” position in this phrase is. The D associated with these expressions can be indefinite, as in (183), and when it is it shows up after the DegP.

(183) How many a book (have you read?)

I’ll assume that English has a silent plural definite article as well. We might parse the indefinites as (184) indicates.

(184) We can leave the denotation I’ve given for many the same if we understand the indefinite determiner to be semantically vacuous.

These DPs provide us with a way of forming representations for how many questions that captures the contrasts in (172) and (173). In one representation, there has been no successive cyclic movement, and many is interpreted in the lower position.
Presupposition: The output of $f$ is a degree

This representation is the only one that will fit creation verbs like *invent*, as we'll see in a moment. As you can see, this representation puts *Diana* within the scope of *she*, and this is the reason we find a disjoint reference effect in the case where *invent* is used.

The other representation is one in which the *how many* phrase QRs first.

---

(185)
This representation doesn't fit well with creation verbs like `invent`. When we work out what (186) means, we'll see that `many` introduces an existential quantification, and the DP in object position is a variable bound by this existential quantification. So, this sentence is going to say that there exist some stories about Diana, and then adds that she is likely to invent them. But inventing stories that already exist is an anomalous thing to say. By contrast, when indefinites are interpreted in the object position of creation verbs, as in (185), they lose their existential force, and that is why (185) is not anomalous. Because this is the only representation that puts `Diana` outside the scope of `she`, it is the only representation in which `Diana` and `she` can corefer. This explains, then, the contrast in (171).

It is useful to compare these examples with ones in which `how` is part of a predicate, not a DP, as in questions like (187).

(187) How happy is Sally?

The idea about adjectives like `happy` is that they are relations between degrees and individuals. So a sentence like (188) has the representation in (189).

(188) Sally is happy.
(189)  
\[
\begin{array}{c}
\text{TP} \\
\text{DP} \quad \text{TP} \\
\Diamond \text{Sally} \quad \text{T} \quad \text{VP} \\
V \quad \text{Sally is } d\text{-degree happy} \\
\text{is} \\
\text{DP} \quad \lambda y. y \text{ is } d\text{-degree happy} \\
\Diamond \text{Sally} \\
\text{Deg} \quad \lambda x. \lambda y. y \text{ is } x\text{-degree happy} \\
\text{d} \\
\text{AP} \quad \text{happy} \\
\end{array}
\]

The *how* in (187) is the same one that we saw in *how many* questions. As in *how many* questions, questions like (187) can get a functional reading.

(190) How happy is every student?
A: Way more than her professor!

So (187) gets an analysis like that indicated in (191).

(191)  
\[
\begin{array}{c}
\text{CP} \\
\text{QP} \quad \text{CP} \\
\text{Q} \\
\text{C} \quad \text{TP} \\
\text{is} \\
\text{DP} \quad \text{TP} \\
\Diamond \text{Sally} \quad \text{T} \quad \text{VP} \\
\text{V} \quad \text{AP} \\
\text{how} \\
\text{D} \quad \text{degree} \quad \text{happy} \\
\text{the} \quad f \\
\end{array}
\]

Now, an interesting difference between examples like these, where *how* is part of an AP, and those like (171), in which *how* is part of a DP, is that in these AP cases there is no possibility of QR, and therefore no possibility of successive cyclic movement. If successive cyclic movement is not possible,
then everything in the moved phrase (except Q) must be interpreted in its lowest position. This predicts that Principle C effects should be unavoidable, and this is known to be the case.

(192)  a. * How satisfied by Sarah1's debate was she1?

       b. * How satisfied because of Sarah1's performance did she1 say you were?

(For a discussion of these cases, see Barss 1986, Huang 1993 and Takano 1995.)
The reason there is no possibility of QR is because the AP that is moved is not a definite description. The cases we have seen where a variable exists in the lower position are both cases in which a definite description is placed in this position. If the position a term moves from is not a position in which a definite description can be interpreted, then we will not have an example of movement that gets translated into a variable-binder relationship. Instead, these cases of movement will all be instances where the moved item behaves, semantically, as if it hadn't moved.

I think this provides us with a place to test the semantics for questions that I adopted. If these wh-APs, which cannot be QRd, are interpreted in their lowest position then we should find that they are incapable of overcoming the interrupting effect of focus sensitive operators. So we should find a contrast between the examples in (193).

(193)  a. How many pictures has only Minsu purchased?

       b. How happy is only Minsu?

It's not clear to me what the fact is here. The question in (193b) is more difficult than the one in (193a), but I'm not sure it's ungrammatical.
I’ve only looked at three movement types here: verb movement, Wh-movement and Quantifier Raising. There are, of course, many other kinds of movements and how they would, or could, fit into the picture I’m developing will have to wait. But I want to end by making a few observations about Argument Movement.

Argument Movement (=194) has a mix of properties that make it different from either QR or Wh movement, and which are difficult to capture.

(194) Smith seems to be available.

From examples like (195), we can conclude that Argument Movement can, like wh-movement, be semantically vacuous.

(195) A good solution doesn’t seem to be available.

≡ ¬ seem a good solution to be available

Our system would give this a representation like (196). (I’m taking some shortcuts here.)
Whatever requires A Movement in this context cannot be semantic. The availability of this interpretation of A Movement, however, seems to be contingent on the kind of quantifier there is in the D.

(197) Every good solution doesn’t seem to be available.

\[ \neg \exists \text{ good solution to be available.} \]

Perhaps this tracks contrasts like (198).

(198) a. There seems to be a good solution available.

\[ \exists \text{ good solution available.} \]

b. \( ?^* \) There seems to be every good solution available.

If we knew what was responsible for the contrast in (198), it could (perhaps) be used to prevent (197) from getting a representation parallel to (196).

It seems that the quantificational DP is interpreted where it is spoken in (197). The same option exists for indefinite subjects like (199).

(199) A solution doesn’t seem to be plausible.

\[ \exists \text{ solution that doesn’t seem plausible.} \]

Might these cases involve a structure which is an overt versions of QR?

(200)\[
\begin{array}{c}
\text{TP} \\
\text{QP}_2 \quad \text{TP} \\
\text{Q} \quad \text{TP} \\
\text{T} \quad \text{VP} \\
\forall \quad \text{V} \\
\text{seem} \quad \text{TP} \\
\text{T} \quad \text{VP} \\
\text{seem} \quad \text{V} \\
\text{to} \quad \text{AP} \\
\text{be} \quad \text{DP} \\
\text{be} \quad \text{AP} \\
\text{NP} \quad \text{D} \\
\text{good solution} \quad \text{the 2} \\
\end{array}
\]

Maybe not. Recall that the NP part of a QR’d phrase remains semantically interpreted in its lower position. This is why there is still a Principle C effect in examples like (201) even when the quantifier takes wide scope.

(201) A different student showed to him every problem with Smith.

\[ \text{him} \neq \text{Smith} \]

But this isn’t the case for A-movement.

(202) My kissing John₁ seems to him₁ to be too easily visualized.
I've chosen an example here where we can't make a derivation in which *John is is “late merged” in. That is, I've chosen an example in which the lowest NP must contain John. We can see that this is the case by looking at a parallel Wh-movement example:

\[(203) \quad \text{* Whose kissing John, does he visualize?} \]

In cases where the A-moved DP is interpreted in the position it is spoken, it seems we want the variable it binds in the lower position to be unrestricted. We'd have to produce something like (204).

\[(204) \quad \text{TP} \]
\[\text{QP}_2 \quad \text{TP} \]
\[\text{Q} \quad \text{T} \quad \text{VP} \]
\[\forall \quad \text{V} \quad \text{TP} \]
\[\text{seem} \quad \text{T} \quad \text{VP} \]
\[\text{to} \quad \text{V} \quad \text{AP} \]
\[\text{be DP} \quad \text{AP} \]
\[\text{NP} \quad \text{D available} \]
\[\text{good solution} \quad \text{the 2} \]

This can't be the same determiner that we have used in building QR because that one requires an NP to semantically combine with.\(^{12}\)

In fact, when the A-moved DP is interpreted in its higher position, there is little reason to believe that movement is involved. We could treat these cases as involving binding of a null pronoun, as in (205).

\(^{12}\) See Takahashi and Hulsey (2009) for an analysis which goes in this direction.
Maybe this is why the “traces” of A-movement, but not wh-movement, can be pronounced when there is a suitable Case marker.

(206)  
   a. The solutions seem to be all of them available.  
   b. * Which solutions did you make all of them available.

And we might be able to capture the other ways in which Argument movement seems to differ from Wh-Movement and QR. Indeed, I think a promising direction is to try to reduce Argument Movement to the “control” relation which exists between the subjects in examples such as (207).

(207) Smith₁ tried [CP PRO₁ to end].
Bibliography


Cable, Seth. 2008. There is no such thing as pied-piping. Unpublished paper, University of Massachusetts at Amherst.


