The representation of syntactic action at a distance

Abstract

It is a common understanding that Merge effectively explains the preponderance of displacement in language. That is, at least since Chomsky 2001, the mechanism that captures displacement (Internal Merge) has been recognized as something that comes 'for free' along with Merge. However, the particular representation of that displacement has been subject to disagreement with some researchers assuming a copy-theoretic view and others a multidominance view. In this paper I show that previous means of adjudicating between the two do not succeed and present two novels means of distinguishing them. The previous arguments comparing the two representations fail on one of two counts. They either 1) rely on interface-dependent notions about which too little is known to be used to distinguish the two or 2) depend on issues of mathematical power that are not a priori relevant. The new arguments presented here rely on syntax-internal notions and interface notions that are on more solid empirical footing. The results of these clearer means of adjudication work out in favor of the copy theory over multidominance.

Keywords: movement, copy theory, multidominance

1 Introduction

Displacement is of course ubiquitous in human language. Linguistic elements are often expressed in positions temporally distal from the words that they are most closely associated with in meaning. In short, meaning does not always straightforwardly follow form. A simple case of this can be seen in the English topicalization construction in (1). Here, *carrots* appears apart from the verb *like* despite sharing a close intuitive relation with that verb.

(1) Carrots, I like.

But to note that displacement is commonplace is to use a loaded term: it suggests evacuation from a proper or unmarked original position. This is not obviously the case and must instead be argued for. The stance implicit in the term was most famously posited by Chomsky (1955,1957) when he first argued for a transformational grammar for human language. He captured the intuitive relation of *carrots* and *like* in (1) by doing a simple thing: he put them together. In less glib terms, Chomsky posited there is a representation causally related to (1) in which the two relevant terms in fact *were* right next to each other and comprised a unit unto themselves. This related representation is baldly sketched in (2) below.

(2) I like carrots.

As is well known, Chomsky argued for a transformational relation between (1) and (2) not out of brute necessity, but out of a drive for a more enlightening explanatory theory.[1] He contended that non-transformational approaches were

[1] Chomsky clearly states that non-transformational accounts of the sentence in (1) are possible and indeed entire productive schools of thought have taken this other route (Johnson and Postal

certainly powerful enough to capture displacement effects, but that they did so in a theoretically unexplanatory fashion. He saw transformations as a promising first step towards a more concise theory of displacement.

But the mere adoption of transformations is no exercise in etiology. It does nothing to shed light on *why* the effects of displacement arise in the first place. Once such concerns came to the fore (Chomsky 1993,1995), transformations and the displacement they effect became a source of puzzlement (see Chomsky 1997). Why should there be displacement when the semantic effects in question could have been taken care of in situ? Why should we expect displacement in the first place?

This puzzlement was relatively short-lived. It became clear that a distinct displacement operation could be reinterpreted as the anodyne result of the basic means of combining two linguistic elements in the first place: Merge (Chomsky 1995). Under this view, displacement is not a surprise but an expectation.[2] Simply put: if there is an operation merge that can take two elements and combine them and the result can in turn be input to that same operation, then it would take a stipulation to the contrary to rule out combining an element with a sub-part of itself. If merge can take A and B to create C as in (3) and C is an object that is subject to the same operation that created it, we should expect it to be able to take B and C to make D as in (4).

(3) Merge(A,B) $\rightarrow_c$ [A B]

1980, Gazdar, Klein, Pullum, and Sag 1985, Pollard and Sag 1994, Bresnan 2001, etc.). In this paper I take transformational grammar as a premise. In particular, I take Merge as an explanation of displacement as a premise (cf. Neeleman and van De Koot 2010)

[2] As a reviewer notes, this cannot be the whole story. In contemporary theory, something is required to spur the internal merge to begin with (say a feature of some type). Without such motivation, the effects of internal merge would never be apparent.

(4) Merge(B,C) → $_{D}$[B $_{C}$[A B]]

In (4) we find our elusive displacement arising without special theoretical dispensation. There is a representation of B that is structurally distant from its original location and depending on when an operation like that in (4) occurs, a representation of B could end up very far from its original position indeed.

This is an important step forward, but the representation in (4) obscures a deep resultant puzzle. There is still a sense in which merge fails to fully account for displacement phenomena. In (4) there are two instances of B equally represented. This parity of representation is clearly not the case in most instances of displacement. For instance, in (1) there is only one overtly instantiated instance of *carrots*. There needs to be some explanation of how we go from two instances of *carrots* at some level of representation to one instance of *carrots* in another.

There are two logically possible explanations. We must either treat positions or elements in positions differently. One option, there are two instances of *carrots* and the lower, original instance is somehow rendered phonologically null. Another option, there is truly only one instance of *carrots* and it is expressed in its most recently derived position. Both of these options have been pursued in the past and I attempt to arbitrate between them here.

The driving question is, in other words, what is the result of undergoing merge twice? Does it result in two objects in their own position or one object in two positions? In this paper I argue that displacement via merge results in two objects each in their own position. Merge does not result in a single token in multiple positions.

Much like Chomsky's original argument in favor of transformations, arguing for this position relies on theoretical concerns as well as the ability for the theory to capture the data in enlightening ways. Basic empirical facts and issues of formal capacity are not suited to adjudicate the two options. Both approaches equally capture the basic empirical facts in principle and differences in their formal power are not clearly conclusive. Instead, it is the shape the theory necessary to account for the facts that will be dispositive. The theory in which there are two copies of a moving element is able to account for certain empirical matters in a more enlightening manner. Before getting into this, I discuss in the next section the basic background to this particular issue.

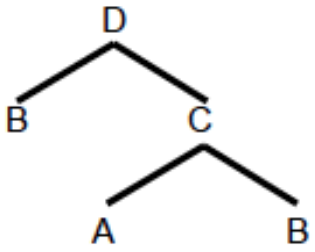
2 The issue

As introduced above, the theoretical advantage of Merge (explaining the ubiquity of displacement) concerns its application to a term that has already been subject to it. The result of this second merge *is* the displacement. So much is clear and in this section I lay out the ways this can be represented.

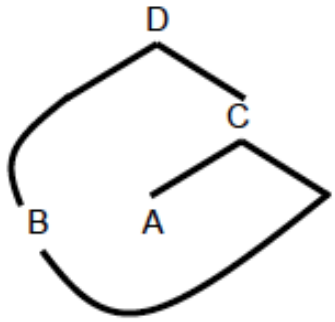
There are two logically possible ways of representing the way in which Merge effects displacement. It can create an entirely new term in what is commonly called the Copy Theory of Movement (CTM) and is discussed in Chomsky 1995 and Nunes 2001,2004 among others. This is sketched in (5) below. The other logical possibility is that no new instance of the moving element is created, but rather the displaced element is introduced into a new position without vacating its old one. This is represented in (6) below and represents Multidominance (MD) accounts as discussed by a large number of researchers including Epstein et al. 1998, Starke 2001, and Gärtner 2002 among others for simple, 'upward' movement. In (5) we find

two distinct instances of the displaced element whereas in (6) we find essentially the same structural relations represented using but one instance of the element in question.

(5)



(6)



The issue at hand is: does (5) or does (6) better represent displacement? Does displacement in the form of re-merge result multidominance or copies?

The answer to these questions will have repercussions elsewhere. The obvious import of the answer lies in the fact that the merge-based explanation of displacement is one of the major theoretical innovations of modern syntactic theory and answering the above questions will of course be inherently important for that very reason. Additionally, the choice between the two representations will be practically important in terms of the demands that the competing representations make on other parts of the grammar. CTM and MD each require different things of

the interfaces and opting for one representation over another forces the interfaces to have certain properties.

Also, in attempting to answer the questions above, I note that comparing the resultant shape of the interfaces given one representation or the other is not on its own a viable means to adjudicate between them. Too little is known about the nature of the interfaces to make a convincing principled comparison. This is especially true given the seeming scant empirical means to distinguish the two approaches in the first place. Instead, adjudication between the two must take one of two forms. One, it must be shown that one representation requires a theoretical impossibility, not merely something ungainly. Here I show that MD analyses require that Merge not operate over terms, but rather positions, which is not possible if one assumes a Merge approach to structure building. Two, it must be shown that, given empirically established interface effects, one approach cannot capture a paradigm without recourse to ad hoc stipulation. Here I argue that (anti-)reconstruction effects show that sometimes scattered interpretations of the displaced element are necessary and that MD accounts cannot handle this. Without arguments like these, our understanding of the interfaces on their own is too rudimentary at this point to serve as a means of arbitration. In section 4 I provide an instance of each type of argument.

Finally, In addition to the authors cited above who focus their attention on 'upward' MD displacement that results in c-command, there is an even larger number of researchers who focus on MD that works akin to the sideward movement of Nunes 2001,2004. That is, MD where the two positions of the element are not in a c-command relation to each other. These works include but

are not limited to McCawley 1982, Ojeda 1987, Blevins 1990, Wilder 1999, Guimarães 2004, Chung 2004, Citko 2005, de Vries 2005, van Riemsdijk 2006, Chen–Main 2006, Gracanin–Yuksek 2007, Bachrach & Katzir 2009, and Citko 2011. In this article I focus solely on 'upward' displacement and do not attempt to distinguish copy-theoretic sideward movement and sideward MD. It is already non-trivial to distinguish CTM from MD in terms of upward movement. Doing the same for sideward movement is beyond the powers of this researcher. Furthermore, the existence of sideward MD is predicated on the existence of upward MD. If the latter does not exist, then neither does the former.

In the next section I rehearse the well-known interface requirements of the competing representations and contend their differences are not dispositive.

3 How not to adjudicate

In this section I will discuss differences between the two approaches but contend that these differences are in principle incapable of adjudicating between the two in any rigorous sense. First I discuss how interface mechanisms deal with the two approaches. Second I discuss the differences in computation power required by the two approaches. The interface mechanisms that each approach requires are different, but these differences hold at a level of representation where there is little firm methodological footing. The differences in power, while provable are not necessarily relevant as has been shown repeatedly in the past.

3.1 Interface mechanisms

Let us recall what displacement effects. For both approaches, CTM and MD, displacement results in two (or more) separate yet derivationally related structural

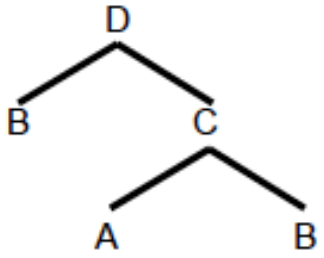
positions that each have pronounceable material in them. However, for each of these approaches, only one of the related structural positions ends up being the one with something pronounced in it. The challenge for CTM and MD is to capture how it comes to be the case that only one positions has the privilege of pronunciation.

Similarly, for both approaches the same two (or more) positions wind up with semantically interpretable material in them. Moreover, the semantically interpretable material in these positions identical as far as Merge is concerned. Merge does not alter the semantics of that which it works over. The challenge for CTM and MD here is that these two positions must somehow be distinguished semantically given the obvious fact that displaced elements frequently serve distinct semantic purposes in their distinct positions.

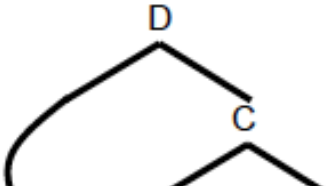
The challenges that the approaches face are extremely similar to one another. Any interface mechanism that would interpret a CTM structure will need to differentially interpret derivationally related *terms* (the atoms that comprise syntactic expressions). Any interface mechanism that would interpret a MD structure will need to differentially interpret derivationally related *positions* (the location of the term in an expression).

To see how this is achieved in action, take the toy cases in (7) below:

(7) a.



b.



A descriptively adequate interpretive mechanism for (7a) would need both a means and a motivation to disregard the lower instance of term B as far as the pronunciation is concerned. One well worked-out approach can be found in Nunes 2001,2004. This work posits that that pronunciation is determined by the number of checked features on a copy. In English, the highest copy is generally going to be the one with the most checked features and as such it will be spelled out. Crucially, this requires a mechanism that cannot allow the pronunciation of unchecked features.

In a toy example, the reasoning works as follows. A wh-word with an uninterpretable feature is merged a second time into the spec,CP position where that feature is rendered interpretable. This is shown in (8a). The idea in Nunes's system is that the copies with uninterpretable features cannot be pronounced. This demands that the lower copy not be pronounced, as seen in (8b).

(8) a. What_{F} did Becky see what_{uF}

b. What_{F} did Becky see ~~what_{uF}~~

This basic approach is what is demanded by CTM theories. They require some means of distinguishing copies of the moved element.

For MD, the interface mechanism needs to be able to distinguish the structurally lower position of B from its higher one and ignore it. Kural 2005 as well

as de Vries 2009 posit a means to achieve this goal via ‘graph traversal’, a common type of algorithm from computer science. Under this approach the sketch in (7b) (interpreted as derived graph) is traced by an algorithm that determines the linear order of the terminals based on which ones were encountered first. In the case of wh-movement in English, the tree traversal algorithm would, by stipulation, hit upon B in its higher position prior to any other. Crucially, this requires a mechanism must somehow know to ignore that lower position of B in terms of pronunciation (B is somehow marked as already pronounced upon each subsequent encounter along the traversal.).

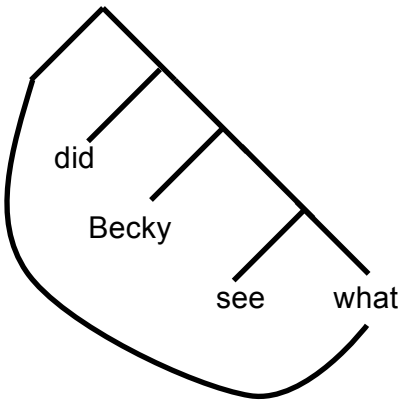
To see how this would work, let’s take the same simply example from (8) above. An example of a simple tree traversal algorithm would be one in which the parser starts at the root node and, for example, investigates left branches before exploring right branches. If the parser hits upon a terminal node, it will spell it out. Once it hits a terminal node, it will retrace its steps until it gets to a node with a right branch, in which case it will explore that path.

For the tree in (9a), starting at the root and exploring the left branch leads the parser to terminal node ‘what’. It then spells out that node, as shown in (9b). Having reached a terminal node, the parser retraces it steps up one node (which happens to be the root). Here it explores the right branch from the root node. Dropping down a node, it explores the left branch and hits upon the terminal node ‘did’, and spells it out to the right of ‘what’, again shown in (9b). It does this through the rest of the tree.

The interesting bit arises when it hits the terminal node ‘what’ a second time. We do not want the parser to spell out this node again, so a constraint is placed on the parser. A simple means of excluding this second ‘what’ is proposed in Frampton 2004 which effectively says that a given terminal node is irrelevant to linearization if

it has a parent node that dominates the local parent node. So in (9a), the lower position of ‘what’ indeed has a parent node (the root) that dominates its local parent (the node immediately dominating ‘see’ and ‘what’. As such, that position does not play a role in linearization and only the higher position is pronounced. In short, the lower position is treated differently than the higher one.

(9) a.



b. what, did, Becky, see

These two approaches are not necessarily the only possible approaches to the challenge, but they are representative and embody the minimally conceptually necessary characteristics of a successful approach. That is, they differentially treat either a term or a position. However, we are not in a position to adjudicate between CTM and MD based on a comparison of these interface strategies. Each of these mechanisms does a perfectly adequate job of meeting the fundamental empirical challenge, either via unchecked features or already-traversed elements, pronunciation will only happen once. As such, they are indistinguishable based on the basic facts.

There could be an argument that one type of mechanism or the other aligns better with the syntactic structure that feeds it. This would take a form like: tree traversal works better with MD, therefore MD. Any gesture in that direction would simply be begging the question since it is exactly the nature of the syntactic structure that is in question.

So, the basic fact that the moving element is usually only pronounced in one position is not a viable venue to tackle this issue. Instead, a comparison must be made based on whether one approach or the other is more plausible *as a PF-interface mechanism*, However, there is no currently known sense in which we should consider one mechanism more plausible in this sense. For example, one mechanism deals in terms, the other positions. Yet, there is no independent interface-based notion that compels us in either direction. Why? There is no sense in which terms and not positions (or vice versa) are the natural purview of the PF-interface. Perhaps in the future some notion will be developed, but as yet no such thing exists. Without this prerequisite, these two approaches are equal. It appears that the only realm of comparison possible is that of the empirical facts. As said before, the approaches are on equal footing there.

The same concern holds for the meaning side of things. First let us rehearse how the two approaches rise to the challenge of differential interpretation. The challenge that faces the CTM account is that two terms that are non-distinct in the syntax must somehow bear differing denotations at the semantic interface. Take the concrete example in (10). The higher instance of *which person* must be interpreted as an operator and the lower instance of it must be interpreted as a variable.

(10) Which person did you see <which person>?

Given that there are two different terms in the CTM version, it is logically possible to treat them different at the LF-interface. One way of doing this has been explored by Fox 2002 as well as Sauerland 2004 and has been dubbed 'trace-conversion' (though it of course works with copies just as well as traces). This proposed interface operation takes the structure implicated in (10) are converts the lower version into a definite description bearing a bindable variable as shown in (11) below (after λ-abstraction):

(11) Which person λx [you saw the person x]

This tactic again of course succeeds in rendering the syntactic results of CTM appropriate for the semantic interpretation they require.

The problem of differential semantic interpretation in different positions is on the surface much more difficult for MD approaches. Any approach that follows the intuitions of trace conversion would require the self-same element to be interpreted differentially.[3]

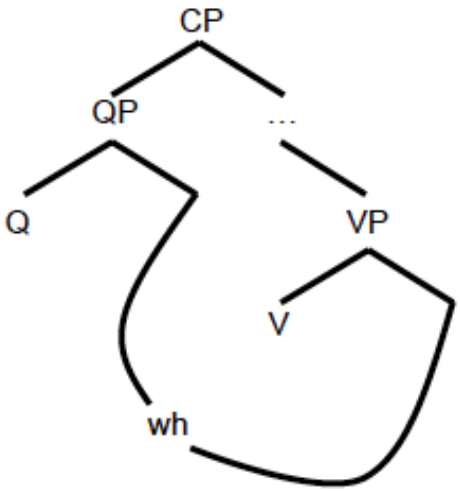
Johnson 2012 skirts this issue by denying that the displaced element needs to be so dually interpreted. Instead he posits that the wh-phrase functions solely as the variable. By analogy with the approach sketched for CTM, under Johnson's analysis the wh-phrase functions like the lower copy in (11) and something else serves as the operator. The operator rule is fulfilled by a Q-head of the sort

[3] This would be possible only if the element that is represented a single object in the syntax is 'copied' in some sense in at the LF-interface. This is of course possible, but it is not without its intuitive problems for the task at hand here. For example, as an argument for MD in the syntax, this sort of mechanism would involve some sort of inverse question begging. Instead of presuming the conclusion, it would require presuming a conclusion that does not follow from the argument. It could also be possible sub-parts of the multiply dominated element could be interpreted at different points. The degree to which this is possible for MD, it is also possible for CTM. So again, we cannot profitably distinguish the two here.

hypothesized by variety of researchers among them Hagstrom (1998), Kishimoto (2005), and Cable (2007, 2010).

Under this approach, the displaced wh-phrase does not find itself multiply dominated in its base position and spec,CP. Instead, the Q-element heads up a QP phrase in spec,CP and the wh-phrase is multiply dominated in its base position and as the complement of the QP. To clarify, this structure is given below in (12).

(12)



Again, we find two approaches that are equally well-positioned to capture the driving empirical fact: the displaced wh-phrase is involved in an operator-variable relationship. This being the case we are again in a position where it is not possible to argue for CTM or MD over one another on empirical grounds. Further, we are once again in a position where arguing for CTM or MD based on how well these approaches align with those approaches begs the question.

Given the facts as they stand here, the only legitimate way to compare CTM and MD is whether the interface-mechanisms required of these approaches are more plausible again *as LF-interface mechanisms*. The CTM approaches requires something like trace conversion and the MD version, as admitted by Johnson

requires “a theory of the syntax/semantics interface that does not force sisters to systematically combine” (Johnson 2012:20). Much like with the PF-interface question, too little is known about the nature of the LF-interface to arbitrate between these two mechanisms. This is not to say that they are in principle indistinguishable, but that they currently are.

In short, CTM and MD approaches to displacement do not in and of themselves solve the displacement problem better than their competitor. They both require interface mechanisms to capture the basic empirical facts. As we have seen, it is possible to devise such mechanisms that are both coherent and empirically adequate. However, arguments in favor of CTM or MD cannot be adduced based on these mechanisms for reasons both empirical and logical.[4]

In fact, the logic goes the other way. If it were possible to somehow conclude that CTM or MD was correct, then it would be possible to make interesting claims about what the nature of the interface mechanisms would need to be. Before that, in the next subsection I will discuss another possible means of arguing for CTM or MD, one that also ends up insufficient.

3.2 Power issues

Another means to potentially distinguish CTM from MD and in turn favor one or another is via formal computational power comparisons. In this subsection I discuss some previous instances of this sort of reasoning and argue that they too miss the mark. Much like with the previous discussion with interface mechanisms, this is not to deny the results or import of these investigations. Rather, I intend to deny that the

[4] Later I provide what amounts to an interface-based argument in favor of CTM, but crucially this argument relies on a more established generalization concerning idioms and binding principle reconstruction.

results ought to be used to determine whether CTM or MD is the correct representation. To the contrary, the results of investigations into the relative power of CTM and MD serve to make clear the unavoidable consequences of the correct representation.

As a guide to our current discussion, it is helpful to return to the issues mentioned at the outset of this paper. Before Chomsky argued in favor of transformations, he used computational power to arbitrate between two conceptions of grammar. Finite state machines were shown to be fundamentally incapable of expressing the sort of patterns found in human language. In light of this result, it was necessary to posit a conception of the grammar that was at least as powerful as a phrase structure grammar. The upshot of this story is to note that this sort of reason is clear of ambiguity or dispute. If there are patterns of human language that are inexpressible assuming CTM or assuming MD, the decision between the two will be simple.

On the other hand, the decision between CTM and MD conceptions of displacement may end up like the distinction between basic phrase structure grammars and those with the additional power of transformations. As noted above, choosing one option over the other was not an issue of formal capacity, but rather one of theory. Phrase structure grammars alone could produce the strings of human language, it is just that transformations were arguably theoretically a *better* option. If turns out that CTM and MD are equal in their capacity to express the patterns we find in human language, then of course that result will not allow a ready decision like the previous case did. It is also important to note that this sort of result also will not entail that such a decision is not possible.

This sort of equality in expressive power occurs in the comparison of a trace theory of displacement (as in Chomsky 1981) versus the CTM that supplanted it. Compare the representations in (13). In (13a) there is an inherently silent trace in the direct object position where as in (13b) there is an unpronounced copy of the moved term.

(13) a. What$_i$ did you see t$_i$

b. What$_i$ did you see <what$_i$>

It is trivially the case that these two options are of equal expressive power: any position that movement could have stemmed from could equally leave a trace or an unpronounced copy. The way to distinguish these two options is via empirical evidence and theoretical concerns. Reconstruction effects are the empirical evidence in favor for CTM (as argued in Chomsky 1993). On the theoretical side, the Inclusiveness Condition (Chomsky 1995) serve to eliminate trace theory as an viable option. Again, even if two options have the same expressive power, they can still be theoretically and empirically distinguished.

An additional note. A few years after transformational grammar was posited, Peters and Ritchie (1973) showed that transformations were not just sufficient to capture the patterns in human language, they were powerful enough to express *any* possible pattern. This immense power was widely considered undesirable (see Gazdar 1985 for example). However, Chomskyan transformational syntax continues, not because of irrational stubbornness, but due to the fact that this mathematical result is not necessarily problematic. Transformations as currently conceived have been greatly constrained so as to moot potential concerns. Moreover, the sort of theoretical explanations made possible by transformations further justify their use.

As a result, we should not necessarily take even massive differences in computational power to be dispositive in the comparison of CTM and MD.

To recapitulate the previous discussion: One, equality in terms of expressive power does not entail that linguistic distinctions (empirical or theoretical) cannot be made. Two, expressive power "overkill" is not necessarily damning. Three, insufficient expressive power is indeed dispositive.

With this background in mind, let us explore the results of some mathematical investigations into CTM and MD. One notable investigation is that of Kracht (2001). He argues CTM and MD are very similar and while not identical, they are "identical for all linguistic purposes" (Kracht 2001:527). He shows this by proving various mathematical mapping relations between the two. As it turns out, the formal mathematical properties of CTM and MD do not relevantly distinguish the two. Anything that can be stated in one can be stated in the other. Because of this, he argues, they are linguistically identical.

This should remind the reader of the trace-theory versus CTM discussion above. There too it is the case that anything that can be expressed in one can be expressed in the other. In other words, the argument that CTM and MD are linguistically identical is not necessarily true. As a general point, It should be well known that when doing cognitively- and biologically-minded linguistics mathematical considerations are not necessarily the final word. Berwick and Weinberg (1982:177) argue persuasively that "mathematical relevance need not imply cognitive relevance." However it is this implication that Kracht attempts to make.

Note that this is not to dispute the basic results of this work (though see Kepser 2010 with some criticism), but to maintain that they are not necessarily the sort of results that help us choose between CTM and MD. The goal here does not

concern formal computational equivalence or lack thereof between CTM and MD. Instead, it is whether a particular choice of encoding is 'better' with respect to syntactic theory and empirical considerations.

The notion of superior encoding relates to another issue raised in Kracht (2001) as well as Gärtner (1999). Both discuss the exponential growth in computational complexity that CTM theories entail in the limit. Kracht in particular proves that in some “worst case” scenarios there are exponentially many more CTM structures than MD structures. In other words, despite their identical ability to encode relations, there are are many more different ways for CTM to achieve this. This may be seen as a reason to reject the CTM approach, but Kracht himself warns against this and maintains that this result does not constitute an argument against CTM. Why? Well, for the same reason that the Peters and Ritchie results were insufficient to render transformations untenable. Immense potential power does not entail psychological implausibility.

In short, much like with the interface mechanisms, the results shown by Kracht do not determine whether CTM or MD is the correct one. However, again much like with the interface mechanisms, his results serve to define the repercussions of the correct choice between CTM and MD, whichever is correct in the end. Still, they do not eliminate the need for empirical and theoretical arbitration. Loosely following Kripke: There is no mathematical substitute for linguistics.

In the next section I present arguments that effectively compare CTM and MD in a way that allows us to make an informed choice between them.

4 How to adjudicate

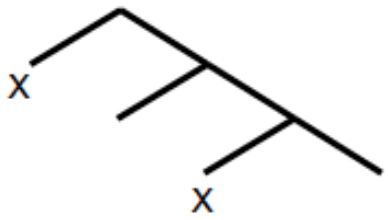
As we saw in the above sections, it is difficult to clearly determine whether CTM or MD is the correct representation of displacement. We know too little about the interfaces to support either representation over another and the mathematical concerns, while interesting, do not on their own force the adoption of one over the other. In this section I provide two means of successfully distinguishing between the two. The first involves fundamental notions of what the target of the Merge options is and the second concerns the possibility of, and constraints on, 'scattered' interpretation.

4.1 The Target of Successive Merges

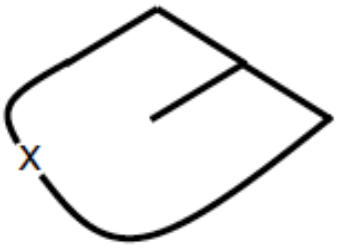
Prior to displacement, as in a structure like that in (14) below, there is only one token instance of X and as such it is at this point that CTM- and MD-style representations are identical. It is only after displacement (second Merge) has occurred that the representations differ along the lines shown in (15).

(14) [... X ...]

(15) a.

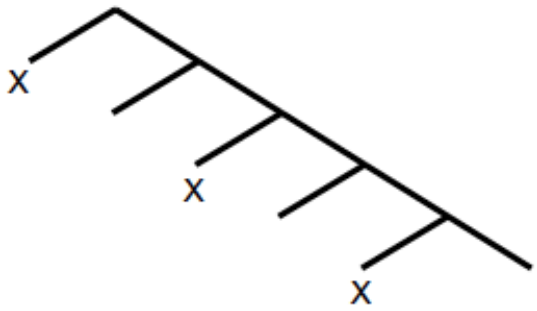


b.

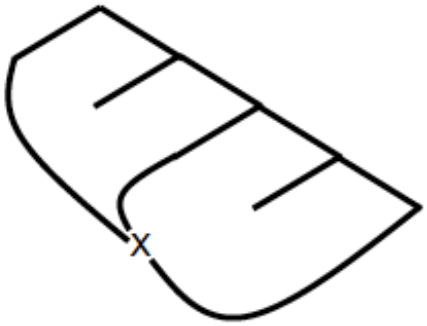


In (15a), there are now two token instances of X whereas in (15b) there is still but one. Of course displacement does not necessarily halt after one hop. Rather, displacement can happen again and again in an iterated fashion as argued by McCloskey 1979, Torrego 1984, Henry 1995, and McCloskey 2001 among others. That is, if X were to continue to be Merged into higher positions, we would be left with representations like those in (16).

(16) a.



b.



However, in (16b) it is always the case that X is *still there* in its base position even though it also finds itself much higher in the structure as well. This contrasts with CTM where there exist higher tokens of X structurally independent of the base position copy. This distinction lay at the heart of the differing interface strategies outlined in the previous section. In those instances, work went into trying to capture how CTM or MD worked with grammatically licit structures and as we saw, there was little to distinguish them by. That is, each system was powerful enough to capture

the licit structures as they pertain to interface interpretation. In this section, I employ a different method.

What other discussions neglect is how the differing interpretations of displacement fare when dealing with grammatically illicit structures and interface-independent notions. In this section, I will argue that MD is not in a position to explain certain ungrammatical sentences based on these narrowly syntactic concerns. In fact, the syntax-internal wedge that will serve to distinguish the two approaches will be Merge, the operation that prompted this discussion of CTM and MD to begin with,

The logic will be as follows. In order to capture the ungrammaticality of certain island violations, MD analyses require something that the system does not allow: the targeting of non-terms by Merge. Because Merge necessarily works over terms, this means that MD representations are incapable in principle of being used to capture island effects.

It has long been known that displacement cannot occur over an arbitrarily great structural distance in a single derivational step. To take a classic example from Ross (1967), displacement cannot occur across a complex noun phrase like in (17).

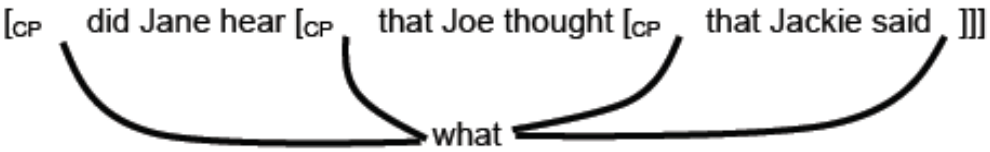
(17) *What did Jane hear the rumor that Joe saw?

Chomsky (1973 et seq.) analyzed this constraint as one of moving too great a distance. The precise formulation of the short-steps constraint has changed over the years, but the basic notion still holds: displacement cannot take place over an arbitrarily great distance.[5]

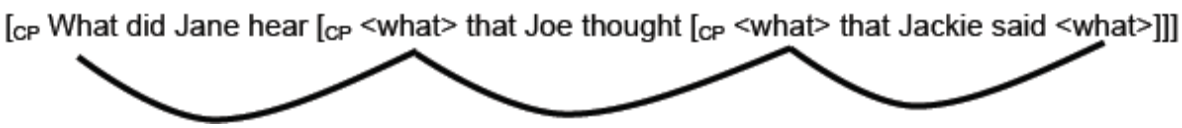
[5] A contemporary interpretation of the prohibition on movement over too great a distance is found in phase-theory (Chomsky 2001, 2008). Under phase-theory, lower structure is periodically

However, under MD accounts, movement relations in fact *do* hold over arbitrarily great distances. An MD-style representation of the sentence in (18) requires that there be a Merge-derived relation between the base position of X and and its final position. This is abstractly sketched in (19).[6]

(18) What did Jane hear that Joe thought that Jackie said?

(19) 

In (19) there is in effect a movement relation that holds across a very great distance. Contrast this with the CTM representation in (20) where there are only local relations.

(20) 

This distinction alone is not sufficient to prefer CTM over MD as it could easily be argued that somehow in virtue of the myriad shorter dependencies in (19) the very long distance dependencies are rendered licit. Perhaps the fact that a series of short hops are instantiated in (19) is sufficient. That is, in virtue of there being a licit 'route' from base position to final derived position, the representation is grammatical.

---

rendered inaccessible for grammatical operations. This causes independent problems for MD analyses as it would require that the moving element be rendered inaccessible to further operations as well: an unwanted consequence. If the moving element were somehow not rendered inaccessible, this would force a nonsensical state of affairs wherein a VP would be rendered inaccessible, but not its constituent parts.

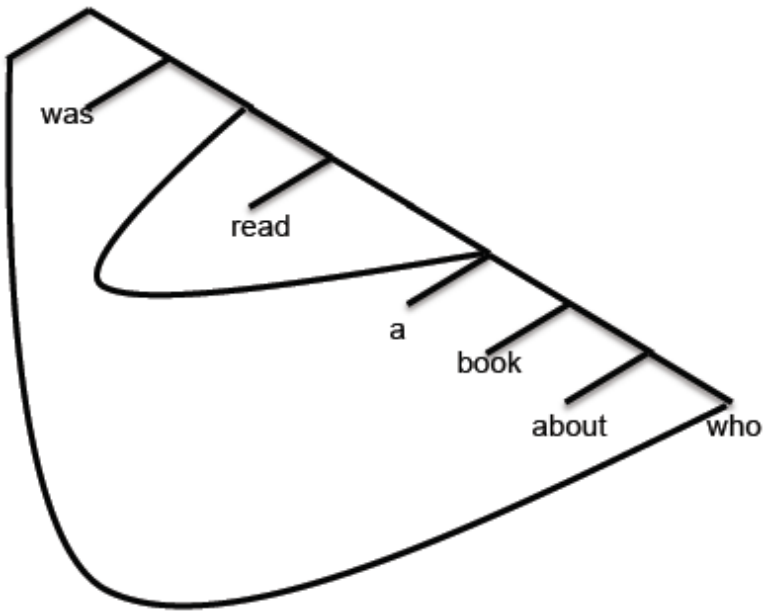
[6] I ignore here landing sites other than Spec,CP for expository purposes.

At each step in the derivation there was a position that X was in that was accessible for further movement, even though there were other position that X was in which were inaccessible. Call this the 'whatever works' interpretation.

This 'whatever works' interpretation of MD movement runs into immediate problems when looking at freezing effects like those in Wexler and Culicover (1980). Such effects are found when movement is attempted out from within an element that itself has already moved. This can be seen in (21) where the subject of the sentence has A-moved from its base position as the internal object of the verb. From within this derived subject, a wh-word has been extracted, leading to ungrammaticality. The sentence in (21) as depicted under MD would be like that in (22).

(21) *Who was a book about read (by Jane)?

(22)



Note that the 'whatever works' approach here makes the wrong prediction. Moreover, it makes the wrong prediction for CTM as well. There in fact exists a licit 'route' between the base position or copy of *who* and its final position: via the base

position or copy of the derived subject. As seen in (23), movement from within the noun phrase in object position is perfectly licit. For some reason, it is only the most recent position that the moving elements are in that matters, for both MD and CTM. This will become more relevant presently.

(23) Who did Jane read a book about?

As such, the 'whatever works' approach does not seem tenable. Another plausible approach would be to restrict operations to just between the root of the tree and 'structurally closest' instantiation of the to-be-moved element. This is a MD version of Shortest Move (Chomsky 1995): create the shortest link. For the representation in (22), the wh-word is effectively in two positions. It is within the larger DP in the lower, PP-complement position as well as within the DP in the higher, subject position. If we adopt the 'structurally closest' stricture, movement of the wh-word is necessarily assessed as if it came from the subject position. Since this closer position is an island for movement, the sentence is correctly ruled out.

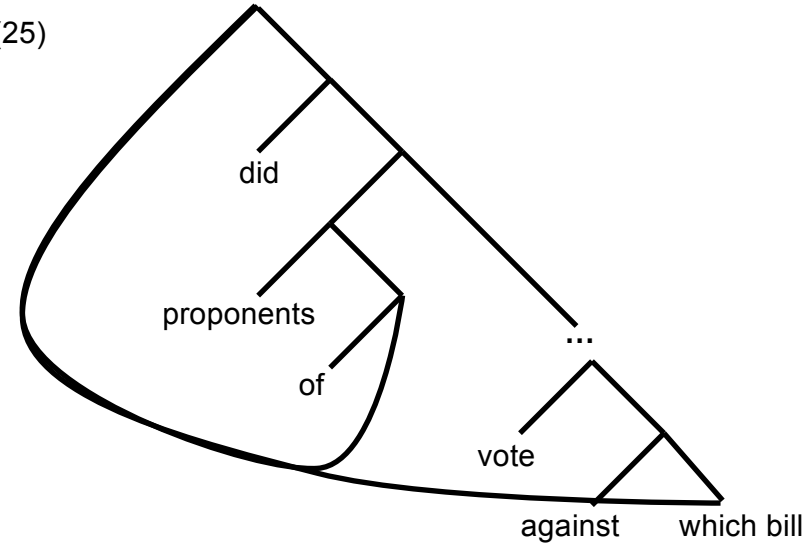
However, even the 'structurally closest' approach cannot be viable. There are instances where extraction from the structurally closest position is also illicit. Take, the following subject-internal parasitic gaps sentence.

(24) Which bill did even proponents of decide they had to vote against?

Following the approach to parasitic gaps found in Nunes 2001,2004, the above sentence is derived by the wh-word beginning its derivational life within the subject. It then sidewards moves into the complement position of *against* before making its way to the left periphery of the matrix clause (crucially before the incipient

subject is merged to the spine of the tree and thus becomes an island). The MD representation is sketched in (25) below.

(25)



In this sort of parasitic gap, the structurally highest position of the moving wh-word is an island for movement as seen in (26) below where there is no alternate position to assess:

(26) *Which bill did even proponents of arrive?

If one were to adopt the 'structurally closest' approach, the assessment of whether the movement is licit would be restricted to the subject-internal position. Since that position is island-internal, the sentence should be ruled out, contrary to fact.[7]

[7] It should be noted, as a reviewer points out, that if the sideward movement approach to parasitic gaps is not assumed, this argument cannot be made.

In sum, it's not that any licit link rules in a sentence, nor is it the case that as long as the shortest link is licit, the sentence is ruled in. Instead it seems that *recency* is what matters. For CTM it is the most recent *copy* of the moving element that can be targeted; for MD it is the most recent *position* of the moving element that can be targeted.[8]

Herein lies the problem for MD accounts. Merge does not operate over positions, but rather terms in the sense of the constants and variables that constitute sentences (Chomsky 1995). In the minimalist program, terms are taken to be elements like *the* and *dog*, but not relations between them.[9] One such relation between terms is that of sisterhood. Sisterhood can be used to identify a position, but neither the sisterhood relation or the position that it defines are terms. This will cause problems for MD accounts.

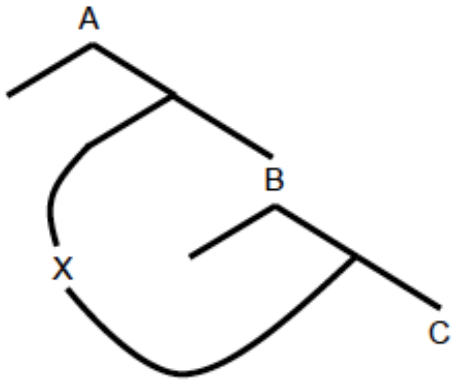
Take the MD representation in (27). Here, the only licit means for X to be Merged with A is for the position "sister of B" to be operated over by Merge. It cannot be the case that that X itself is targeted for Merge because X is also in its too distant base position. Recall that it is simply not enough to say that Merge with X is licit in virtue of there being some licit extraction site. As we saw with the freezing effects above, the existence of a licit extraction site is necessary but not sufficient. If Merge targeted X, it would in effect be targeting an inaccessible term. Only the most recent

[8] For CTM analyses, the notion that only the most recent copy is relevant for further Merge could be (and is) easily implemented in phase-theory wherein previous copies are rendered inaccessible throughout the derivation. In such a case, the most recent position can be determined and the term in that position can enter into the Merge operation.

[9] That operations target terms in positions is not a novelty of Merge. Rather, terms have always been that which grammatical operations target going back through move alpha (Lasnik and Saito 1994) on through to Chomsky 1955.

position can be operated over. The position (sister of B) is not a term and as such cannot be operated over by Merge.

(27)



Because MD requires Merge to operate over non-terms, it cannot be maintained as a possible representation of displacement. For MD accounts to work, it must be the case that Merge be altered so as to be able to operate on more than just grammatical terms.

It is perhaps an option to amend Merge such that it can operate over positions (defined in terms of sisterhood: {X,Y}, X and Y being sisters) instead of or in addition to terms. However, given Bare Phrase Structure, there is no narrow syntax distinction between terms and their positions. Previously, under X-bar-theoretic conceptions of phrase structure (Chomsky 1970 and Jackendoff 1977), structural positions were in fact reified things (Spec,XP, for instance, existed independent of the term in that position.). As such, it would theoretically possible to target a position as such. This is not the case under Bare Phrase Structure. The Inclusiveness condition (Chomsky 1995:225) prohibits novel entities like positions to

be syntactically reified and because of this, Merge can not be amended so as to be able to target them.[10]

So if it is not possible to amend Merge to save MD, perhaps Merge should be scrapped in favor of something different. This option renders the entire comparison of CTM and MD moot as it is premised upon Merge's existence. In short, successfully ruling in long-distance successive cyclic movement while at the same time ruling out sub-extraction from a moved element cannot be done by adopting both Merge and MD in principle. Contrast this with CTM implementations of displacement which are in principle capable of making those distinctions while maintaining Merge.

4.2 Scattered Interpretation

Above I presented a theoretical reason why CTM is to be preferred over MD. In this section I present what is at heart an empirical case. As seen above, such arguments are not easy to make in any decisive way. Here an interface-related argument that is not only sound, but one that could serve as a template for other interface arguments is presented. The argument relies on developing an empirical generalization that can serve as the stable fulcrum for an argument.

The basic empirical point is found in Chomsky 1993. He notes that in a sentence like (28) there is a correlation: when *Bill* is antecedent of the reflexive, the sentence is ambiguous between an idiomatic reading in which *take a picture* means *photograph* and one where it means *take possession of*. In contrast, when *John* is the antecedent the sentence only has the second non-idiomatic interpretation.

[10] Note that this does not preclude the reification of positions outside of narrow syntax at the interfaces.

(28) John wondered which picture of himself Bill took

Chomsky accounts for the contrast by positing differing LF interpretations of the string in (28). For us, the crucial interpretation is the one where *John* is the antecedent and the idiomatic reading is not possible. For such an interpretation, the LF is like that in (29). Here, the reflexive covertly adjoins to the T-head that *John* had already overtly Merged with, thus ensuring its construal with *John*.

(29) [John$_i$ [himself$_i$ +T^0] wonders [$_{CP}$ which picture of ~~himself~~$_i$ [Bill took ~~which picture of himself~~$_i$]]]

Chomsky argues that the idiomatic reading is ruled out in such an instance because of three things. One, the reflexive must move to the matrix T and be interpreted there at LF for the relevant construal with *John* to arise. Two, it is not possible to interpret a copy of *picture* without interpreting the copy of *himself* that it dominates. That is, there can be no scattered interpretation of *picture* and *himself.* Three, in the framework at hand, DS no longer exists and as such the requirement that idioms be constituents at that level is reanalyzed as holding of LF instead. Because of this, the idiomatic reading requires that *picture* (and *himself* in turn) be interpreted in its base position. This precludes the simultaneous idiomatic interpretation and matrix subject construal of the reflexive.

Chomsky is not entirely technically explicit here, but the basic interface-theoretic notion is clear: The interpretation of a given head at LF necessitates the interpretation of that which it dominates. When the head is interpreted low, so too should its dependents. It should be clear from the previous discussions that MD approaches can easily capture this interface generalization. As far as the reflexive is

concerned, one of its positions is sufficiently local to *John* and one to *Bill*. Further, one position of *picture* feeds an idiomatic interpretation and one does not. We can encode the interpretive dependences in the MD idiom as in (30).

(30) If position 1 of element X is interpreted, then all elements Y that X dominates are also interpreted in position 1.

Something like (30) prevents the scattered interpretations of constituents for MD, mimicking the analogous prohibition against scattered deletions discussed above in the context of the CTM. More particularly, (30) would prohibit interpreting the higher position of the *himself* and the lower position of *picture* in (28).

Again, we find ourselves in a position where CTM and MD are equally able to account for the data, given certain assumptions: If the head is interpreted, so too must be its complement. However, something like what we have in (30) cannot be the last word. Sometimes scattered interpretations are in fact required. An example of this can be found in what are commonly known as Lebeaux-effects (Lebeaux 1988, 1991 and Chomsky 1995). That is to say, that for the following sentences, it is still possible to get the idiomatic readings. Nevertheless, the adjunct to the wh-phrase is not interpreted low down in the idiom: There are no Principle C effects.

(31) a. Which picture that Bill_i hated did he_i say that Mary took?

b. Which habit that Bill_i hated did he_i say that Mary finally kicked?

CTM accounts can rely on the late adjunction of the relative clause to capture these anti-reconstruction effects (as in Stepanov 2001 and others). After movement of the wh-phrase to it non-base position, an adjunct can Merge with it and not the copy left in situ. The higher copy can be manipulated independently of the lower one

and if the relative clause is appended to the higher copy only, no Principle C effects are predicted. MD accounts can surely also resort to late adjunction, but this is going to have an unwanted effect: the adjunct is late-adjoining low as well as high and the lack of Principle C effects is not predicted.

Perhaps the injunction against scattered interpretations in (30) could be altered so as to make reference to segments and the non-standard assumption that the relative clause adjoins to the DP 'which picture' not the NP 'picture'. In this case, the moving DP will not *fully* dominate relative clause adjunct. But even then, we could further embed the adjunct such that it would be fully dominated by something:[11]

(32) Which picture of [the dog that Bill$_i$ hated] did he$_i$ say that Mary took?

What injunction could be made such that an MD account could handle these anti-reconstruction effects? Perhaps a 'whatever works' sort of approach that would absolve any Binding Principle violation if at least one occurrence obeyed them. This would falsely predict that non-adjunctions would also show anti-reconstruction effects:

(33) *Which picture of Bill$_i$ did he$_i$ say that Mary took?

One could claim that idiom-internal adjuncts are immune to binding theory strictures. Though this would also make the incorrect predictions:

[11] A reviewer notes that this test might not be entirely useful since Nunes 2001 argues that reconstruction-driven Principle C effects are not as strong when the R-expression is embedded more generally, even within nominal complements which should obligatorily reconstruct with their host. However, it is hard to construct the correct test with the picture idiom since the picture would need to be of a noun that takes a complement clause (like 'rumor' or 'claim') and such nouns are not generally things that can be photographed.

(34) *He$_i$ took some pictures that Bill$_i$ hated.

In short, the scattered interpretation facts can be cleanly captured under CTM but cannot be so captured by MD theories. CTM allows for the independent manipulation of terms in a way that cannot be said of MD. That is, late adjunction of adjectives allows for anti-reconstruction effects in certain instances but obligatory reconstruction for certain complements. The same cannot be said for MD. Again, the distinction comes down to dealing with terms or positions. In this case, we assume that adjuncts attach to terms and not to positions. If the adjuncts above could adjoin to positions to the exclusion of the term in them, then the problem for MD would not arise. However, this is not possible in a Merge-based system the deals in terms. This is not to say that MD accounts cannot in principle somehow capture this data, but much like Chomsky's argument from transformations, that approach would be less enlightening.

5. Conclusion

Merge as an explanation of displacement is an important conceptual step forward in the explanation of grammatical properties. The otherwise problematic fact of syntactic action at a distance is rendered much tamer by this re-interpretation of structure building. Coupled with other current guiding principles in syntax, it has lead to a question about the result of an object undergoing Merge more than once.

The two logically possible options, each reasonable in its own right, have been assumed, adopted, and argued for in the past, but here we have seen that there has been little clear basis for those moves. The clear differences between the copy theory and multidominance do not readily translate into strong arguments in

favor of one over another. Each approach asks something of the interface, but too little is known about the interface for that to be a deciding factor. Each approach carries with it its own expressive power, but much like other such instances in the past, we cannot use these differences to make the case for one over the other.

Instead I have shown that to adjudicate between the two requires discussion of interface-independent notions (such as what is targetable by Merge) or interface generalizations that enjoy some empirical support. When comparisons of that sort are made, it becomes possible to argue for one approach over another. In the instances relayed above, it was shown that a copy-theoretic conception of displacement is preferable.

6. References

Bachrach, Asaf, and Roni Katzir. 2009. Right node raising and delayed spellout. In Grohmann (ed.) *InterPhases: Phase-theoretic investigations of linguistic interfaces.* 283-316. Oxford University Press.

Berwick, Robert C., and Amy S. Weinberg. 1982. "Parsing efficiency, computational complexity, and the evaluation of grammatical theories." *Linguistic Inquiry* 165-191.

Blevins, James. 1990. Syntactic complexity: Evidence for discontinuity and multi-domination. Diss. University of Massachusetts.

Bresnan, Joan. 2001. Lexical Functional Syntax. Blackwell.

Cable, Seth. 2007. *The grammar of Q: Q-particles and the nature of Wh-fronting, as*

revealed by the Wh-questions of Tlingit. Diss. Massachusetts Institute of Technology.

Cable, Seth. 2010. *The grammar of Q: Q-particles, wh-movement, and pied-piping*. Oxford University Press.

Chen–Main, Joan. 2006. On the generation and linearization of multi-dominance structures. Diss. John Hopkins University.

Chomsky, Noam. 1955. Logical structure of linguistic theory (LSLT). Ms. Parts of 1956 revision published in New York: Plenum, 1975, and Chicago: University of Chicago Press, 1985.

Chomsky, Noam. 1957. *Syntactic structures*. Walter de Gruyter.

Chomsky, Noam. 1973. Conditions on transformations. In: Anderson, S.R., Kiparsky, P. (Eds.), A Festschrift for Morris Halle, Holt, Rinehart, and Winston, New York, pp. 232–286.

Chomsky, Noam. 1970. Remarks on nominalization. In: R. Jacobs and P. Rosenbaum (eds.) *Reading in English Transformational Grammar*, 184-221. Waltham: Ginn.

Chomsky, Noam. 1981. Lectures on Government and Binding, Foris, Dordrecht.

Chomsky, N. 1993 , A minimalist program for linguistic theory, in K. Hale & S. J. Keyser, eds, `The view from Building 20', MIT Press, Cambridge, MA, pp. 1-52.

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

Chomsky, Noam. 1997. New Horizons in the Study of Language.

Chomsky, Noam. 2001. "Derivation by phase. Ken Hale: A life in language, ed. by Michael Kenstowicz, 1-52.

Chomsky, Noam. 2008. On phases. in Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud, eds. Robert Freidin, Carlos Otero, and Maria Luisa Zubizaretta, 133–166. Cambridge, Mass.: MIT Press.

Chung, Dae-Ho. 2004. Multiple dominance analysis of right node raising constructions. *Language Research* 40:791–812.

Citko, Barbara. 2005. "On the nature of merge: External merge, internal merge, and parallel merge." *Linguistic Inquiry* 36.4: 475-496.

Citko, Barbara. *Symmetry in syntax: Merge, move and labels*. Vol. 129. Cambridge University Press, 2011.

Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima & Hisatsugu Kitahara. 1998. A Derivational Approach to Syntactic Relations. New York: Oxford University Press.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. LI 33:63-96.

Frampton, John. 2004. Copies, traces, occurrences, and all that: Evidence from Bulgarian multiple wh-phenomena. Ms., Northeastern University, Boston.

Gärtner, Hans-Martin. 1999. "Phrase linking meets minimalist syntax." *Proceedings of WCCFL*. Vol. 18.

Gärtner, Hans-Martin. 2002. Generalized Transformations and Beyond: Reflections on Minimalist Syntax. Berlin: Akademie Verlag.

Gazdar, Klein, Pullum and Sag. 1985." *Generalized phrase structure grammar*. Blackwell.

Goodall, Grant. 1987. *Parallel structures in syntax: Coordination, causatives, and restructuring*. Cambridge: Cambridge University Press.

Gracanin-Yuksek, Martina. 2007. About sharing. Diss. MIT.

Guimarães, Maximiliano. 2004. Derivation and representation of syntactic amalgams. Diss. University of Maryland.

Hagstrom, Paul. 1998. *Decomposing questions*. Diss. Massachusetts Institute of Technology, 1998.

Henry, Alison. 1995. Belfast English and Standard English: Dialect variation and parameter setting. New York: Oxford University Press.

Jackendoff, Ray. 1977. *X-bar-Syntax: A Study of Phrase Structure*. Linguistic Inquiry Monograph 2. Cambridge, MA: MIT Press.

Johnson, David E. and Paul Martin Postal. *Arc pair grammar*. Princeton: Princeton University Press, 1980.

Johnson, Kyle. 2012. "Towards deriving differences in how Wh Movement and QR are pronounced." *Lingua* 122.6: 529-553.

Kepser, Stephan. "On monadic second-order theories of multidominance structures." *The Mathematics of Language*. Springer Berlin Heidelberg, 2010. 117-128.

Kishimoto, Hideki. "Wh-in-situ and movement in Sinhala questions." *Natural Language & Linguistic Theory* 23.1 (2005): 1-51.

Kracht, Marcus. 2008. "On the logic of LGB type structures. Part I: Multidominance structures." *Logics for Linguistic Structures* 201: 105.

Kural, Murat. 2005. "Tree traversal and word order." *Linguistic Inquiry* 36.3: 367-387.

Lasnik, Howard, and Mamoru Saito. 1994. *Move alpha: Conditions on its application and output*. No. 22. MIT press.

Lebeaux, David. 1988. Language acquisition and the form of the grammar. Doctoral

dissertation, University of Massachusetts, Amherst.

Lebeaux, David. 1991. Relative Clauses, Licensing, and the Nature of the Derivation. In: Rothstein, S. D. (ed.) Perspectives on Phrase Structure: Heads and Licensing. San Diego: Academic Press.

McCawley, James. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13:91–106.

McCloskey, James. 1979. *Transformational syntax and model theoretic semantics: A case study in Modern Irish*. Dordrecht: D. Reidel.

McCloskey, James. 2001. "The morphosyntax of WH-extraction in Irish." *Journal of Linguistics*: 67-100.

Neeleman, Ad, and Hans Van De Koot. 2010. "A local encoding of syntactic dependencies and its consequences for the theory of movement." *Syntax* 13.4: 331-372.

Nunes, Jairo. 2001. "Sideward movement." *Linguistic inquiry* 32.2: 303-344.

Nunes, Jairo. 2004. Linearization of Chains and Sideward Movement. volume 43 of Linguistic Inquiry Monographs. MIT Press, Cambridge, Massachusetts.

Ojeda, Almerindo E. 1987. Discontinuity, Multidominance, and Unbounded Dependency in Generalized Phrase Structure Grammar: Some Preliminaries in Discontinuous Constituency. *Syntax and semantics* 20: 257-282.

Peters, Stanley, and Robert. Ritchie. 1973. "On the generative power of transformational grammars." *Information Sciences* 6: 49-83.

Pollard, Carl and Ivan A. Sag. 1994. Head-Driven Phrase Structure Grammar.

Chicago: University of Chicago Press.

van Riemsdijk, Henk. 2006. Grafts follow from merge. In Mara Frascarelli (ed.), *Phases of Interpretation* (Studies in Generative Grammar 91), 17–44. Berlin: Mouton de Gruyter.

Ross, John R. 1967. Constraints on variables in syntax." *doctoral dissertation, Massachusetts Institute of Technology*.

Sauerland, Uli. 2004. “The interpretation of traces”. Natural Language Semantics 12:63-127

Starke, Michal. 2001. Move dissolves into Merge. PhD dissertation, University of Geneva.

Stepanov, Arthur. 2001. Late adjunction and minimalist phrase structure. *Syntax* 4.2: 94-125.

Torrego, Esther. 1984. "On inversion in Spanish and some of its effects." *Linguistic Inquiry*: 103-129.

de Vries, Mark. 2005. Merge: Properties and boundary conditions. Linguistics in the Netherlands 22:219–230.

de Vries, Mark. 2009. "On multidominance and linearization." *Biolinguistics* 3.4: 344-403.

Wexler, Kenneth, and Peter Culicover. 1980. Formal principles of language acquisition.

Wilder, Chris. 1999. Right node raising and the LCA. In Bird, Carnie, Haugen, and Norquest (eds.) *Proceedings of WCCFL*. Vol. 18. Cascadilla Press,

Somerville MA.