

March 2017

Embedded Late Merge and the Theory of Displacement¹

Danny Fox, MIT

Previous work has argued that extraposition of an NP modifier has consequences for the scope of the QP that takes NP as complement (Guéron and May 1984, Fox and Nissenbaum 1999, Fox 2002). Fox and Nissenbaum account for the relevant observations under the assumption that extraposition results from a specific interaction between covert movement (Quantifier Raising) and the operation of Late Merge (LM) proposed by David Lebeux (Lebeaux 1988).

In this paper I argue that the facts are more involved than had previously been assumed. The relevant complications arise under two circumstances: (a) when no QP takes the modified NP as complement and (b) when the modified NP is dominated by two different QPs. In such cases, I argue that the normally attested consequences for scope can in principle hold for any dominating QP. I point out that this observation can be derived if LM can apply in embedded positions. When such embedded LM is combined with the view that the tail of a movement chain is interpreted as a restricted variable (“Trace Conversion”, TC) new predictions are made which I attempt to corroborate.

If arguments for LM, and in particular for embedded LM, are successful, the precise theoretical consequences are not entirely clear. In particular, one might ask whether LM per se needs to be postulated or whether there is another way of deriving the required structures. More specifically, we might ask whether the arguments I will be making are consistent with a multi-dominance view of movement (see Bachrach and Katzir 2009, Johnson 2012, 2014 and Sportiche 2015).

Assume that there is no movement rule and that a single operation of merge yields multi-dominance structures – structures in which a single constituent might have more than one mother. In other words, assume (along the lines of much recent literature)² that multi-dominance can be derived not only by a rule of Internal Merge and is therefore not limited to cases where one mother of a multi-dominated phrase dominates all other mothers (cases where one position c-commands all others). Such an assumption might license the derivation of the necessary syntactic representations (LM structures) without a counter-cyclic operation of late merge. However, this comes with an obvious cost: the constraints on displacement (e.g. locality) cannot be linked directly to movement (Internal Merge), since there will be no specific rule responsible for displacement. We, thus, cannot say that the source of locality follows from requirements on Internal Merge.

¹ For valuable comments and questions, I would like to thank Noam Chomsky, Luka Crnić, Kyle Johnson, Roni Katzir, Jon Nissenbaum, Chris O’Brien, David Pesetsky, Ivy Sichel and Dominique Sportiche.

² See Citko (2005), de Vries (2005) and references therein.

(For example, it would not suffice to say that an object cannot be Internally Merged, unless it is made visible by agreement.)

I, therefore, suggest (in the spirit of Johnson 2012, 2014) that the constraints on displacement follow more indirectly. When multi-dominance arises (when a single constituent has more than one position in a linguistic representation), I will adopt what I think is a rather simple assumption, namely that linearization (which is required for spell-out) fails, unless a specific device is in place which tells the phonology which of the two positions occupied by a single constituent is to be ignored or deleted (Nunes 2001, 2004, Johnson 2012, 2014).³ I will add to this the assumption that phonological deletion (of the relevant sort) is parasitic on agreement, which is, in turn, conditioned by locality. So movement does not exist as a special rule, and, hence, LM does not exist. Multi-dominance, however, does exist, and subsequently the structures that I will be arguing for can be derived. Merge creates many multi-dominance structures that are not attested. As usual, the hope is that this problem of over-generation can be dealt with by independently motivated constraints (in particular, TC in semantics, linearization in phonology and probably a syntactic residue, e.g. the case filter).

Sections 1-5 deal with extraposition from DPs that contain non-restrictive adjectives. It begins with an observation made in Culicover and Rochemont (1990) that, at first glance, suggests that embedded LM of the sort I will be arguing for is not available. I will present evidence that this interpretation of the data is incorrect. Specifically, I will argue that the basic observation is predicted if embedded LM can apply freely as long as traces are interpreted by TC. In sections 6-13 I will be examining the predicted interaction between embedded LM and TC for constructions in which the NP modified by the extraposed material is dominated by more than one QP. I will be arguing that, as long as the final structure is interpretable by TC, the attested word order can be derived by QR of any one of the dominating QPs (followed, of course, by embedded LM). Finally, in section 14-15, I will present a way of deriving the required structures by a single cyclic rule of Merge, i.e. without reference to the counter-cyclic rule of late merge.

1. Putative Locality

Culicover and Rochemont (1990, C&R) argue that extraposition is subject to a very specific locality condition, one that allows a modifier to extrapose from NP only if it modifies NP itself (i.e. is daughter of NP rather than of some nominal constituent properly dominated by NP). To understand the empirical claim, I will take a close look at an example discussed in Johnson (1991).⁴ Consider first the two interpretations of the sentence in (1) which are derived from two possible structures for the NP *alleged mouse from Mars*.

- (1) John saw an alleged mouse from Mars (yesterday).

³ As we will see, there might be specific circumstances where this does not hold. See note 30

⁴ For additional examples, see Culicover and Rochemont (1990), p. 63.

- (1') a. John saw an $[_{NP} \text{alleged} [[\text{mouse}] [\text{from Mars}]]]$
 $\exists x [\text{alleged}[\lambda w. \text{mouse}(w, x) \ \& \ \text{from-Mars}(w, x)] \ \& \ \text{J. saw } x]$ (*alleged* > &)
 b. John saw an $[_{NP} [\text{alleged mouse}] [\text{from Mars}]]$
 $\exists x [\text{alleged}[\lambda w. \text{mouse}(w, x)] \ \& \ \text{from-Mars}(x)] \ \& \ \text{J. saw } x]$ (& > *alleged*)

Under one interpretation, (1'a), the sentence involves existential quantification over individuals about which there is a particular allegation, namely that they are mice from Mars. This interpretation results from a structure in which *from Mars* is conjoined (by predicate modification) with the noun *mouse* forming the argument of *alleged*. Hence we will characterize the interpretation as one resulting from a structure in which *alleged* takes scope over conjunction/modification (*alleged* > &). Under the alternative interpretation, (1'b), the relevant allegation pertains only to mousehood. Under this interpretation, the sentence involves existential quantification over individuals who meet two requirements: (a) they must be alleged mice and (b) they must (in actuality) be from Mars. In other words, the second interpretation results from a structure in which *from Mars* is conjoined with the phrase *alleged mouse* (& > *alleged*).

C&R's locality claim is that extraposition of the modifier, as in (2), is only possible when this modifier is sister of the phrase *alleged mouse*, i.e., when the NP structure is the one in (1'b). The precise statement of the constraint depends, of course, on the theoretical perspective one takes on extraposition. If, for example, the extraposed PP is moved to the right, the constraint might be stated as a requirement that rightward movement not cross two nominal nodes. If, alternatively, extraposition involves LM of a PP to a QR-ed phrase (as in Fox and Nissenbaum 1999), the constraint might restrict the depth of embedding for LM (as in Sauerland 1998 and Tada 1993), e.g. it might state that an LM-ed phrase must modify the NP complement of the QR-ed phrase (that is, rule out what I've called embedded LM). And, of course, there are other possibilities (e.g. that of C&R, or Johnson 1991).

- (2) John saw an alleged mouse yesterday from Mars.
 a. ~~$\exists x [\text{alleged}[\lambda w. \text{mouse}(w, x) \ \& \ \text{from-Mars}(w, x)] \ \& \ \text{J. saw } x]$~~ (*alleged* > &)
 b. $\exists x [\text{alleged}[\lambda w. \text{mouse}(w, x)] \ \& \ \text{from-Mars}(x)] \ \& \ \text{J. saw } x]$ (& > *alleged*)

Paraphrase of only available meaning (if locality is correct): *There is something John saw which is alleged to be a mouse and in reality is from Mars.*

Be that as it may, the empirical argument is based on an interpretive contrast between (1) and (2). Specifically, it is observed that (2), in contrast to (1), is not judged to be true when there is just one individual that John saw such that this individual is presupposed to be a mouse and there is some skepticism about an allegation that this mouse is from Mars. Since this state of affairs verifies (1'a) and not (1'b), the observation supports the locality claim.⁵

⁵ Another way to argue for C&R's claim (suggested to me by Jeremy Hartman) is to consider a noun which, on its own, would normally not provide the content of any controversial claim about an individual in

2. Pesetsky's Empirical Claim

David Pesetsky (p.c. spring 2000) has objected to this argument, pointing out that, although the sentence in (2) cannot be true in the characterized situation, it is wrong to claim that the reading in (2)b is the only one available. Specifically, he claimed that conjunction/modification can be interpreted in the scope of *alleged*, as long as mousehood is alleged rather than presupposed.

Although, at the time, I was suspicious of the judgment, I think that the sentences in (3) and (4) can serve to verify Pesetsky's intuition. If the locality claim were correct, these sentences should only receive an anomalous interpretation – contrary to the judgment of many speakers I've consulted.

- (3) John saw an alleged alien yesterday from Mars.

Predicted meaning:

$\exists x$ alleged $[\lambda w.$ alien(w, x)] & from-Mars(x) & J. saw x (allege & from Mars)
#There is something John saw which is alleged to be an alien and in reality is from Mars.

- (4) a. John saw an alleged New Yorker yesterday from the Upper East Side.
 b. John saw an alleged antique yesterday from the 15th century.

In (3-4), the extraposed PP entails the head noun, at least given presupposed background assumptions (*from Mars* \Rightarrow *alien*; *from the upper east side* \Rightarrow *New Yorker*; *from the 15th century* \Rightarrow *antique*). If conjunction had scope over *alleged* (as C&R claim), we would require of the relevant individual that it be in the (actual world) extension of the predicate denoted by the PP modifier. But then, by entailment, it would also have to be in the (actual world) extension of the predicate denoted by the head noun, and use of the adjective *alleged* would be inappropriate (since *alleged P* cannot be appropriately predicated of an individual x if the speaker believes that P is true of x). A speaker,

the domain of quantification. When such a noun is in the scope of the adjective *alleged*, the scope would (at least under normal circumstances) have to contain additional material, and if C&R are right, this material would not be able to extrapose. Candidates for such nouns suggested by Hartman are *person*, *area*, and *thing*:

- (i) a. # He alleged that John is a person
 b. #I met an alleged person yesterday.

Subsequently C&R predict the following contrast:

- (ii) a. I met an alleged person from France yesterday.
 b. #I met an alleged person yesterday from France.

I trust that the reader will be able to see that the contrast would also be predicted by the alternative to C&R that I will be arguing for.

uttering (3), for example, would be attributing two contradictory properties to the same individual: (a) that of being from Mars, and (b) that of being an alleged alien (and hence, at least by implicature, not a real alien).⁶ (3), however, is a perfectly natural sentence in which the speaker raises skepticism about an allegation that the person John saw is an alien and more specifically about the details of this allegation according to which this alien is from Mars.⁷

In other words, if we were to characterize the effect of extraposition on meaning in (2) as one resulting from locality, we would be making wrong predictions for (3) and (4). We conclude that C&R are right that extraposition has an effect on meaning in cases such as (1), but that this effect has been incorrectly characterized. Our task, then, is to provide an appropriate characterization and see whether it can give us a hint as to the roots of the effect.

3. Embedded Late Merge and the Meaning of Traces

The facts observed in (2-4) appear to be rather puzzling. From (2) we learn that the domain of quantification for an object QP of the form *an alleged N...*, from which a modifier is extraposed, must be made up of individuals that are not presupposed to be in the extension of N (*mouse* in (2)). We might state this as a conclusion that the predicate denoted by *alleged N* must hold of all individuals in the domain of quantification, in

⁶ From now on, I will be talking (at times) as if the inference that the nominal argument of *alleged* is not true of the relevant individual follows from the basic meaning of the NP, i.e. as if *alleged* takes a predicate *P* and returns a predicate that is false of individuals that *P* is true of:

(i) $\llbracket \text{alleged} \rrbracket^w(P) = \lambda x. P(x)(w) = 0 \ \& \ \lambda w'. P(x)(w')$ is an allegation in *w*.

But the same facts would follow in essentially the same way if we eliminate the first conjunct, $P(x)(w) = 0$, and derive an inference that *P* is false of the relevant individual (or at least that one might think it is false) more indirectly (e.g. as a scalar implicature). In section 9, we will see considerations that might prefer such an indirect derivation (see note 23).

⁷ The judgments reported here should also be tested with other non-intersective adjectives. Jeremy Hartman (p.c.) points out a class of adjectives that are non-intersective, yet differ from *alleged* in entailing the truth of their complement: *undisputed/proven/confirmed/acknowledged/obvious/known*. These adjectives show a similar effect, which is likewise predicted by the current account, though explaining this would require a more nuanced discussion. So, for example, *John saw an undisputed mouse yesterday from Mars*, strongly prefers an *undisputed* > & parse, whereas this preference disappears in *John saw an undisputed alien yesterday from Mars*. I think the account we propose in the paper will extend to such cases, the moment we are precise enough about the semantic properties of *undisputed*, those that make it a non-intersective adjective in the first place. For example, for *x* to be called an undisputed alien, it cannot be common ground that *x* is an alien. Similar negative inferences follow for all of the adjectives listed above, and these inferences suffice for the account to be extended.

As for other non-intersective adjective, Mitcho Erlewine points out that *former* shows a similar effect to the one described for *alleged*, but that *fake* behaves differently, a contrast that I will have to leave to another occasion. Roni Katzir suggests that this contrast might be related to the arguments made in Partee (2009) that *fake* is a subjective adjective. See note 23 for another possibility.

(i) a. I interviewed a former New Yorker yesterday from Brooklyn and asked her why she moved away.
b. #John was arrested for selling fake antiques yesterday from the 15th century.

which case it will be teaching us that conjunction has scope over *alleged*. From (3) and (4), however, we learn that the extraposed modifier (*from Mars* in (3)) need not hold of individuals in the domain of quantification. This leads to the opposite conclusion, namely that conjunction is in the scope of *alleged*. This apparent contradiction is restated in (5).

- (5) **Let β be a modifier “extraposed” from the QP *an alleged N*. And let x be an individual in the restrictor of QP.⁸**

Property 1 – inferred from (1)b:

$\llbracket \text{alleged} \rrbracket (\llbracket N \rrbracket)$ must be true of x . Hence β (and conjunction) cannot be in the scope of *alleged*.

Property 2 – inferred from (3) and (4)

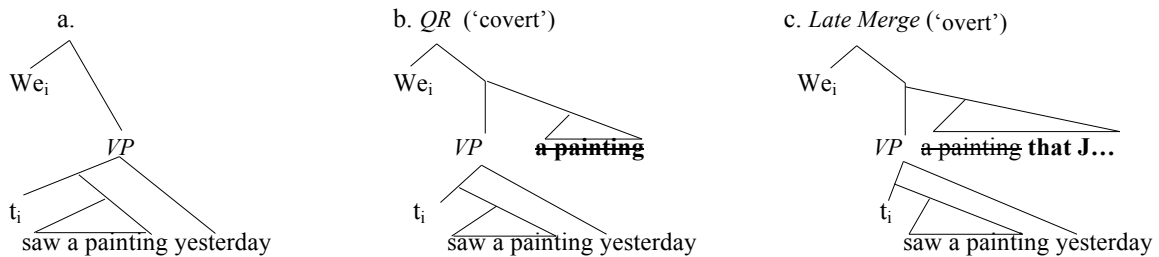
$\llbracket \beta \rrbracket$ need not be true of x . Hence, β (and conjunction) can be in the scope of *alleged*.

Property 1 and 2 cannot both be satisfied, unless there are two occurrences of *alleged* in the structure each of which satisfies one but not the other property. I would, therefore, like to take the data to argue for this conclusion, which, in turn, follows from – and thus supports – the analysis of extraposition argued for in Fox and Nissenbaum (1999) and Fox (2002). [See also Bhatt and Pancheva 2004, Fox and Johnson 2015, Hulsey and Sauerland 2006, and Johnson 2012.]

3.1. Extraposition, Late Merge, and Trace Conversion

Fox and Nissenbaum (1999), following an earlier proposal by Guéron and May (1984) and a related proposal in Reinhart (1991), argue that extraposition of NP modifiers involves Quantifier Raising (QR) in its derivation. Specifically, they propose that the derivation involves QR followed by LM of an adjunct (proposed by David Lebeaux for *wh*-movement).⁹

- (6) We saw a painting yesterday that John talked about.



⁸ e.g., an individual in the set denoted by the NP argument of the existential quantifier.

⁹ Nothing in this paper relies on specific claims about the nature of the derivation. What is important, in the context of this paper, is the final LF representation. For different views about the nature of the derivation which agree with Fox and Nissenbaum (1999) about the relevant properties of the LF representations, see Johnson 2012 and Sportiche 2015. See also section 14-15 below.

talked about are paintings. The arguments for TC have therefore been rather indirect, coming from Binding Theory, the parallelism conditions on ellipsis, focus, reconstruction, and the explanation of conservativity (Bhatt and Pancheva 2004, Chomsky 1993, Erlewine 2014, Fox 1999, 2002, Heim 2012, Romoli 2015, Sauerland 1998, 2001, 2004).

Before we move on, it is useful to point out that nothing in what I said relies crucially on the nature of derivations. What is crucial is that there be a way to derive LF representations in which part (but not all) of a QPs restrictor appears in the trace position and that extraposition tells us which part this is (the part which has not been extraposed). The claim that the derivation involves LM is not crucial, but, given its familiarity, it is a useful way to communicate the relevant property of the final presentation. As stated in the introduction, at the end of the paper, I will entertain the possibility that the LM representations are derived without late merge, i.e. with just one cyclic operation of Merge that can create multi-dominance representations (which are phonologically uninterpretable, unless PF deletion applies to designated occurrences of multi-dominated phrases).

3.2. An Account of C&R's Observation

The interpretive effects of extraposition observed in (2-4) are predicted by the account of extraposition outlined above, as long as there are no relevant locality conditions on LM (i.e. as long as embedded LM is possible). Specifically, if an extraposed PP can LM either inside or outside the scope of *alleged*, we have (after TC) the following two structures (with the irrelevant adverb *yesterday* omitted).

- (10)a. [An alleged [N PP]]
 λx John saw [the_x alleged N]
 b. [An [[alleged N] PP]]
 λx John saw [the_x alleged N]

Since (10a) is a possible structure in which conjunction/modification has scope below *alleged*, we account for (3-4), and more generally for Property 2 in (5). Since TC ensures that under both structures (that is, irrespective of scope) the_x *alleged* *N* is a constituent (which does not include the modifier), we account for C&R's observation in (2), and more generally for Property 1 in (5).

To see this in greater detail, let's start with the two structures available for (3).

- (3')a. [An alleged [alien from Mars]]
 λx John saw [the_x alleged alien]
 b. [An [[alleged alien] from Mars]]
 λx John saw [the_x alleged alien]

The structure in (3'b) is anomalous for the reasons mentioned above, namely it ascribes two contradictory properties to the same individual: that of being an alleged alien (hence not a real alien, see note 6) and that of being from Mars (hence a real alien). The structure in (3'a) accounts for the attested interpretation. It asserts that John saw someone who is alleged to be an alien from Mars (hence not a real alien from Mars). It also presupposes about this individual (a presupposition introduced by the trace) that it is an alleged alien (hence not a real alien, see note 6). This presupposition, although not tautological, is very easy to accommodate since the source for skepticism about an allegation that someone is an alien from Mars is very plausibly rooted in skepticism about the possibility that this person is an alien in the first place.

Let's now have a parallel look at the two structures available for (2):

- (2') a. [An alleged [mouse from Mars]]
 λx John saw [the_x alleged mouse]
- b. [An [[alleged mouse] from Mars]]
 λx John saw [the_x alleged mouse]

The structure in (2'b) is this time not anomalous: it simply asserts that John saw an individual who is truly from Mars and is alleged to be a mouse. This might correspond to the salient interpretation of the sentence. The structure in (2'a) is also possible. It asserts that John saw someone who is alleged to be a mouse from Mars (hence not a real mouse from Mars). But it also presupposes about this individual that it is an alleged mouse (hence not a real mouse). This presupposition might be a little harder to accommodate than in (3'a) since the source of skepticism about an allegation that something is a mouse from Mars is probably more plausibly confined skepticism about origin alone. But nevertheless, I think this is a possible interpretation, as originally pointed out by Pesetsky. Imagine, for example, that we see a picture of a creature that doesn't look like a mouse at all and someone tells us that it is indeed a mouse, though an oddly shaped one. This person furthermore tells us that it is oddly shaped because it comes from Mars. We are suspicious about both allegations, and send John to check things out. John sees the creature and remains skeptical. I think that the following day, we could describe this state of affairs with (2), as Pesetsky had claimed.

What we've seen is that the presupposition that comes from the trace accounts for C&R's observation. Using a bound variable with the presupposition that the individual that the variable refers to (under a modified assignment function) is alleged to be a mouse is what makes it impossible to assert the sentence in the situation characterized at the end of section 1. In other words, whether or not modification is in the scope of *alleged*, TC introduces a presupposition that is incompatible with a situation in which the speaker believes John saw a mouse and is skeptical about the allegation that this mouse is from Mars.¹³

¹³ The presupposed introduced by TC would likewise explain the facts in footnote 5.

To repeat, the two properties in (5) follow from the claim that there are two occurrences of the non-intersective adjective *alleged* and of the noun. When the modifier is LM-ed, *alleged* and the noun must form a constituent at the trace position, and the resulting presupposition (that the relevant individual is not necessarily in the actual world extension of the noun) accounts for property 1. In the Scope position (where the modifier is LM-ed) we have two possible constituency structures, and one of them (in which modification is in the scope of *alleged*) accounts for property 2.

Traces, given TC, are definite descriptions that contain variables and their binding is analogous to other anaphoric relationships in which such definite descriptions partake. (*Mary read every book* after QR is analogous to *Every book is such that Mary read that book.*) And indeed, what we've seen with extraposition arises more generally with such anaphoric relationships. (11a) is just a repetition of (2), which we were, at first, tempted to think requires a reading in which modification is outside the scope of *alleged*. (11b) shows that the same effect arises when presuppositions are triggered overtly: when there is an anaphoric relationship between the DP *an alleged mouse from Mars* and a DP which transparently introduces the presupposition that, under TC, is introduced covertly in (11a).

- (11)a. John saw an alleged mouse yesterday from Mars.
- b. Right after John saw an alleged mouse from Mars, he talked about the alleged mouse to various reporters.

(12a) is a repetition of (3), which showed that it is wrong to think that extraposition requires a reading in which modification is outside the scope of *alleged*. (12b) shows that the same holds under anaphora.

- (12)a. John saw an alleged alien yesterday from Mars.
- b. Right after John saw an alleged alien from Mars, he talked about the alleged alien to various reporters.

The facts in (11b) and (12b) are trivial consequences of the theory of presupposition. With the covert structures that result from LM and TC, the account extends automatically to extraposition.

4. Prediction for Condition C

The LM theory of extraposition is an extension to covert movement of a proposal made by Lebeaux (1989) for overt movement (to account for the interaction between overt A-bar movement and Condition C of the Binding Theory). Specifically, Lebeaux suggested that A-bar movement can circumvent a Condition C violation only if material that dominates the offending r-expression is LM-ed.

(13)Overt Movement Circumventing Condition C

- a. *He₁ is eager to sell the most expensive artifact from John's₁ excavation.
- b. [Which artifact from John's₁ excavation]₂ is he₁ eager to sell t₂?

(14)Lebeaux's LM Derivation of (13b)

- a. He₁ is eager to sell which artifact –wh-movement + I-to-C →
- b. [Which artifact]₂ is he₁ eager to sell [Which artifact]₂? –LM →
- c. [Which artifact from John's₁ excavation]₂
is he₁ eager to sell [Which artifact]₂? –TC→
- d. [Which artifact from John's₁ excavation]
 $\lambda x.$ is he₁ eager to sell [the_x artifact]?

As pointed out to me by Jeremy Hartman, the account I proposed for C&R's observation leads to very specific predictions when coupled with Lebeaux's proposal. Specifically, when Condition C circumvention requires LM into an NP that contains the adjective *alleged*, non-trivial presuppositions are predicted by TC. More specifically, we predict environments where Condition C requires LM to pattern with extraposition.

To understand the prediction, consider the two questions in (15). (15a) favors a parse in which the modifier *from his recent excavation* is in the scope of the adjective *alleged*.

- (15)a. Which alleged artifact from his₁ recent excavation is John₁ eager to sell to the museum before it is discovered for what it is?
- b. (#)Which alleged artifact from John's₁ recent excavation is he₁ eager to sell to the museum before it is discovered for what it is?

Based on Lebeaux's proposal, (15b) (in contrast to (15a)) requires LM. Consequently, based on what we said in the previous sections, the trace in (15b) will introduce the presupposition that the relevant individuals in the domain of quantification are *alleged to be artifacts* (i.e. that they are not presupposed to be artifacts), rather than the more natural presupposition that they are believed to be artifacts but that there is some skepticism about an allegation that they were found in John's recent excavation (see note 5). This difference in possible presupposition should, in principle, be noticeable.¹⁴

5. Embedded Late Merge and the Nature of Movement

As mentioned above (at the end of 3.1.), in ordinary cases of QR in which LM does not apply, traditional structures and those that result from TC receive the same interpretation. In such cases, TC triggers a presupposition at the trace position (that the individual denoted by the trace is a member of the QPs restrictor). But this presupposition projects as a tautology, since quantificational constructions at most demand that elements of the QPs restrictor satisfy the presupposition of the scope. If LM of the standard sort applies

¹⁴ The judgments are not straightforward, and although some speakers reported a contrast in the predicted direction, I think that at this stage I can only point to the relevant prediction.

(i.e. unembedded), nothing changes. Under this scenario, the presupposition triggered by TC is simply weaker than it would be without LM and therefore continues to project as a tautology. More specifically, the restrictor of the quantifier is modified by the LM-ed phrase to be a sub-set of the set of individuals that satisfy the presupposition of the scope. Hence, once again, the presupposition is necessarily satisfied.¹⁵ In other words, whether or not LM applies, the restrictor of the quantifier is guaranteed to entail the presupposition of the scope triggered by TC.

But this state of affairs is not guaranteed to hold generally, if embedded LM of the sort we've postulated in (10b) is allowed. We've seen that LM of a modifier within the immediate scope of *alleged* creates a restrictor (e.g. [*alleged [mouse from mars]*]) that does not entail the presupposition triggered by TC (*alleged mouse*), at least not when implicatures are factored in (see note 6). The latter presupposition needs to be accommodated for the sentence to be assertable, and it is this accommodation that I argued is behind C&R's observation.

If any of this is right, the obvious question to ask is whether there are restrictions on this form of embedded Late Merge. This question is one I would like to return to in the near future. What I have tried to argue in recent talks (see Fox 2014) is that embedded LM is not restricted by syntax at all and that what determines when it can occur is simply the question of whether the resulting presupposition can be accommodated. If this is correct, there are non-trivial consequences for the theory of movement, which I also hope that I can return to.

If any of this is right, the obvious question to ask is whether there are restrictions on this form of embedded LM. What I would like to argue now is that embedded LM is not restricted by syntax at all and that what determines when it can occur is simply the question of whether the resulting presupposition can be accommodated. If this is correct, there are non-trivial consequences for the theory of movement, which I will return to in section 14.

6. Exceptionally non-local Extraposition

In this section we will see cases of extraposition that appear to involve violations of a well-known locality condition on movement, namely the Complex NP Constraint, CNPC. We will also see that the distribution of extraposition out of complex NPs (CNP) is not uniform, but is affected by properties of material within the CNP. In later sections, we will see that the apparent violations of locality follow from the possibility of combining local QR with deeply embedded LM. I will argue that properties of material in the CNP are expected to affect acceptability, given the observations made above about the potential consequences of embedded LM for the presuppositions triggered by TC.

¹⁵ The situation might be different for non-conservative quantifiers, since presuppositions are likely to project differently. See Bhatt and Pancheva (2004), Fox (1999, 2001), and Romoli (2015).

6.1. Apparent Non-locality

It is well known that movement out of CNPs leads to unacceptability, (16). But parallel extraposition is – relatively speaking – acceptable, (17).

- (16) a. I bought a car owned by Fred.
 b. *By whom did you buy a car owned?
- (17) ? We [[[looked at [a house owned by someone]] yesterday] who teaches at UCLA].

If extraposition involved movement of the extraposed relative clause, this contrast would be mysterious. Under QR + LM, it also seems surprising, since the relevant environment seems to be a scope island:

- (18) #We [[[looked at [a house owned by every faculty member]]]]
 compare: *We looked at a house **jointly** owned by every faculty member.*

The sentence in (18) entails the existence of a single house that is owned by every faculty member, and, thus, conflicts with world knowledge (common assumptions about ownership, as apposed to joint ownership). If the universal quantifier could take matrix scope, an alternative interpretation would be available for the sentence, one that would be natural. It is thus reasonable to conclude that QR is subject to the CNPC (as is standardly assumed, see Ruys and Winter 2011, Lechner 2015, i.a.).

One could, of course, claim that the indefinite, *someone*, in (17) is special and can move by QR out of a CNP. But this (although arguably motivated on independent grounds, see Schwarz 2001) would not explain the counter-example to Williams's generalization exemplified below:

- (19) I'll [[read [every paper that was recommended by any linguist] next year] who teaches at UCLA].

If extraposition of the RC in (19) involves QR of the NPI *any linguist*, we would expect the requirements of the NPI (that it be in a downward entailing environment) not to be satisfied, contrary to fact.¹⁶

¹⁶ This argument would not go through if the relative clause could license the polarity item independently of its external argument (what Dayal 1998 has called sub-triggering). But if this were possible, we would expect the following to be acceptable.

- (i) *I read anything yesterday that John had recommended.

One, of course, still needs to understand why sub-triggering is incompatible with extraposition. If sub-triggering requires reconstruction of the head-noun to its base position, things would follow from the observation of Hulsey and Sauerland (2006) that LM is incompatible with head-raising diagnostics. See Sportiche (2015) for further discussion.

6.2. The local environment of the source NP

What I would like to point out next is that when the CNPC appears to be violated, certain manipulations of material internal to the complex NP can affect acceptability:

- (20) a. I'll [[explain [a paper that was recommended by a linguist] when we meet] who teaches at UCLA].
 b. *I'll [[explain [a paper that wasn't recommended by any linguist] when we meet] who teaches at UCLA].
- (21) a. I'll [[read [every paper that was recommended by any linguist] next year] who teaches at UCLA].
 b. *I'll [[read [a that wasn't recommended by any linguist] next year] who teaches at UCLA].
 c. ??/* I'll [[read [a paper that was recommended by every linguist] next year] who teaches at UCLA].
 d. ??/* I'll [[read [every paper that was recommended by every linguist] next year] who teaches at UCLA].¹⁷

Our puzzle is thus twofold. On the one hand, we would like to understand the apparent insensitivity of extraposition to the CNPC. On the other hand, we have to explain why this insensitivity depends on properties of the CNP. In the following sections, I would like to argue that embedded LM provides an answer. Extraposition is derived by QR + LM. But if embedded LM is possible, the QRed phrase need not be the local QP (the one that takes the source NP as complement). Instead the complex QP that dominates the local QP can QR followed by deeply embedded LM (merge with the local NP). This leads to apparent violations of the CNPC, with properties of the CNP relevant for evaluating the presupposition triggered by TC.

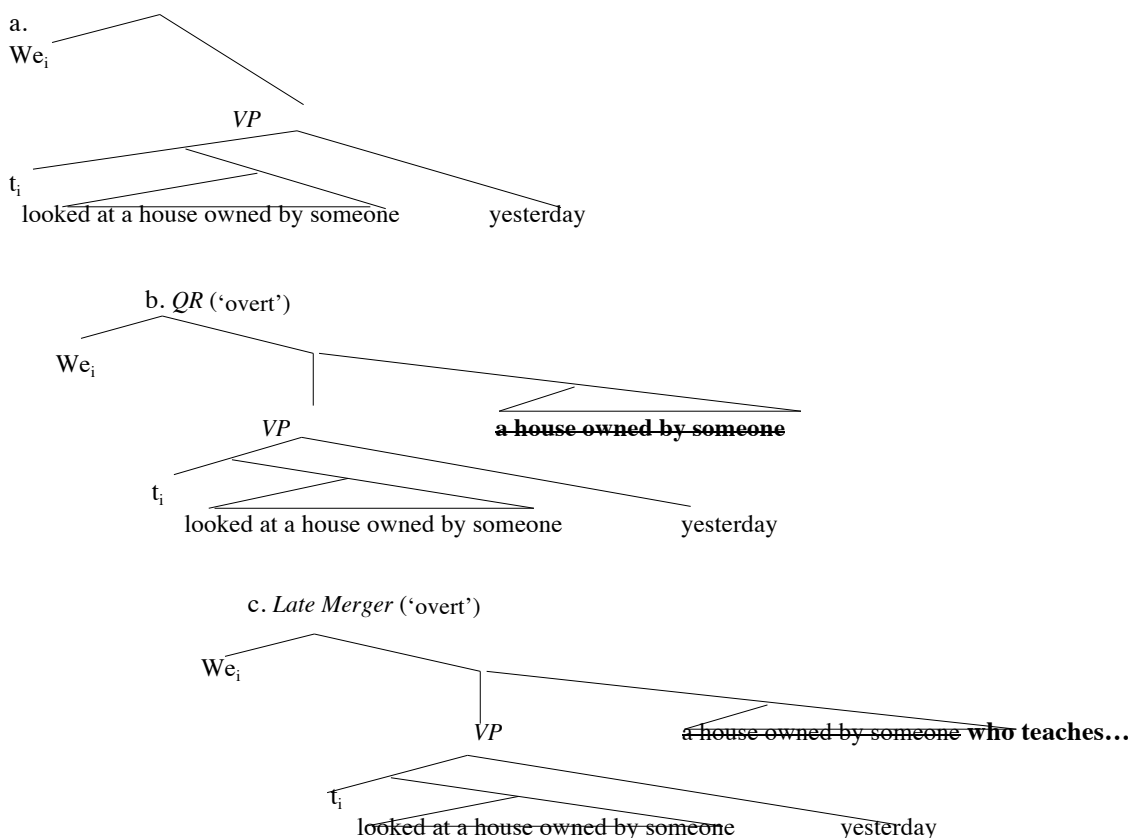
7. Derivation of Exceptionally non-local Extraposition

In section 3 we argued that embedded LM is, in principle, possible. Specifically, we argued that LM can apply not only to the NP complement of a QP that has undergone QR, but also to the nominal daughter of this NP complement (thus accounting for the acceptability of (3) and (4)). What I would like to point out now is that the apparent violations of the CNPC we've seen in section 6 are derived if LM can apply in even more deeply embedded positions. Specifically, the sentence in (17) can be derived by QR of

¹⁷ Speakers report that (c) and (d) are not as bad as (b). The difference would follow from the approach I will be developing, if violations of the complex NP constraint were not as devastating as violations of the condition on NPI licensing or the uninterpretability that results in these particular cases from embedded LM (as we will see). In the (c) and (d) cases, an interpretable structure would result from QR of the source QP followed by simple LM. In the (b) cases, only embedded LM would be available (since the source QP must be in the scope of its licensor), and embedded LM as we will see leads to uninterpretability in DE environments.

the complex NP *a house owned by someone* followed by LM of the RC to the deeply embedded QP, *someone*.¹⁸ The derivation is given below:

(17)' We looked at a house owned by someone yesterday who teaches at UCLA.



In order to see whether we derive the appropriate interpretation we have to compute the meaning of the QP that is formed by embedded LM and apply it to the meaning its sister. For the sake of simplicity, we can, of course, assume (as we have in (9)) that the non-quantificational subject *we* is reconstructed to its base position:

(17)" **Interepretation of (17) (by Trace Conversion)**

$$\begin{aligned}
 & \llbracket \text{A house owned by someone who teaches at UCLA} \rrbracket \\
 & \quad (\lambda x \llbracket \text{we looked at } [\text{the}_i \text{ house owned by someone}] \rrbracket^{1 \rightarrow x}) \\
 = & \llbracket \text{A house owned by someone who teaches at UCLA} \rrbracket \\
 & \quad (\lambda x: x \text{ is a house owned by someone. we looked at } x)
 \end{aligned}$$

¹⁸ Whether this form of embedded LM is possible can also be tested by looking at overt movement if we have ways of investigating the properties of the trace (e.g. Condition C of the Binding Theory, see section 4). Judgments, however, are not as straightforward as one would hope for. For conflicting reports, see Tada (1993), Sauerland (1998a, section 2.2.) and Sportiche (2015).

Given embedded LM, the presupposition triggered by TC is not *guaranteed* to project as a tautology. However, in this particular case it does: although the LM-ed modifier is not conjoined with an NP to form the QPs restrictor (as it is in non-embedded LM), the restrictor that is formed by LM nevertheless entails the presupposition triggered by TC. The presupposition triggered by TC is the presupposition that x (the individual denoted by the trace under the relevant assignment function) is a house owned by someone. Since this presupposition necessarily holds for every individual that is in the restrictor of the quantifier (of every house owned by someone who teaches at UCLA) the interpretation we derive is equivalent to what we would get by traditional representations.

So we have seen that embedded LM provides an account for apparent violations of the CNPC. A complex QP can undergo QR, with LM applying to an NP that is deeply embedded within this complex QP. The structure ends up interpretable, as long as the presupposition triggered by TC projects as a tautology (or can be accommodated as in (3) and (4)). This set of assumptions leads to rather intricate predictions, which I will attempt to investigate below (showing in particular that the facts in (20) and (21) are understood, once we look at the presuppositions that would be triggered by TC in these cases).

One set of predictions pertains to the head of the chain formed by QR, namely that this need not be the local QP but can be a different QP that dominates the local QP. Another set of predictions pertains to the tail of the chain, namely that it will include everything at the head of the chain with the exception of the LM-ed material. The specific predictions I will be looking at (of both varieties) pertain to scope, presuppositions triggered by TC, tense interpretation and Condition C of the binding Theory.

But before we get there, I need to highlight a crucial additional assumption which is implicit in Fox and Nissenbaum (1999) and subsequent work on QR and LM, namely that not every constituent can undergo QR. If every constituent could undergo QR we would not account for WG. For example, we would not derive the observation made in (8) if VP could undergo QR. My working hypothesis is going to be that only quantificational phrases can undergo QR.¹⁹

8. Scope

WG states that extraposition has consequences for the scope of the source QP (which we have also referred to as the local QP), see footnote 10. But if embedded LM is possible, this is no longer predicted. Extraposition need not affect the scope of the source QP because the relevant sequence of words can sometimes result from QR of a properly containing QP as in (17)'. Still extraposition should always have consequences for the scope of *some* QP that dominates (or is identical to) the source QP.

As discussed, (19) already provides a counter-example to WG, as it was formulated in (7), and thus motivates the following (weaker) reformulation.

¹⁹ This is a consequence of Scope Economy (Fox 2000, Mayr and Spector 2010), and hence is independently motivated. However, one could imagine various alternatives that would suffice for current purposes. What I think is crucial for this paper is to rule out QR of VPs and NP complements of D.

- (22) **Modified Williams's generalization (MWG):** When an adjunct β is “extraposed” from a “source QP” α , there must be a QP α' which is either identical to or dominates α (α' reflexively dominates α) and the scope of α' must be at least as high as the attachment site of β (the extraposition site).

Additional counter-examples to WG, which lead to the same conclusion, are provided in (23), where, once again, an NPI would not be in a downward entailing environment if WG were correct.

- (23) a. I'll [[talk to you about [every paper that was recommended by any linguist] when we meet] who teaches at UCLA].
 b. I'll [[do [nothing that was prohibited by anyone] when I am a student] who would have any authority over me].

The acceptability of these sentences is expected with embedded LM, since the source QP need not QR to the position where the extraposed material is linearized; instead the higher QP (the one that dominates the source QP) can QR. But the argument for embedded LM would be stronger if it were based on non-trivial consequences of MWG for cases in which MG is counter-exemplified. To test such consequences, we would have to look at the scope of the non-local QP.

Stated somewhat differently, Fox and Nissenbaum (1999) and subsequent work focused on cases in which only one QP dominates the source QP. For such cases, whether or not embedded LM is possible has no consequences, and, in particular, WG and MWG are equivalent. When two distinct QPs dominate the source QP, embedded LM allows the non-local QP to QR, and thus predicts counter-examples to WG. That the local QP need not have wide scope has been verified, but now we would like to examine the scope of the non-local QP.

8.1. Setting the Stage

To set the stage, I would like to introduce an environment in which the relevant scopal relationships, which are normally quite difficult to detect by introspection, can be controlled for by plausibility considerations. Consider the two sentences in (24), which differ minimally in the boldfaced material, a difference that has clear consequences for scope assignment. The natural interpretation of (24a) involves wide scope for the negative QP (*nothing prohibited*) over the *without*-Phrase. Under this interpretation, the sentence states that there is nothing prohibited that John did without (as a consequence) being reprimanded for doing it [i.e. that he was reprimanded for everything prohibited that he had done]. The sentence can also be interpreted with narrow scope for the negative QP (below *without*). Under this less natural interpretation, the sentence states that John did nothing prohibited and that he was not reprimanded (as a consequence of doing nothing prohibited). This interpretation is probably less natural because no one would expect anyone to be reprimanded for avoiding what is prohibited. In any event, the status of this second interpretation of (24a) is not going to be important for our purposes.

(24) **Plausibility Considerations Determine Scope:**

- a. John did nothing **prohibited** without being reprimanded.
(preferred reading: *nothing* > *without*)
- b. John did nothing **required** without being reprimanded.
(*without* > *nothing*; # *nothing* > *without*)

What is important is that this second scopal relationship (narrow scope for the negative QP) is the only plausible option in (24b). For (24b), in contrast to (24a), wide scope for negative QP is entirely implausible. The sentence states that John did nothing required and (nevertheless) was not reprimanded. In other words, it states, that the school was not particularly harsh in enforcing its requirements. Had the negative QP received wide scope, the sentence would have stated that there is nothing which is required that John did without (as a consequence) being reprimanded for it, i.e., that the school penalized John every time he did something which was required, and that, of course, is not particularly plausible. The same contrast can be seen in (25), which differs from (24) just in the presence of a complex *by*-phrase, a difference that will turn out to be useful in distinguishing WG from its modification in (23).

(25) **Plausibility Considerations Determine Scope:**

- a. John did nothing **prohibited** by anyone who teaches in this school without being reprimanded. (preferred reading: *nothing* > *can*)
- b. John did nothing **required** by anyone who teaches in this school without being reprimanded. (*can, without* > *nothing*; # *nothing* > *can, without*)

8.2. Where WG and MWG are Equivalent

Before we get to our crucial test case, consider the contrasts in (26) and (27), which instantiates both WG and MWG. In both the (a) and the (b) cases, extraposition of the relative clause rules out a parse in which the source QP receives narrowest scope (below the *without*-Phrase). In the (a) cases, this scopal requirement is consistent with a natural interpretation, whereas in the (b) cases it conflicts with the plausibility considerations outlined above. More specifically, in the (b) cases extraposition leads, by WG, to an implausible interpretation (suggesting that students are penalized when meeting the school's requirements).

- (26) a. John did nothing, without being reprimanded, that is prohibited.
- b. #John did nothing, without being reprimanded, that is required.
- (27) a. John can do nothing, without being reprimanded, that is prohibited.
- b. #John can do nothing, without being reprimanded, that is required.

Since no distinct QP dominates the source QP, the contrast follows under MWG as well and cannot serve to distinguish the two generalizations. The same holds, of course, for (28) and (29).

- (28) a. John did nothing, without being reprimanded, prohibited by anyone who teaches in this school.
- b. #John did nothing, without being reprimanded, required by anyone who teaches in this school.
- (29) a. John can do nothing, without being reprimanded, prohibited by anyone who teaches in this school.
- b. #John can do nothing, without being reprimanded, required by anyone who teaches in this school.

8.3. Distinguishing WG from MWG

Consider now the facts in (30) and (31). The (a) sentences, although difficult to parse, provide counter-examples to WG of the sort we've seen in (19) and (23), and likewise argue that MWG is a better generalization. More specifically, here, just like in (19) and (23), we see that the source QP (the NPI) need not take scope at the extraposition site, since this would place it outside the scope of its licensing operator. However, in these sentences, in contrast to (19) and (23), we can test the consequences of extraposition for the scope of the non-local QP (the one that properly dominates the source QP).

- (30) a. John did nothing prohibited by anyone, without being reprimanded, who teaches in this school.
- b. #John did nothing required by anyone, without being reprimanded, who teaches in this school.
- (31) a. You can do nothing prohibited by anyone, without being expelled, who teaches in this school.
- b. #You can do nothing required by anyone, without being expelled, who teaches in this school.

For the NPI to be licensed (and for the CNPC to be satisfied) the derivation must involve QR of the negative QP *nothing required/prohibited...* followed by embedded LM of the relative clause to the source NP. The consequences of QR for scope (MWG) are corroborated: we see that the negative QP must take scope over the *without*-Phrase and that subsequently there is a contrast between the (a) and the (b) sentences which parallels what we've seen in (26) and (28).

9. Presupposition Triggered by Trace Conversion

In section 3.2. we've seen that the presupposition triggered by TC can have non-trivial semantic consequences when embedded LM takes place. What I would like to demonstrate now is that these predicted consequences account for the observations made in section 6.2. Specifically I would like to show that in all the unacceptable cases of non-local extraposition we've looked at, the kind of embedded LM that would be needed results in non-trivial presuppositions which, when accommodated, lead to a violation of independently motivated conditions on language use.

Consider the contrast in (20) repeated below.

- (20) a. I'll [[explain [a paper that was recommended by a linguist] when we meet] who teaches at UCLA].
 b. *I'll [[explain [a paper that wasn't recommended by any linguist] when we meet] who teaches at UCLA].

Consider, now, the two LF representations that would be associated with the two sentences, once local QR and embedded LM apply (ignoring the irrelevant adverb *when we meet*).

(20)' **LF representations for (20) (by local QR + embedded LM)**

- a. [A paper that was recommended by a linguist who teaches at UCLA].
 I'll explain [a paper that was recommended by a linguist]
 b. [A paper that wasn't recommended by anyone who teaches at UCLA]
 I'll explain [a paper that wasn't recommended by anyone]

Next we need to consider the interpretive consequences of TC for these structures derived by embedded LM. First consider (20)'a. In this case, just as in (17)" above, TC has no interesting consequences.

(20a)" **Interpretation of (20a)' (by Trace Conversion)**

$$\begin{aligned} & \llbracket \text{A paper that was recommended by a linguist who teaches at UCLA} \rrbracket \\ & (\lambda x \llbracket \text{I'll explain [the}_1 \text{ paper that was recommended by a linguist}] \rrbracket^{1 \rightarrow x}) \\ = & \llbracket \text{A paper that was recommended by a linguist who teaches at UCLA} \rrbracket \\ & (\lambda x: x \text{ is a paper that was recommended by a linguist. I'll explain } x) \end{aligned}$$

The presupposition triggered by TC, in this case, is the presupposition that x (the individual denoted by the trace under the relevant assignment function) is a paper that was recommended by a linguist. Since this presupposition necessarily holds for every individual in the restrictor of the quantifier (for every paper that was recommended by a linguist who teaches at UCLA) the interpretation we derive is equivalent to what we would get by traditional representations.

Compare this with (20)'b. (20)'b will also be interpreted by TC, but this time the resulting presupposition is not at all trivial.

(20b)" Interpretation of (20)'b (by Trace Conversion)

$$\begin{aligned}
& \llbracket \text{A paper that wasn't recommended by anyone who teaches at UCLA} \rrbracket \\
& \quad (\lambda x \llbracket \text{I'll explain [the}_1 \text{ paper that wasn't recommended by anyone}] \rrbracket^{1 \rightarrow x}) \\
= & \llbracket \text{A paper that wasn't recommended by anyone who teaches at UCLA} \rrbracket \\
& \quad (\lambda x: x \text{ is a paper that wasn't recommended by anyone. I'll explain } x)
\end{aligned}$$

The presupposition triggered by TC, this time, is the presupposition that x (the individual denoted by the trace under the relevant assignment function) is a paper that wasn't recommended by anyone. But this does not necessarily hold of every individual in the restrictor of the quantifier (of every paper that wasn't recommended by anyone who teaches at UCLA). So if the sentence is to be acceptable, the speaker would have to accommodate the presupposition, as in the construction discussed before (section 3.2.).

The problem, however, is that, once the presupposition is accommodated and the common ground contains the information that the relevant paper wasn't recommended by anyone at all (not just by anyone who teaches at UCLA), we are hard pressed to understand why the speaker uttered the sentence in the first place.²⁰ Specifically, there is a simpler alternative, which differs only in the absence of the relative clause, namely (32), which would have served the same communicative purpose; this simpler alternative and the sentence (20)b would make the same communicative contribution once the presupposition triggered by TC is accommodated.²¹

(32) I'll explain a paper that wasn't recommended by anyone (when we meet).

²⁰This discussion might benefit if we established first a specific theory of presupposition projection. With such a theory I could state explicitly what presupposition is accommodated. It would be simplest for my purposes to assume that presuppositions project universally (as apposed to by the Strong Kleene recipe). I think, however, that we can do without a specific theory by pointing out what it is clear enough, namely that for the speaker to believe that an existential sentence is true, the speaker has to believe that there is an individual that satisfies both the restrictor and the scope of the quantifier, and that this requirement suffices to explain why the sentence is "speaker equivalent" to the simpler alternative in which the relative clause *who teaches at UCLA* is omitted.

²¹ David Pesetsky points out the potential relevance of certain restrictive relative clauses that don't seem to restrict the NP they modify. If the restrictive RC in (i), for example, truly does not restrict the NP it combines with, its contribution to the meaning of the sentence would be less-straightforward than what is assumed in our discussion and one should ask whether we still rule-out embedded LM (which (ii) indicates we need to):

- (i) John will meet a person that didn't read any paper **that ever existed**.
- (ii) *John will meet a person that didn't read any paper next year **that ever existed**.

I think, however, that the relative clause clearly contributes to the meaning of (i) and an understanding of its contribution would probably lead to the prediction that (ii) should be ruled out by both manner and Magri's considerations (just like (20b)). A simple way of understanding the contribution of the relative clause is to assume that (like ordinary restrictive relative clauses) it (formally) must exclude elements from the domain of quantification. In order for it to do this, the domain needs first to be *widened* to include elements that wouldn't otherwise be considered, in our example to include (even) papers that never existed.

So this provides can account for the unacceptability of (20)b. For the sentence to be uttered, a presupposition must be accommodated to the common ground, and once the presupposition is accommodated, (20)b and (32) are “contextually equivalent”. (20)b, thus, ends up violating two independently motivated conditions on language use. One of these conditions (sometimes referred to as the Maxim of Manner) states that a sentence is unacceptable when it is contextually equivalent to a simpler alternative (Katzir 2011, Meyer 2013, 2014, i.a.).

The other condition states that a sentence is unacceptable when it is contextually equivalent to an alternative (relevant for the computation of scalar implicatures) that is logically stronger (Magri 2009, 2011, to appear). Under this second condition, there is another way of talking about the contrast in (20), which might be useful to examine carefully. (20)a has two components of meaning, assertion and presupposition, which live happily with each other.

(20a)''' The Good Case (informally):

There is a paper that was recommended by a linguist who teaches at UCLA, s.t.

- (a) it is presupposed that this is a paper recommended by a linguist
- (b) it is asserted that I (the speaker) will explain this paper when we (speaker and addressee) meet.

(20)b, by contrast, has three components of meaning, assertion, presupposition and Scalar Implicature, and two of these components (the presupposition and the Scalar Implicature) contradict each other.

(20b)''' The Bad Case (informally):²²

There is a paper that wasn't recommended by anyone who teaches at UCLA, s.t.

- (a) it is presupposed that this paper wasn't recommended by anyone
- (b) it is asserted that I'll explain this paper when we meet.
- (c) Scalar Implicature: the relevant paper is going to be recommended by a linguist (who does not teach at UCLA)

These two accounts of the contrast in (20) are based on independently motivated and closely related conditions on language use. I leave it to the reader to see that both accounts explain all of the bad cases discussed in 5.2. In all of these cases, like in (20b), the embedded LM-ed relative clause appears in a downward entailing environment that is embedded by the restrictor of the QR-ed phrase. As such its absence in the trace position (resulting from LM) leads (by TC) to a logically stronger presupposition than the one derived in its presence. Once this stronger presupposition is accommodated, the lethal

²² To give a full account of the conflict, I would have to go over the theory of obligatory Scalar Implicatures developed by Magri (2011), and doing that will require more space than I have available. I would also have to go over the theory of alternatives that explains why the Scalar Implicature in (20b)'''c is expected (Katzir 2007).

competition with a simpler and stronger alternative arises. In the next section, we will have the opportunity to go over one additional example.²³

As in section 3 (ex. (12)), we are reassured by the fact that a parallel contrast arises in anaphoric relationships in which the covert objects we've been postulating occur overtly.

(33) **Parallel Anaphora Contrast**

- a. John read a paper that is going to be recommended by a linguist who teaches at UCLA. And he loved this paper that is going to be recommended by a linguist.
- b. John read a paper that isn't going to be recommended by any linguist who teaches at UCLA. #And he loved this paper that isn't going to be recommended by any linguist (at all).

10. Interaction Between Scope and Presupposition

Consider the sentence in (34a).

- (34) a. By tomorrow, John will find a picture of every boy who studies with him.
 $\exists x[\forall y[\text{boy}...(y) \rightarrow \text{picture}(x,y)] \& \text{Find}(J, x)]^{24}$ $(\exists > \forall)$
 $\forall y[\text{boy}...(y) \rightarrow \exists x[\text{picture}(x,y) \& \text{Find}(J, x)]]$ $(\forall > \exists)$
- b. John will find a picture of every boy, by tomorrow, who studies with him.
 $*\exists x[\forall y[\text{boy}...(y) \rightarrow \text{picture}(x,y)] \& \text{Find}(J, x)]$ $*(\exists > \forall)$
 $\forall y[\text{boy}...(y) \rightarrow \exists x[\text{picture}(x,y) \& \text{Find}(J, x)]]$ $(\forall > \exists)$

In this sentence, a universal QP is base-generated within the restrictor of an existential QP. There are consequently two possible scopal interpretations: the universal QP can take scope internal to the restrictor of the embedding QP, $(\exists > \forall)$, or it can take matrix scope, $(\forall > \exists)$. Under $(\exists > \forall)$, the sentence asserts that there is a single picture in which every boy who studies with John is depicted and that John will find this picture by tomorrow. Under $(\forall > \exists)$, the sentence asserts that for every such boy there is a picture depicting him that John will find by tomorrow. What we see in (34b) is that extraposition requires sentential scope for the universal quantifier, $(\forall > \exists)$.

This observation conforms to WG: the relative clause is extraposed and the source QP (headed by the universal quantifier) must take scope over the adverb and subsequently cannot remain within the restrictor of the existential QP as required for $\exists > \forall$. But this state of affairs does not follow as directly if embedded LM is possible. Under embedded

²³ A Magrian account of the contrast in (20), in contrast to one based on the maxim of manner, would predict that embedded LM should not only be impossible in DE contexts but also in non-monotonic contexts. So if the Magrian account is correct, we would need to treat *alleged* and *former* as upward monotone operators with the effects of embedded LM following as an SI (see notes 4 and 5).

²⁴ These formulas are not complete: the adverb *by tomorrow* is omitted and the ellipsis sign ... should be filled with information from the relative clause '*& studies(y, with John)*'.

First we see in (37), just like we saw in (19) and (23) above, that the presence of a dominating QP makes it unnecessary for the source QP to take wide scope. If the source QP in (37) had to take scope at the extraposition site, SoT would not be licensed and the attested “simultaneous” interpretation would not be available. Specifically, it would be impossible for the predicate *pregnant* to be evaluated at the time of marriage.

(37) **Non-Local extraposition – no effect relative to material in a dominating DP**²⁵

- a. You told me that you wouldn’t mind dating a man who **had** married a woman who *was* pregnant.
- b. You told me that you wouldn’t mind dating a man who **had** married a woman, when we met, who *was* pregnant.

Time of marriage can be interpreted as contained in time of pregnancy whether or not there is extraposition.

This is predicted: extraposition can (in this case must due to the CNPC) be derived by QR of a QP that properly contains the source QP; In this particular case, the QR-ed phrase is the complex QP, *a man who had married a woman*; QR is followed by embedded late merge, which creates an SoT environment for the relative clause. More specifically, in the final representation *was pregnant* is inside the scope of past tense on *had*, and can subsequently receive a “present interpretation” relative to the time of marriage (or more accurately, the time at which *married* is evaluated).

But, again, what is more interesting is that in such cases we predict that the dominating QP will have to take scope at the extraposition site, and, thereby, will be outside the SoT environment of a tense operator that is interpreted below the extraposition site. This prediction is corroborated in (38).

(38) **Non-Local extraposition – there is an effect relative to material outside of dominating DPs**

- a. You told me that you wouldn’t mind dating a man who **was** married to a woman who *was* pregnant.
- b. You told me that you wouldn’t mind dating a man who **was** married to a woman, when we met, who *was* pregnant.

Extraposition (in (b)) makes it impossible for time of dating to be contained in time of marriage or in the time of pregnancy.

In both sentences in (38), just like in (37), the tense on the extraposed relative clause can receive a simultaneous interpretation relative to a higher tense inside the complex QP (pregnancy can be evaluated in the time of marriage). However, here we can detect an effect of extraposition on SoT. In (38a), the time of marriage (and subsequently pregnancy) can be evaluated at the time of dating, but in (38b) it cannot. (To get clear

²⁵ In (37) and in (38) I boldface the relevant “high” occurrences of past tense that can license sequence of tense on italicized “low” past tense.

judgments it might be useful to consider that the speaker and the addressee are too young to get married, say in kindergarten.)

This, again, is predicted: extraposition, as we said, is derived by QR of the QP that properly contains the source QP, *a man who was married a woman*, followed by embedded LM, which allows for “present interpretation” of *was pregnant* relative to the time of marriage, but does not allow such an interpretation relative to time of dating, since, given QR, *was pregnant* is outside the scope of the past tense on *would* (which determines the evaluation time for *dating*.)

12. Condition C of the Binding Theory

Embedded LM has a variety of predictions that we tried to test in the previous sections. Before discussing possible theoretical consequences, I would like to talk about another clear and obvious prediction, which, unfortunately, I haven’t been able to test as successfully as I would like to. It is well known, since Tarlidsen (1981), that extraposition can have consequences for condition C of the Binding theory:

- (39) a. ?? I told him_i about a new argument that supports John’s_i theory.
 b. I told him_i about a new argument the other day that supports John’s_i theory.

This contrast is derived by the theory of extraposition proposed here if coupled with a parsing preference for a structure that does not involve LM.²⁶ The string of words in (39a) can be mapped by the parser to a syntactic structure that does not involve LM. This is the preferred structure, and since it violated Condition C, the sentence feels somewhat deviant (for some speakers). (39b), by contrast, cannot be parsed without LM and, since the resulting LF does not violate condition C, the sentence is judged as acceptable.

But embedded LM makes additional predictions. Specifically, it is predicted that when LM is embedded enough to place a relative clause (that contains an r-expression) within the c-command domain of a co-indexed element, Condition C would be violated. Evidence that this prediction is correct was presented in Fox (2002) based on the following contrast.

- (40) a. ? I told him_i about your new argument the other day that supports John’s_i theory.
 b. * I told you about his_i new argument the other day that supports John’s_i theory.

There are additional predictions pertaining to cases in which the source QP is embedded within another QP. When locality and/or scope considerations teach us that the larger QP has to undergo QR, the predictions for Condition C are rather intricate. Unfortunately, the data is not as clear as I would like it to be. In any event the predictions are the following:

- (41) a. You told me that she_i wouldn’t mind dating a man who had married a woman, when we met, who *was* Sue’s_i boss.

²⁶ Something like a parsing preference would be needed, as far as I can see, for any description of the facts, unless string vacuous extraposition (in (39a)) were to be blocked within grammar, see Fox (2000: p. 76).

- b. You told me that you wouldn't mind dating a man who she_i had introduced to a woman, when we met, who was Sue's_i boss.

13. Summary

We began this paper looking at extraposition from NPs that contain the non-intersective adjective *alleged*. We claimed that, although extraposition of this sort has interpretive consequences, these couldn't be accounted for by a locality condition on extraposition of the sort assumed by Culicover and Rochemont. Instead, they follow from LM which can occur within the scope of *alleged*, with non-trivial presuppositions predicted by our assumptions about the interpretation of chains (TC). The hypothesis we, then, investigated is that there are no syntactic constraints on the depth of LM, and that apparent locality conditions, in general, result from the projection of non-trivial presuppositions (triggered in the trace position given TC). Finally we looked at a set of predictions that pertain to the LF position of the source QP and QPs that dominate it. All in all, I think it is fair to say that we've ended up with some non-trivial evidence for the LF representations derived by LM. The question I would like to ask in the remainder of this paper (following Bachrach and Katzir 2009, Johnson 2012, 2016 and Sportiche 2015) is whether LM is literally needed or whether there is, instead, a way to derive the required LF representation that is (a) not counter-cyclic and (b) consistent with a multi-dominance view of movement.²⁷ I will argue that a positive answer is conceivable, by sketching a system in which a "movement chain" is generated by a single operation (Merge) that is capable of yielding representations in which a single phrase has more than one mother.²⁸

²⁷ Sportiche (2015) suggests that the LM structures are derived by a combination of Internal Merge and interpretive "neglect" of material that has more than position in the final representation. To get the effects of LM in the case of overt movement, he assumes that the semantic component can (up to interpretability) neglect material at the tail of an overt movement chain. To derive extraposition, he assumes that the phonological component can neglect some material at the head of a chain and some at the tail, with (M)WG derived in conjunction with an independent ban on reconstruction for rightward movement. However, to derive the effects of embedded LM on the presupposition of traces (facts discussed in sections 3 and 9), one would have to develop Sportiche's system in such a way that phonological neglect (at the head of the chain) entailed semantic neglect of the (non phonologically-neglected) material at the tail of the chain. A similar issue arises for simpler observation about extraposition, e.g. its effects on condition C of the binding theory (Taraldsen 1981), or on Parallelism (Fox 2002).

What I will propose here is mostly inspired by Johnson's work (Johnson 2012, 2016). The main difference between what I propose and what Johnson proposed pertains to the way in which the linearization conflicts that arise from Multi-dominance are negotiated when syntactic structures are linearized.

²⁸ My goal is rather modest – to argue that LM is not, in principle, incompatible with arguments for multi-dominance, by providing a particular "proof of concept". As a by-product, I will be highlighting one particular way of thinking about things. I, of course, hope that it might turn out to be useful, but my main purpose is to explain why I think (following Bachrach and Katzir and Johnson) that LM is in principle compatible with representations of displacement in which a single constituent can have more than one structural position.

14. Movement as Multi-Dominance

Following Bachrach and Katzir (2009), Johnson (2012, 2016), and Fox and Johnson (2016), I would like to raise the possibility that LM, although useful in describing a class of LF representations, does not exist. The relevant representations are ones in which the head of a “movement chain” shares properties with the tail, but is not identical to it (in that it contains LM-ed material). If representations of this sort are to be derived cyclically by Merge, we must specify conditions under which the interpretive components (phonology and semantics) can treat different linguistic objects as members of the same “chain”. In particular, the “LM” representations must be derived by merge of a QP at the base position and another QP at the scope position. These QPs, though not identical, must be similar enough to be treated by the interpretive components as members of a “movement chain” (in semantics to trigger TC at a low position, and in phonology to trigger deletion of certain positions in structures involving multi-dominance).

To illustrate, consider the simple case of extraposition in (6) and the LF representation derived by QR + LM in (6').

(6) We saw a painting yesterday that John talked about.

(6)' LF of (6) [word order irrelevant]

[[a [painting [that John talked about t]]] [we saw a painting yesterday]]

I would like to assume that an LF representation of this sort involves multi-dominance with *painting* having two mothers *a painting* (at the tail of the QR chain) and *painting that John talked about* (at the head of the QR chain). How can such a representation be derived? One possibility, suggested in Sportiche (2015), is that it is derived by movement (internal merge) of the QP *a painting that John talked about* followed by semantic “neglect” of the relative clause at the trace position. But this proposal relies on independent neglect at the phonological component to derive the word order of extraposition, and therefore does not explain the correlation with semantic neglect; in other words, it does not explain the effects of extraposition on the presupposition triggered by the trace, or on Condition C and Parallelism in ACD constructions (see note 27). So, instead, I would like to suggest that the structure is derived directly by Merge.

I will assume, following Chomsky (1995), that lexical items in a linguistic representation are marked in a way that differentiates structures in which different tokens of the same lexical item occur in different positions from otherwise identical structures in which it is a single token that occupies two positions. For example, in the sentence *every boy likes every boy*, I will adopt the common assumption that the subject and the object are occupied by different phrases. Specifically, I will assume that the primitive elements that make up the subject and the object are distinct tokens of the same lexical items (which, might be distinguished, for example, by different numerical indices, see (42a) below).

So grammar forms structures by a single operation of Merge and in such structures one phrase marker can, in principle, have more than one mother. (The same token of a

lexical items can be merged with two distinct phrases, and likewise for complex phrases.) When this happens, however, the structure will yield contradictory phonological instructions (as pointed out by Nunes 2001, 2004, assuming the copy theory of movement, see Johnson 2012). For example, if the same phrase *every boy* were to appear as object of V and as subject of vP, *boy* would have to precede *every* (because subjects in English precede objects) and would have to follow *every* (because determiners precede nouns).

Nevertheless, to account for displacement we might want to allow for multi-dominance. And if so, there must be principles that would tell the phonology when it can ignore/delete (or neglect to use Sportiche's term) one of the multiple positions that a phrase can occupy. And it is reasonable to hope that there is a natural way to articulate these principles that would yield the various properties of displacement, most notably locality. With this as a goal, I will assume that phonological neglect is conditioned by syntactic agreement, thereby allowing me to adopt some of the claims that have been made in the recent literature about the roots of locality.

To state these ideas somewhat more specifically, I'll assume a basic recursive procedure that derives headed multi-dominance structures:²⁹

(42) Recursive Definition of a language with multi-dominance

Let LEX be a set of lexical items.

The set of potential phrases projected from LEX is the minimal set L such that

- a. **Tokens of lexical items are phrases:** $\forall l \in \text{LEX} \forall n \in \mathbb{N} \langle l, n \rangle \in L$
- b. **External Merge:** If $X \in L$ and $Y \in L$ ($X \neq Y$), $\text{Merge}_X(X, Y) \in L$
 (* $\text{Merge}_X(X, Y)$ is a projection of X with two daughters X and Y*)

The definition in (42) characterizes an infinite set of phrases, L. For each lexical item, a token of that lexical item is a phrase in L. And for any two phrases a node that has the two phrases as daughters with one phrase determining the label is a phrase as well. My second claim (which I think follows from maximally simple ways of stating mappings from structure to linear order) is that multi-dominance structures cannot be linearized unless all but one of the positions of every lexical item is ignored in the linearization process (Nunes 2001, 2004, Johnson 2012, 2016).³⁰ My third claim is that such phonological neglect is parasitic on agreement, which I assume (following Chomsky 2000, 2001 and subsequent literature) is a relationship between lexical items (heads) bearing a certain specification of features (and subject to locality, intervention and perhaps PIC). Finally, my fourth claim is that the structural configuration that triggers phonological neglect (for which agreement is crucial) also triggers the required interpretive process, namely TC.

²⁹ Whether or not projection/labeling is determined by a separate algorithm (as in e.g. Chomsky 2013) is not critical to my current purposes.

³⁰ My proposal is in principle compatible with the possibility that certain constructions involving multi-dominance can be linearized without any phonological neglect – as suggested by Wilder (1999) for RNR (See also Bachrach and Katzir, to appear). See O'brien (2017) for a proposal that might be compatible with what I am entertaining here.

Consider a simple case in which a head X^0 merges with a phrase W ($=[\dots YP_1\dots]$) and X^0 projects, forming X' :

$$(43) \quad X' = [X^0 [\dots YP_1\dots]]$$

Assume that now X^0 enters into an agreement relation with the head of YP_1 , Y^0 . In Chomsky (2000) this is the configuration that allows internal merge of YP_1 with X' . But if (42) is the basic recursive procedure, internal merge has no special status and is allowed to apply independently of agreement. What I would like to suggest, then, is that agreement in (43) – although not required for internal merge to apply – is necessary for its successful linearization. Moreover I would like to state things in such a way that would also allow for the generation of an LM structure such as (6)'. Specifically, I suggest that agreement between X^0 and Y^0 is the trigger for phonological neglect: if X^0 agrees with Y^0 in (43), Merge of any projection of Y^0 with X' will be followed by phonological neglect of various positions (by a principle of chain pronunciation that will be spelled out).

So assume that X' is merged with YP_2 which is a projection of Y^0 :

$$(44) \quad [YP_2 [X' X^0 [\dots YP_1\dots]]]$$

There are two possible scenarios to consider, one in which YP_2 is identical to YP_1 (simple internal merge) and one in which it isn't (among other things, cases of LM). For each such scenario, we need to consider two forms of phonological neglect, which correspond to what is traditionally termed overt and covert movement.

Let S be a structure in the set L defined in (42). Every phrase X dominated by S is associated by a set of “positions” or “occurrences”. Following Collins and Stabler (2016, section 4), a position of a phrase, X , in a structure, S , is a path that goes from X to S , that is, a sequence of phrases $\langle a_1, \dots, a_n \rangle$ where $a_1 = X$, $a_n = S$ and a_{j+1} is a mother of a_j , for any j . I'll assume that linearization is a mapping from a set of positions to a linear order of the terminals.³¹ A phrase can be linearized without contradiction only if at the point of linearization every position of a terminal node but one is deleted (or neglected), but see note 30.

What I've proposed above is that when (44) is formed, agreement between Y^0 and X^0 triggers this required phonological neglect:

$$(45) \quad \text{Let } X' \text{ be a complex phrase (not a head). If in } \text{Merge}_{X'}(X', YP_2) \text{ the head of } X' (X^0) \text{ agrees with the head of } YP_2 (Y^0), \text{ phonological neglect is triggered.}^{32}$$

³¹ Let a_1 and b_1 be two lexical items in a structure in the set L defined in (42) with positions $\langle a_1, \dots, a_n \rangle$, $\langle b_1, \dots, b_m \rangle$. It is easy to see that $\exists j \leq n, \exists k \leq m \forall i \in [j, n] \forall i' \in [k, m] [a_i = b_{i'}]$. This means that traditional statements determining the relative ordering of sisters (e.g. a head precedes/follows its complement) can yield the relative ordering of lexical items based on positions (as long as every lexical item has just one position).

³² Johnson (2012), Sportiche (in press) and Fox and Johnson (2016) assume that restrictors of quantifiers get to their position by movement. This can be captured in the system entertained here if the head of QP , Q^0 , agrees with the head of its NP complement N^0 , and if agreement is taken to be an equivalence

Phonological neglect occurs whenever X' is merged with a phrase YP_2 and the head of X^0 agrees with Y^0 . If a necessary condition for agreement is c-command, the relevant configuration must involve sharing³³ of Y^0 (as in (44)) and if sharing can't be linearized as is, we can think of (45) as a local method for determining that neglect is required.

What we need, next, is a definition of phonological neglect. As mentioned, I would like to suggest that neglect involves the phonological deletion of various positions of terminal nodes and, more specifically, that there are two forms of neglect, one corresponding to “overt movement” and the other to “covert movement”:

- (46) Let $[YP_2 [X' X^0 [...YP_1...]]]$, be a phrase, in which X^0 agrees with Y^0 , the head that is shared by YP_1 and YP_2 .³⁴

Overt-Neglect:

For every terminal, t , dominates by YP_1 and for every position of t , P , if $X' \in P$, delete P .

Covert-Neglect:

For every terminal, t , dominates by YP_1 and for every position of t , P , if $X' \notin P$, delete P .

Finally – though this is not crucial for anything I've said here – one might combine the above with a proposal that David Pesetsky and I made in unpublished work (Fox and Pesetsky 2009). We proposed that the distribution of overt and covert movement is dependent on the way that X' and YP_2 are linearized (in a given language), specifically that phrases are linearized based on the position that would place them furthest to the left.³⁵ This idea, which is designed among other things to account for the fact that extraposed NP modifiers appears on the right, can be incorporated to the current setting with the following statement:

- (47) Let $[YP_2 [X' X^0 [...YP_1...]]]$, be a phrase in which X^0 agrees with Y^0 , the head that is shared by YP_1 and YP_2 :

If X' follows its sister:

For every terminal, t , dominates by YP_1 and for every position of t , P , if $X' \in P$, delete P .

If X' precedes its sister:

relationship (e.g. feature sharing) : Q^0 agrees with N^0 , the head that under Sportice's account triggers NP movement (say T^0 agrees with N^0) and subsequently with Q^0 .

³³ X and Y *share* A if they both dominate A and neither dominates the other. If in $\text{Merge}_X(X', YP_2)$ the head of X' (X^0) agrees with the head of YP_2 (Y^0), it follows that X' and YP_2 share Y^0 .

³⁴ A fully explicit statement of these definitions would have to specify how the various players (X^0 , Y^0 , YP_1 and YP_2) are identified. I hope that the reader can fill in the necessary pieces.

³⁵ Among the considerations we entertained (besides the fact that extraposition appears on the right) is the observation that languages like English don't have multiple specifiers on the left and SOV languages appear to have scrambling and scope rigidity.

For every terminal, t , dominates by YP_1 and for every position of t , P , if $X' \notin P$, delete P .

Finally, we can also think of the configuration in (44) as the trigger for TC:

- (48) Let X' be a complex phrase (not a head)³⁶. If in $\text{Merge}_{X'}(X', YP_2)$ the head of X' (X^0) agrees with the head of YP_2 (Y^0), Interpret $\text{Merge}_{X'}(X', YP_2)$ by Trace Conversion.

- (49) **Trace Conversion:** $\text{Merge}_{X'}(X', YP)$ is interpreted by TC, as follows

$$\llbracket \text{Merge}_{X'}(X', YP) \rrbracket = \llbracket YP \rrbracket \lambda z. \llbracket X'(i/Y^0) \rrbracket^{i \rightarrow x}$$

where i is a numerical index that does not appear in X' and $X'(i/Y^0)$ is a modification of X' differing only in that the highest position of the head of YP in X' bears the index i .³⁷

This together with an appropriate semantics for indexed determiners (along the lines of e.g. Elbourne 2005)³⁸ derives the desired results.

- (50) $\llbracket D_i \rrbracket^g = \lambda P: P(g(i))=1. g(i)$

15. Sample Derivations

In this final section, I would like to go through a few sample derivations involving both overt and covert movement, with and without LM. In doing so I will be making a few choices that are not crucial for my purposes while ignoring properties of derivations that are not pertinent (e.g. successive cyclic movement through the edge of spell-out domains).

15.1. Overt Movement

Consider a simple constituent question such as (*Guess*) *Which book Mary read*. One step in the derivation of such a question involves an interrogative C^0 merging with TP and agreeing (under c-command) with the head of the *wh*-P (*which boy*) dominated by TP. The next step involves Merge of C' with the same *wh*-Phrase, as in (51) (“internal merge”). Since the external specifier of C^0 is linearized to the left of its sister (at least in English), we have overt-neglect: all positions of the terminals dominated by *which boy*

³⁶ I will not be considering here the possibility that $X' = X^0$.

³⁷ A head h is higher than h' if the first phrase that dominates h and has a different label (i.e. is a projection of a different head) dominates h' .

³⁸ If the proposal in Fox and Johnson (2016) is correct, the entry in (50) would need to be modified in ways that should not affect what is discussed here.

(paths up the tree) that contain C' are deleted, and *which boy* is, consequently, linearized at the beginning of the sentence.

- (51) [which book [C' C⁰ [TP Mary read which book]]]

Linearization:

Which book Mary read.

Suppose, instead, that what merges with C' is not *which boy* but an entirely distinct phrase the head of which agrees with C⁰, say *which novel*, as in (52).³⁹

- (52) [which novel [C' C⁰ [TP Mary read which book]]]

Linearization:

**which novel Mary read.*

Since here, just as in (51), C⁰ agrees with *which* (and since the specifier of CP is linearized here on the left), all the positions of the terminals dominated by *which book* that contain C' will be deleted, and since the terminal *book* doesn't have an additional position, this terminal will fail to be linearized. So if we adopt a principle that requires every terminal to be linearized, we can block a derivation of this sort:

- (53) **Full Interpretation (phonology):** At the point of linearization, every terminal must be associated with at least one position.

More generally, (53) ensures that in the case of overt movement, the “head of the chain” will dominate all of the terminals dominated by the tail. But the head of the chain can contain additional material, as in the simple cases of LM identified by Lebeaux (or in more complicated cases involving embedded late merge):

- (54) [which book that Fred bought [C' C⁰ [TP Mary read which book]]]

³⁹ Assuming that every novel is a book, the presupposition triggered at the tail of the chain will be satisfied. If not, other examples can be chosen to make the same point.

Linearization:

Which book that Fred bought Mary read.

15.2. Covert *wh*-movement

Consider now a multiple *wh*-question in English (e.g. *which girl read which book yesterday*) and assume that there is a step in the derivation in which the interrogative C^0 agrees with both *wh*-phrases in its scope. Assume further that next two steps in the derivation involve internal merge of the two *wh*Ps as multiple specifier of C^0 (respecting their original c-command relationship as seen overtly in Bulgarian, Rudin, Boskovic, Richards).^{40,41}

- (55) [which girl [C^* [which book [C' C^0 [$_{TP}$ which girl read which book yesterday]]]]

Linearization:

Which girl read which book yesterday.

Since the external specifier of C^0 is linearized to the left of its sister C^* , we have overt-neglect: all positions of the terminals dominated by *which girl* (paths up the tree) that contain C' are deleted, and *which girl* is, consequently, linearized at the beginning of the sentence. Following Fox and Pesetsky (2009), I assume that internal specifiers in English (in contrast to Bulgarian) are linearized to the right of their sister. We thus have covert-neglect: all positions of the terminals dominated by *which boy* that do not contain C' are deleted, and *which boy* is, consequently, linearized in situ.

So this is the simple case of covert internal merge. But what happens if we try to merge different *wh*Ps in [Spec, CP]. In order for two position to be treated as a chain, I have proposed that that they must share a head (the head C^0 must agree – under c-command – with the head of the specifier). So let's consider a case where the element in [Spec, CP] is distinct from the element in the base position but shares a head with it, say late merger. This gives us another way of talking about the Fox and Nissenbaum derivation of extraposition:

- (56) [which girl [C^* [**which book that Fred bought** [C' C^0 [$_{TP}$ which girl read **which book** yesterday]]]]]

⁴⁰ C' and C^* are intermediate projections of the same head C^0 . I use different symbols to so that I can refer to each separately.

⁴¹ I will say nothing here about the principles that constraint the relative c-command of the multiple specifiers ("Tucking-in" order). All I can do at the moment is express the hope that the correct account of this effect will be compatible with the outlines of my proposal.

One speculation that I hope to investigate further is based on the idea that when interrogative C has multiple specifiers [C $_{TP}$] has two binder indices appended to it (and is interpreted as a two place function $\lambda x. \lambda y. f(x)(y)$, as in Barker 2007, Beck 2000, Nissenbaum 2000 and Sauerland 1998b). This would require a rephrasing of the TC rule in (49). But it will have the advantage of allowing us to derive tucking in from a more general requirement that the relationship between the binder indices and the heads they bind will involve nested (rather than crossing) dependencies.

Linearization:

Which girl read which book yesterday that Fred bought.

The word order is that of extraposition: all terminals dominated by *which book* have two positions and are pronounced in the base position (since the other positions would lead to linearization further to the right, given Fox and Pesetsky's assumption, and as we see by a consideration of where the relative clause is pronounced); the terminals dominated by the relative clause have only one position and are linearized based on this position.

Finally we need to consider cases where [Spec, CP] is even less similar to the base position, namely the covert analog of (52), e.g. (57). A structure such as (57), in contrast to (52), is not ruled-out by phonological consideration of Full Interpretation – (53). Furthermore it is interpretable by TC (see note 39).

- (57) [which girl [_{C*} [which novel [_{C'} C⁰
[_{TP} which girl read which book yesterday]]]]

Linearization:

**Which girl read which book yesterday novel.*

If what I'm suggesting here is correct, there must, therefore, be another way of ruling out such a structure. To this effect, I would like to point out that Takahashi and Hulsey's (2009) principle that rules-out late merger of NPs in A-bar movement (what they've called whole sale late merge) could serve as the missing piece. Specifically, Takahashi and Hulsey's principle would rule out (57), where the NP *novel* is not case-licensed.

15.3. Covert QR

Covert QR should work just like covert *wh*-Movement. But if David Pesetsky and I are correct about the overt/covert distinction, the high position of a covertly moved QP must be linearized to the right of its scope (else the movement would be overt). We proposed that this holds because covertly moved QPs, just like covertly moved *wh*Ps, are internal specifiers (and internal specifiers are more generally linearized on the right in English). Specifically we claimed that QR involves Merge of a QP dominated by VP (or, here, a constituent that shares a head with it) as an internal specifier of TP.⁴² Before this can happen (under the set of assumptions entertained here), T⁰ must agree with the QP. So I will assume that T⁰ can be enriched to contain a quantificational feature (as in Chomsky 1995, chapter 4). We, thus, have the structure in (58) in which T⁰ agrees with *every*, after which T' can merge with *every book (that Fred recommended)*, yielding (59) before the subject *Mary*, which agrees with T⁰ in phi features, is merged in the external specifier position:

- (58) [_{T'} T⁰ [_{VP} Mary read every book]]]
(59) [every thing [_{T'} T⁰ [_{VP} Mary read every thing]]]

⁴² This assumption is not crucial for our current purposes, but it does derive the covertness of QR along with a few additional properties (see Fox and Pesetsky 2009).

- (60) [Mary [_{T*} every book that Fred recommended [_{T'} T⁰ [_{VP} Mary read every book yesterday]]]]

Linearization:

Mary read every book yesterday (that Fred recommended).

If internal specifiers are in general linearized on the right of their sisters in English, the higher positions of the terminals dominated by *every book* will be deleted (the positions that do not contain T'), and the word order of extraposition would be derived.

References

- Bachrach, Asaf and Roni Katzir. 2009. Spelling out QR. *Proceedings of Sinn und Bedeutung* 11, E. Puig-Waldmüller (ed.), Barcelona: Universitat Pompeu Fabra, pp.63-75.
- Bachrach, Asaf and Roni Katzir. To appear. Linearizing Structures. *Syntax*.
- Barker, Chris. 2007. Parasitic Scope. *Linguistics and Philosophy* 30 (4), 407–444.
- Beck, Sigrid. 2000. Star Operators. Episode 1: Defense of the Double Star". In Kiyomi Kusumoto and Elisabeth Villalta (eds): *UMass Occasional Papers in Linguistics* 23: Issues in Semantics. Amherst, MA: GLSA, 1-23.
- Bhatt, Rajesh, and Roumyana Pancheva. 2004. Late Merger of Degree Clauses. *Linguistic Inquiry* 35: 1-45.
- Bošković, Željko. 1997 "Superiority effects with multiple wh-fronting in Serbo-Croatian" *Lingua* 102: 1-20.
- Bošković, Željko. 1998. Multiple wh-fronting and economy of derivation. *Proceedings of the West Coast Conference on Formal Linguistics* 16, 49-63. Stanford: Stanford University
- Chomsky, Noam. 1995. *The Minimalist Program*, Cambridge, MA: MIT Press.
- Culicover and Rochement. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge University Press.
- Chomsky, Noam. 2000. Minimalist inquiries. In R. Martin, D. Michaels and J. Uriagereka. (Eds.) *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. (pp. 89-156.) Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz. (Ed.) *Ken Hale : A life in language*. pp. 1-52. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2013. Problems of projections. *Lingua* 130. 33-49.
- Citko, Barbara. 2005. On the nature of merge: External merge, internal merge, and parallel merge. *Linguistic Inquiry* 36: 475–496.
- Collins, Chris and Edward Stabler. 2016. Formalization of Minimalist Syntax. *Syntax*.19:1, March 2016, 43–78
- Dayal, Veneeta. 1998, ANY as inherently modal, *Linguistics and Philosophy* 21, 433–476.
- de Vries, Mark. 2005. Merge: Properties and boundary conditions. *Linguistics in the Netherlands* 22. 219–230.

- Elbourne, Paul. 2005. *Situations and Individuals*. MIT Press. Cambridge, MA.
- Fox, Danny. 1999. Reconstruction, Variable Binding and the Interpretation of Chains. *Linguistic Inquiry* 30: 157–196.
- Fox, Danny. 2000. *Economy and Semantic Interpretation*. MIT Press.
- Fox, Danny. 2001. The Syntax and Semantics of Traces. Talk presented at the University of Connecticut linguistics colloquium.
<http://lingphil.mit.edu/papers/fox/UCONN.pdf>.
- Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.
- Fox, Danny. 2003. On Logical Form. In *Minimalist syntax*, ed. R. Hendrick, 82–123. Oxford: Blackwell.
- Fox, Danny, and Jon Nissenbaum. 1999. Extraposition and scope: A case for overt QR. In *WCCFL 18*, Sonya Bird, Andrew Carnie, Jason D. Haugen, and Peter Norquest, eds., 132–144. Somerville, Mass.: Cascadilla Press.
- Fox, Danny, and Jon Nissenbaum. 2000. Extraposition and Covert Movement. Talk Presented at New York University. <http://lingphil.mit.edu/papers/fox/NYU.pdf>
- Fox, Danny & Kyle Johnson. 2016. QR is restrictor sharing. In Kyeong-min Kim, Pocholo Umbal, Pocholo Block, *West Coast Conference on Formal Linguistics*, 1–16. Somerville, MA: Cascadilla Proceedings Project.
- Fox, Danny, and David Pesetsky. 2009. Rightward movement, covert movement, and cyclic linearization. Talk delivered at Ben Gurion University, July 2009.
<http://web.mit.edu/linguistics/people/faculty/fox/Ben-Gurion-7-09.pdf>
- Guéron, Jacqueline. and Robert May. 1984. Extraposition and Logical Form. *Linguistic Inquiry* 15: 1-31.
- Hulsey, Sarah and Uli Sauerland. 2006. Sorting out relative clauses. *Natural Language Semantics*, 14:111–137.
- Heim, Irene. 2012. Functional Reading without Type Shifted Nouns. Unpublished ms. MIT. <http://semanticsarchive.net/Archive/TUwYzUwZ/>.
- Johnson, Kyle. 1991. Extraposition from NP is extraposition of NP. In *Proceedings of the Formal Linguistics Society of MidAmerica II*.
- Johnson, Kyle. 2012. Toward deriving differences in how *Wh* movement and QR are pronounced. *Lingua* 122:529–553.
- Johnson 2016. Towards a Multidominant Theory of Movement. Talk Presented at University College London.
http://people.umass.edu/kbj/homepage/Content/Multi_Movement.pdf
- Katzip Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy*. 30:669–690.
- Lebeaux, David. 1988. *Language acquisition and the form of grammar*. PhD thesis, UMass Amherst.
- Lebeaux, David. 1991. Relative Clauses, Licensing and the Nature of Derivations. In *Phrase Structure, Heads and Licensing*, S. Rothstein and M. Speas eds., *Syntax and Semantics* 25, Academic Press.

- Lechner, Winfried. 2015. The Syntax - Semantics Interface. In Alexiadou, Artemis and Tibor Kiss (eds.), *Syntax - An International Handbook of Contemporary Research*. Berlin/New York: Mouton de Gruyter, 1199 - 1256.
- Magri, Giorgio. 2009. A Theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, 17.3: 245-297.
- Magri, Giorgio. 2011. Another argument for embedded scalar implicatures based on oddness in downward-entailing environments. *Semantics & Pragmatics*, 4.6: 1-51.
- Magri, Giorgio. To appear. Blindness, Short-sightedness, and Hirschberg's contextually ordered alternatives: a reply to Schlenker (2012). To appear in *Linguistic and Psycholinguistic Approaches on Implicatures and Presuppositions*; Salvatore Pistoia-Reda and Filippo Domaneschi (eds.); Palgrave Macmillan.
- Mayr, Clemens, and Benjamin Spector. 2010. Not too strong! Generalizing the Scope Economy Condition. In *Proceedings of Sinn und Bedeutung 14*, ed. by Martin Prinzhorn, Viola Schmitt, and Sarah Zobel, 305–321. Vienna.
- Meyer, Marie-Christine. 2013. *Ignorance and Grammar*: MIT PhD dissertation.
- Meyer, Marie-Christine. 2014. Deriving Hurford's constraint. In T. Snider, S. D'Antonio & M. Weigand (eds.), *Semantics and Linguistic Theory (SALT)*
- Nissenbaum, Jon: 2000, *Investigations of Covert Phrase Movement*, Doctoral Dissertation, MIT, MITWPL, Cambridge.
- O'brien, Chris. 2017. Compositionality and the linearization of multidominant structures: Deriving the right-edge restriction on RNR. Unpublished ms. MIT.
- Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry* 32: 303–344. DOI: <http://dx.doi.org/10.1162/00243890152001780>
- Nunes, Jairo. 2004. *Linearization of chains and sideward movement* (Linguistic Inquiry Monographs 43). Cambridge, MA: MIT Press.
- Partee, Barbara 2009. The dynamics of adjective meaning. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue 2009"* (Bekasovo, May 27-31, 2009), ed. A. E. Kibrik et al, 593-597. Moscow: Russian State Humanities University. [[PDF of proceedings paper](#)]
- Richards, Norvin 1997. *What moves where when in which languages?* PhD. Dissertation. MIT.
- Richards, Norvin. 1998. The Principle of Minimal Compliance. *Linguistics Inquiry* 29. 599-629.
- Rudin, Catherine. 1988. On Multiple Questions and Multiple WH-fronting. *Natural Language and Linguistic Theory* 6(4):445-501.
- Romoli, Jacopo. 2015. A structural account of conservativity. *Semantics-Syntax Interface* 2(1): 28-57
- Rullmann, Hotze & Sigrid Beck. 1998. Presupposition projection and the interpretation of which-questions. *Proceedings of SALT* 8.
- Ruys, Eddie, and Yoad Winter. 2011. Scope ambiguities in formal syntax and semantics. In: Dov Gabbay and Franz Guenther (eds.), *Handbook of Philosophical Logic*, Volume 16, 159-225. Amsterdam: John Benjamins.
- Schwarz, Bernhard. 2001. Two kinds of long distance indefinites", unpublished manuscript, Stuttgart University. <https://4559e6bf-a-62cb3a1a-s->

sites.googlegroups.com/site/bernhardschwarz2003/indefinites.pdf?attachauth=ANoY7cp3MmeI9VA425PLPVaNSF2QDyIo4k6JtLwyqPlyZpjucRTp5jvzJ39625z-QajMMh9ihEh5Z017rACFV1tQgluQ51Tbyehg4gdX-yXwwOQhou2PF-xI0bLw6r7gEizxXTvoPSR2UuC-4tt0o51oVL65J-s396Ko_vRXTYw4GtHaHU_0cT_opMCEIJtGGQDnordf4QHK_OieiSf4s7s42_Cd0TGtLIMkE_D7CZLQBTWpzVF-QPI%3D&attredirects=0Sportiche, Dominique. 2015. Neglect. Unpublished ms. UCLA.

- Sauerland, Uli. 1998a. *The Meaning of Chains*, PhD dissertation. MIT.
- Sauerland, Uli. 2001. A Contrast to a Trace', in K. Megerdooimian and L. Bar-El (eds.), *Proceedings of WCCFL 20*, pp. 498–509. Cascadilla Press, Somerville. 2004).
- Sauerland, Uli: 1998b, 'Plurals, derived predicates, and reciprocals', in U. Sauerland and O. Percus (eds.), *The interpretive tract*, MIT Working Papers in Linguistics, Cambridge, MA: 177–204.
- Sportiche, Dominique. 2015. Neglect.
- Sportiche, Dominique. In press.
- Tada, Hiroaki. 1993. *A/A-bar Partition in Derivation*. PhD dissertation. MIT.
- Takahashi, Shoichi, and Sarah Hulsey. **2009**. Wholesale late merger: beyond the A/A distinction. *LinguisticInquiry* 40:387-426.
- Taraldsen, Knut Tarald. 1981. The theoretical interpretation of a class of marked extractions. in A. Belletti, L. Brandi and L. Rizzi (eds.), *Theory of Markedness in Generative Grammar: Proceedings of the 1979 GLOW Conference*, Scuola Normale Superiore di Pisa: 476-516.
- Williams, Edwin. 1974. *Rule Ordering in Syntax*. PhD dissertation. MIT.