11 Comments on the paper by Roberts[1]

Sabine Iatridou

Playing devil's advocate is not entirely an easy task. In addition, my own particular difficulty with this very interesting paper is that I do not share all of the background assumptions that Roberts makes. But rather than question them, I decided, for the sake of the argument, to accept *all* the assumptions that he makes. Furthermore, I shall not introduce any additional data. I will only deal with the internal structure of his proposals.

1 Introduction and general comments

Roberts' paper should be seen as within "a research program which aims to assimilate head movement as far as possible to XP-movement." Roberts notes that there are basically two differences between XP-movement and X-movement in ECP discussions like Rizzi (1990b): first, traces of XP-movement have to be both antecedent-governed and head-governed, while traces of X-movement only have to be antecedent-governed. Second, there are two types of XP-movement, based on a functional difference (A vs. A', or following Chomsky & Lasnik (1991) L-related versus non-L-related[2]) but no such split exists for X-movement, which has been treated as being of one type. He wants to eliminate these differences and support two claims:[3] that traces of X-movement should also be head-governed and that the notion of L-relatedness divides both XP-movement and X-movement into two types, so that all the boxes in the table in (1) are filled (where a head is L-related if it appears as a feature of a lexical head):

(1)

	L-related	~L-related
(a) XP	+	+
(b) X	+	+

There is a trivial sense in which (1) is true: of course there are some functional heads that appear as features on a lexical head (for example on the verb, like Tense and Agreement) and some that do not (like Negation and the Complementizer). But this is not what Roberts intends with (1). His goal is to show that the distinction in (1b) is operative in relativized minimality effects, the way (1a) is. This means that an α-L-related head can structurally intervene between two β-L-related heads[4] without causing problems of antecedent-government.[5]

In abstraction, such a situation would yield sentences which violate the Head Movement Constraint (HMC) yet are grammatical, as in (2), where Z moves over Y to X, leaving trace t_Z behind:

(2) X/Z .....Y.....t_Z...

Put somewhat differently, if ever a sentence appears to violate the HMC but is still grammatical, two conditions are satisfied:

(3) a. Z is able to skip over (i.e. ignore) Y
b. the trace *t* of Z satisfies the ECP

(3a) is Roberts' claim about there being two categories of heads, members of each of which can move over heads of the other variety. Moreover, he says that this division into two types is based on the same distinction that distinguishes maximal projections, namely L-related and non-L-related. So, (3a) is satisfied when both X and Z are L-related while Y is non-L-related, or when both X and Z are non-L-related while Y is L-related. (3b) addresses whether head-traces should be both antecedent-governed and head-governed. Roberts answers in the affirmative. This means that the derivation in (2) (grammatical sentences which violate the HMC) will be possible if—in addition to (3a)—Y is a head-governor .

For the most part, Roberts' paper discusses cases where (2) is apparently true and tries to show that (3a,b) hold. I think his discussion of (3b) is interesting and I will have nothing further to say on his position that head-government holds in X-movement. But the case for (3a) needs some refinement. In contrast, I will argue that Roberts' position that (1b) is operative in minimality effects does not seem superior even to alternatives that his own system provides. For example, recall that Roberts follows Roberts (1991) in accepting excorporation as a possible movement under certain conditions;[6] then an apparent violation of the HMC could have the derivation in which Z moves into and excorporates from Y, yielding the same linear string as (2) would:

(4) ...X/Z .... Y/t'_Z.....t_Z...

It could then be argued that instead of (3a,b), (4) should satisfy (5a,b):

(5) a. Z can excorporate from Y
b. the traces *t* and *t'* of Z should satisfy the ECP (which can again be brought down to satisfaction of both antecedent- and head-government).

The possibility of the (4)/(5) alternative also shows that (3a) and (3b) are not necessarily associated, given that one can find (3b) desirable without accepting (3a).

But before going over the specific cases that Roberts analyzes as (2)/(3), I would like to make a couple of general points. In Roberts' approach, there is an assumption (made explicitly in the quote at the beginning of this section) that we *want* (1) to hold, i.e. that we *want* X-movement to be as similar as possible to XP-movement. But why should such a state of affairs be a priori desirable? This is unclear to me. Such a much coveted simplicity or symmetry often (though not always) seems to be a by-product of the illusion that we can have intuitive access to how the physical world is put together; whereas all we have access to is which *theory* seems simpler or more symmetrical, hence we are often in pursuit of symmetrical representations like (1). The underlying belief is that Nature is a mathematician and does everything in a unifying way. But that is far from obvious and without denying that some such intuitions may yield positive results, it is entirely possible that the "truth," if in fact we are equipped to find what it is, is not going to conform to many of our aesthetic criteria of what makes a neat picture. Such a driving force is, of course, not necessarily particular to linguistics. Consider an example resulting from the similarity between Newton's law and Coulomb's law. According to Newton's law, the force F exerted between two bodies of mass m1 and m2 respectively, at a distance R is proportional to $(m1.m2)/ R^2$. According to Coulomb's law, the force exerted between two charges q1 and q2 at distance R, is proportional to $(q1.q2)/R^2$. But in the case of Coulomb's law there is a sign ('-' or '+') associated with the charges that determines whether the force is attractive or repellent. In Newton's law there is no such factor. One can imagine finding this difference between charges and mass a priori objectionable and arguing that since electrical charges come in two varieties, positive and negative, so too must mass and then try to look for a way to categorize the objects around us as having positive or negative mass. However, there

are enough differences between charges and mass so that one need not *automatically* consider this particular difference a weakness of the theory.[7] Similarly, there are enough differences between maximal projections and heads that it is entirely possible that they should differ with respect to being split along L-relatedness dimensions. It is easy to think of differences that one does not want to eliminate, even within the rest of Rizzi (1990b), which Roberts does not appear to want to change. Just to mention one, XPs can be referential (or D-linked); referential XPs can connect to their antecedents through binding chains by virtue of the fact that they have a referential index; non-referential chains, on the other hand, can only enter the more fragile government chains. Will we want to find referential (D-linked) and non-referential (non-D-linked) heads and binding versus government head-chains? Do we find it in principle objectionable that this distinction does not extend to heads? It seems to me that there are enough differences between heads and maximal projections, as well as between the chains that they form, so that one more difference (the lack of minimality effects in head-chains due to the L-relatedness dimension) should not be necessarily considered a weak point of the theory. In other words, I disagree with Roberts that (1) is an a priori desideratum but would be happy to accept (1), if it turns out that there are specific arguments for it.

So what would constitute support for (1)? It should crucially include arguments that there are constructions whose correct analysis is that of (2)/(3). This, in turn, means finding support for two positions: (a) that heads fall into two groups when judged on the basis of certain well-defined behavioral criteria and (b) that these two groups are best characterized as L-related and non-L-related. Although Roberts presents these two positions as one, it is important to keep them apart because both are necessary for (1) to be the final picture. It is entirely conceivable, for example, that heads do fall into two groups when it comes to certain aspects of their syntactic behavior *without* their distinction being along the (non-)L-relatedness line. In fact, in his paper, Roberts has another distinction that cuts heads into two categories, namely the (in)ability to head-govern.

To summarize so far, I take Roberts' goals to be the following: to find arguments for (1) in the analysis of certain constructions along the lines of (2) and (3). This includes showing that we should accept (1), not because we always wanted to and have finally found some phenomena that are consistent with it, but because we are forced to on the basis that both (a) and (b) above prove to be required additions to the theory.

2 Some specific cases

In this section I will focus on some of the specific points that arise in relation to the four constructions that Roberts analyzes as arguing for (1), (2) and (3). These constructions are Long Head Movement (LHM), Clitic Climbing, Long Movement of Infinitives, and AUX-to-COMP.

2.1 Long head movement

Long Head Movement is a construction in which a (non-finite or participial) verbal head moves to a sentence-initial position by skipping a (finite) verbal head, in apparent violation of the HMC:

(6) a. V[Prt]/[-fin]$_i$ ... Aux[+fin] ...t_i...

b. Darte he un exemplo (Old Spanish/ Roberts' (10b))
give-you (I) will an example
'I will give you an example'

Roberts argues that this movement of the untensed verb does not take place for morphological reasons and that within his system this means that the chain created would be non-L-related.[8] He now has to show that this chain can be broken by an intervening non-L-related head. He claims to find such evidence in the fact that negation blocks LHM. In other words, while the representation in (6) can characterize well-formed sentences, the one in (7) cannot:

(7) * V[-fin]$_i$ ...Neg... t_i...

The unacceptability of (7) versus the acceptability of (6a) is said to show the desired result: since LHM is a non-related chain, negation (a non-L-related head) causes a relativized minimality effect in the non-L-related chain (by preventing the verb from antecedent-governing its trace), while the finite verb (which is L-related) does not. But does this argument hold? In order to be convincing, Roberts has to show that negation does not block L-related head movement in the relevant languages. He has no such cases; in fact, this part of the argument is not addressed. If it is not shown that negation causes problems *only* for non-L-related chains, it is natural to wonder whether negation causes the ungrammaticality of LHM for a reason entirely different from the one given. It is conceivable, for example, that in these languages negation blocks head movement over it not because it is non-L-related but because it is not a head-governor. As a result, the trace left below nega-

tion would violate the ECP. It seems to me, then, that before we can accept Roberts' evaluation of LHM with respect to his proposal, he has to show that negation blocks only *some* head-chains.

Even if we assume that the ungrammaticality of (7) is due to negation being a non-L-related head, the question arises as to why the moving head should want to skip negation in the first place. Within Roberts' own proposal, the fact that negation is non-L-related means that it cannot be ignored in the formation of a non-L-related chain. But why can it not form part of the chain? In other words, what would prevent the verb from moving on to negation and then excorporating from it? This possibility is not excluded—or even mentioned—with respect to LHM, but in the section on clitic-climbing a related derivation is ruled out, effectively by fiat. I will return to this point in the relevant section.

Roberts claims that LHM shows that (3b) is also satisfied and that, in fact, LHM is possible only if AGR is a head-governor. I had some difficulty with this argument due to a combination of my not understanding some of Roberts' underlying assumptions and his not being terribly explicit about his phrase structure. I would like to make some points explicit, in case a reader has a similar difficulty. Roberts says that LHM is possible only if AGR is a proper head-governor. In a representation like (8) (his (19)):

(8) $[_C$ V[-fin]$_i$ $]$ $[_{AGR}$ (cl-) Aux[+fin] $]$ ...t_i...

The [-fin] verb can move only if its trace(s) are head-governed. This might very well be so but it is unclear to me how he derives from this that AGR is the actual head-governor of one of them. Roberts follows Belletti (this volume) in having AGR above T(ense) and also assumes that the infinitival verb first moves through T to pick up infinitival morphology. This seems to imply that there are two T nodes: one with the finite morphology that ends up on the auxiliary, and one with the infinitival morphology that ends up on the infinitival verb. He then also has two AGR nodes, so that the presupposed (although not quite explicated) phrase structure appears to be as in (9):

(9) C AGR_1 T_1 AUX AGR_2 T_2 V

In other words, in LHM, AUX is supposed to move from its position to T_1 and then to AGR_1, while the verb moves to the head of T_2 and from there, skipping AGR_2 due to its L-relatedness, directly to C. It is only within these assumptions that one can accept the conclusion that LHM

shows that AGR is a head-governor. If one does not accept both (9) *and* that LHM is *specifically* T_2-to-C-movement, the same conclusion cannot be drawn.

Finally, Roberts does not attempt to exclude a derivation in which there is L-related movement of the verb to AGR_2, before the "long" part of LHM starts, the "long" part being the non-L-related part. After all, the movement of V-to-T_2 forms an L-related chain, it could continue in L-related movement to AGR_2, and the non-L-related part of LHM would start in the last part of the derivation, when the target is C. But also he does not address the possibility of the verb adjoining to and excorporating from AGR_2. Such an adjunction to an L-related head is permitted in his proposal and treated as non-L-related movement, so that under this scenario, the non-L-related part of LHM would still start when the verb leaves T_2. In fact, he makes use of such movement in his treatment of clitic-climbing.[9] Since either of these alternatives is in principle permitted by his own account, it is hard to simply accept the claim that AGR_2 is skipped over. But managing to show that AGR_2 *is* skipped over in non-L-related chains constitutes his goal (3a), a very crucial part of his proposal. Therefore it seems that this part of his discussion of LHM leaves some open questions.

2.2 Clitic-climbing

Another construction that Roberts analyzes as arguing for his position is clitic-climbing. His argument goes as follows: clitic-climbing is a case of non-L-related movement. Therefore it can be blocked by an intervening non-L-related head. Specifically, the intervening head he considers is negation and, as is well known, negation blocks clitic-climbing. This he then takes as proving his case.

Let us consider each step at a time. First he needs to show that clitic-climbing is a case of non-L-related movement. He basically does this by proposing that clitics adjoin to AGR and that all adjoined positions are non-L-related positions, as mentioned in the previous section. But, according to Roberts, although clitics have to adjoin to AGR, nothing forces them to adjoin to the *local* AGR. When they do not, in fact, clitic-climbing results:

(10) $[_{AGR1}$ cl_i $[_{AGR1}]]$ ... T_1 ... V $[_{CP}$ $[_C$ $t_i']$ ... AGR_2 $[_{TP2}$ t_i $[_{TP2}]]$ VP'

(11) lo voglio fare
it want do

'I want to do it'

In (10) the clitic starts out as a maximal projection (DP) (from where is not specified), adjoins to TP and then starts moving as a (D) head. It skips the local AGR_2, moves into and excorporates from C,[10] and then moves to the higher AGR_1. The clitic can go from C to AGR_1 by skipping T_1 because T_1 is an L-related head and the clitic is forming a non-L-related chain. The movement in (10), Roberts says, is possible partly due to the ability of AGR_2 to head-govern the trace t_i in TP_1 and partly to the ability of V to head-govern the trace t_i' in C. The former is a parameter setting of AGR in Italian, and the latter is a lexical property of restructuring verbs. Again, accepting many details of this scenario is crucial to accepting Roberts' more general conclusion. For example, the lower AGR_2 must be skipped and not moved through. If it were moved through, there would be a trace left there that should be head-governed, and that would force Roberts to say that C is a head-governor and head-governs the trace in AGR_2, something that he clearly cannot have (cf. also the discussion in his section 6 where he explicitly rules out Italian C as head-governor).

But skipping AGR is not entirely a simple matter since it goes against the general spirit of minimality that a moving element must move through all potential landing sites (Roberts accepts this, as in a later section where he says that "...movement never goes further than the nearest appropriate landing site — essentially the idea behind deriving Relativized Minimality from economy"). And since the local AGR_2 is a potential landing-site, one would expect that it must function as a move-through site as well. Roberts does not address this point from the general (aforementioned) angle. He claims to be able to exclude movement through AGR_2 by appealing to (12) (his 23):

(12) Clitics adjoin to AGR, possibly via successive-cyclic movement

It is unclear to me how exactly (12) will derive the desired result. Roberts takes successive cyclic movement to be excorporation. Excorporation is possible only from a non-m-selected position, i.e. only if movement into that position did NOT take place for reasons of morphology. But movement of the clitic to AGR is, following his (12/23), exactly adjunction/non-L-related movement, NOT m-selection. Therefore one would expect excorporation to be possible, yet he doesn't permit it. He says in fact, that "what [12] states is that once adjoined to AGR, clitics do not further excorporate (although they may move with AGR)." But it is entirely unclear to me how (12), in combination with his theory of excorporation, yields the content of the aforementioned quote.

So, to summarize so far, it is crucial for Roberts that the clitic NOT move through the lower AGR on its way to the higher AGR. But the way he prevents this movement seems unclear. Notice that he does not say that AGR_2 is skipped because it is an L-related-head and clitic-movement forms a non-L-related chain. And it makes sense that he does not use this argument because, since the clitic is aiming towards an AGR/L-related head, he cannot have it at the same time skip other AGR-heads on the basis of their L-relatedness. It wouldn't make sense to appeal to the L-relatedness of AGR since there is a way that an L-related head can form part of a non-L-related chain, namely by having the moving element adjoin to it, which would create a non-L-related chain segment.[11]

But what other way is there to exclude movement of the clitic through the lower AGR? Maybe on the basis of the following point, which was made above: if the clitic moved through AGR_2, the trace left there would have to be head-governed. The only element that is in a position to do that is C. But C is not a proper head-governor. Therefore the trace in the lower AGR would not be head-governed and so that derivation would be out. But although this scenario blocks movement into and out of the lower AGR_2, it does not automatically permit skipping over it. Doing that would obviously amount to not creating a trace because that trace would violate the ECP and thereby permit, for example, movement out of islands, clearly a highly undesirable result. So, I don't see how Roberts can have the clitic skip over AGR_2, but I do see that he definitely needs it to do so.[12]

Let us grant the last point and continue on the assumption that there is some factor, as yet undiscovered, that would permit the clitic to skip the local AGR_2. Now the point is to prove that negation blocks this movement. Clearly it does:

(13) a. *Lo voglio non fare
b. Voglio non farlo

Roberts takes this to show that negation is a non-L-related head. But here the same point can be made as in LHM: negation blocks clitic-climbing, but does it do so because it is a non-L-related head? To show this, Roberts would have to show that negation does NOT block L-related movement. He does not attempt to do this. (In Italian, as far as I know, the verb always follows *non*. This will make it very hard to

show that negation does not block L-related movement as well. Indeed, this fact could very well be taken to show that negation blocks also L-related movement, precluding the desired theoretical result.) On the other hand, it is entirely possible that, for example, negation blocks movement over it because the trace left just below it (in Roberts' tree this would be the trace in TP) is not properly head-governed, negation not being a head-governor.

But let us grant Roberts one further step and assume for the time being that his treatment of clitic-climbing does show that negation is a non-L-related head because it blocks clitic-climbing by blocking antecedent-government of the trace in T by the trace in C. Another obvious question arises at this point, which also came up in the discussion of LHM in section 2.1 above: if Neg is a non-L-related head, why doesn't it form part of the non-L-related chain, rather than block it? In other words, why couldn't the clitic adjoin to and excorporate from the Neg-head? Unlike with LHM, here Roberts foresees this question and quickly preempts it by postulating (14) (his 28):

(14) Excorporation from non-L-related heads is impossible.

But (14) is custom-made for this case; as far as I can see, there is no motivation given for it other than to make possible Roberts' position that negation is skipped over. In fact, now the explanation of the data does not rely any more on the correctness of the position that there are (non-) L-related heads, but on the correctness of (14), for which there is no independent evidence given. Imagine, for example, that (14) did not hold but the rest of the story is as Roberts wants it. In other words, imagine that negation is a non-L-related head and the clitic can move through it (i.e. adjoin and excorporate). All such a derivation does is provide one more head the clitic has to stop at on its way up. But this would mean that, contrary to fact, negation would not block clitic-climbing and that, nevertheless, this derivation would be fully compatible with the desired result that there are (non)-L-related heads.[13] It seems then, that Roberts' basic proposal that there are two varieties of heads could be right even if negation did not block clitic-climbing. In other words, the behavior of negation with respect to clitic-climbing cannot be taken as an argument *for* a certain theoretical claim, since the exact opposite behavior of negation is still compatible with that very same theoretical claim.[14] Unfortunately, Roberts does not provide evidence for (14) other than somewhat turning the argument on its head by *presupposing* that

heads come in two varieties and that Neg is a non-L-related head, and arguing that since negation blocks clitic-climbing, it *must be* that (14) holds. But deriving the necessity for (14) *from* the existence of (non-)L-related heads is of course not the same as arguing *for* the existence of such heads.[15]

But let us again take the account as far as it goes by accepting what Roberts would want us to. Then at least two questions come to mind, one more general than the other. The more general one is the following: earlier in the paper Roberts adopts his (1991) proposal according to which "excorporation is possible from an adjoined position or from a position into which a head has freely substituted, but not from an m-selected position into which a head has been substituted to satisfy m-selection." Movement to satisfy m-selection is a subset of L-related movement. What are the cases of permissible excorporation we are left with upon the introduction of (14)? Excorporation from non-L-related heads is out. Excorporation from L-related heads when movement happened for reasons of m-selection is out. What about excorporation from an L-related head to which movement happened for a reason other than m-selection? One such case is movement of a clitic to AGR. He argued before that "no adjoined position is an L-related position," which means that (14) is applicable in the case of clitic climbing,[16] and indeed he assumes that the clitic cannot leave AGR once it is adjoined to it. So, if I understand correctly, excorporation out of a position adjoined to an L-related head also appears to be out. This leaves only one case: excorporation from an L-related head into which movement was of the "freely substituting" variety (which according to Roberts (1991) is possible only when the host head is "radically empty"). It would be interesting to see what empirical cases fall in this category.[17]

The second point, of a more technical nature, revolves around the actual chain created in clitic-climbing, namely our (10). Recall that the clitic is said to move up by successive cyclic movement, which Roberts construes as 'excorporation'. The first (head-)trace that the clitic leaves is in D after the DP/clitic has moved to TP. Since this is the original trace of the clitic/head, I assume the notion of excorporation is irrelevant. Then the clitic moves into and excorporates from C. But C is supposed to be a non-L-related head and by (14) the clitic is not supposed to be able to excorporate from it. How this can be resolved within the assumptions of the paper (this apparent contradiction is not addressed) is unclear to me.

2.3 "Long" movement of infinitives

According to Roberts, long movement of an infinitive is a construction in which an infinitival verb moves to T, skips over AGR and moves into a higher position, as in (15a) (his (34)):

(15) a. $Inf_i\ [_{AGR}\ cl_j\ AGR\ ]\ ...\ [_{TP}\ t_j\ [_{TP}\ [T\ t_i\ ]\ ...$

b. Gianni ha deciso di farlo / *lo fare
G. has decided to do it / *it do

Roberts also says that it is unknown what this higher position is, as well as what triggers this movement, but since it does not take place for m-selection, it is non-L-related movement.[18] The main motivation of his discussing this construction is to show that AGR is a head-governor in Italian: the traces in TP and T are licit, which means that they are head-governed (and, of course, antecedent-governed). This may be a correct point to make about (15). But the side-comments of this section (the ones about (non-)L-relatedness) are again up for discussion. He assumes that AGR is skipped because it is an L-related head and movement of the infinitive is non-L-related. But how do we know that the infinitive does not move to AGR and excorporate from it?[19] This is permitted since movement to it did not take place for reasons of m-selection and since AGR is an L-related head and excorporation from it is permitted (see the discussion (14) in the previous section and fn. 15). So I don't think this section provides evidence for (3a), that is, for the (non)-L-related distinction, but just for condition (3b)(head-government).[20]

2.4 AUX-to-COMP

AUX-to-COMP is a construction in which a participial (or infinitival or marginally subjunctive) verb moves to C, giving the appearance of subject-AUX-inversion. When there is a clitic, it is carried along by the verb (Roberts' 36 and 44):

(16) a. $[_C\ V+T+AGR+Cl]\ [_{AGR}\ t'']\ [_T\ t']\ [_{VP}\ ...\ t...\ ]$

b. Avendo Gianni fatto questo, ...
having John done this
c. *Avendo Gianni lo fatto

having John it done
d. Avendolo Gianni fatto
having it John done

According to Roberts, the derivation of a sentence like (16d) is as (16a): the verb leaves its VP, moves to T, moves to AGR and picks up the clitic, and finally moves to C.[21] This is possible partly because Italian infinitival AGR is a head-governor and head-governs the trace below it, namely the trace in T. This proves the point of this section, but notice that it doesn't show anything about the (non-)L-related distinction. Roberts recognizes that even though this is the formation of a non-L-related chain, it cannot be maintained that the L-related AGR was skipped because the auxiliary verb must have been there to pick up the clitic. But since AGR is adjoined to, it forms a good non-L-related link in a non-L-related chain. This again raises the issue of excluding the possibility that AGR was, in fact, not skipped, but adjoined to and excorporated from.[22]

2.5 Overview

One of the main points of the four constructions discussed in this section is that they are supposed to show the correctness of (2) and (3a) by showing that the HMC is violated because α-L-related movement can skip over β-L-related heads. Specifically, they are all cases of non-L-related heads moving over L-related heads. The reverse (L-related heads skipping over non-L-related heads) is not discussed because there are supposed to exist additional constraints on the locality of L-related chains. I will return to this briefly in the final section. I argued, however, that (3a) was not convincingly argued for, partly because the possibility of movement into and excorporation from the allegedly skipped over head (i.e. (4)/(5a)) was not excluded.

3 Final comments

As argued in section 2, Roberts does not quite prove that apparent violations of the HMC have the derivation in (2) (repeated below as (17)) rather than the one in (4) (repeated as (18)):

(17) $...X/Z\ .....Y.....t_Z...$

(18) $...X/Z\ ....\ Y/t'_Z.....t_Z...$

I already described (18) as a possibility of adjunction of the verb (Z) to and excorporation from AGR (Y) when head-movement of the verb to X is non-L-related. The movement to Y would also constitute non-L-related movement, given that Roberts assumes adjunction to a head to always create a non-L-related chain. However, there is another scenario that may yield (18), namely one in which movement of Z to Y is L-related, while further movement of Z from Y to X is non-L-related. In case Z is also L-related, this would not automatically yield major problems, but imagine the case in which Z is a non-L-related head. Then the chain formed by movement from Z to Y to X would constitute a case of "improper" movement (non-L-related to L-related to non-L-related). With maximal projections, such ("improper") movement is excluded on the basis of Binding Condition C effects: the deepest trace (a variable left in the original position of Z) would be bound (by the trace in Y) inside the domain of its operator (the element in the non-L-related position X). But it is questionable whether these same considerations make much sense in the world of heads. The notion of non-L-relatedness just means that the head in question is not a feature on a lexical head. From there to assume that a non-L-related head has operator-like properties, with a variable and a range etc., is a far cry. And unless this can be shown, it is unclear whether improper head movement (i.e. an L/non-L chain) should be excluded on the same principles as the related XP case. And maybe it might not have to be excluded at all. This would mean that there is one more alternative to Roberts' (2)/(17). Interestingly, this alternative is still fully compatible with the position that there are (non-)L-related heads. But rather than take this to mean that non-L-related movement skips L-related heads, as Roberts does, we can take it to mean that unlike XP-movement, non-L-related heads can move through L-related positions without problem. The considerations that rule out such movement for maximal projections are due to independent properties of maximal projections and are not applicable to head-movement.

If (4)/(18) rather than (2)/(17) is the correct representation, it would mean that there are no actual violations of the HMC, at least not in the constructions Roberts discusses. And this could either be taken to be just a matter of fact of the particular constructions, or to be derived from more general Economy considerations of shortest move, provided that the traces left are head-governed and antecedent-governed, as Roberts himself would want anyway.[23]

On the other hand, with respect to L-related movement, at least the

cases where it is motivated by morphological reasons, Roberts argues that considerations of m-selection make it never violate the HMC. But why should movement for morphological reasons automatically be local? Of course, by calling it "selection" one builds in the locality requirement. But if movement takes place merely to put features that need to be picked up on a head, why should it be the closest head that picks them up? In a way, one can see that the locality requirement should not be taken too literally in that AGR is said to take as complement not the element on which it has to appear (the verb) but T. Besides, by pointing out in his footnote 9 that Neg does not block movement of the verb from T to AGR, Roberts himself recognizes that movement for reasons of morphology does not always and necessarily result in strict locality.

Finally, with respect to the sections on the history of French and cross-Romance generalizations, I found Roberts' discussion very interesting and I think he makes a good case for considering four phenomena which are at first blush dissociated (clitic-climbing, long infinitive movement, AUX-to-COMP and null subjects) as surface realizations of one particular property. He argues that this property is the ability of AGR to head-govern. But even if he is wrong in this identification, Roberts has certainly set the stage in such a way that an investigation of any one of these phenomena will have to include its connections to the other four. This type of unification always constitutes a significant advance.

Notes

1. I would like to thank Michael Hegarty for helpful discussions.
2. In Chomsky & Lasnik (1991), a position is defined as being L-related when it is the specifier or complement of a feature of a lexical head L.
3. Roberts gives four reasons to reformulate the conditions on X-movement. The two in addition to the ones mentioned in the main text are (a) the locality on X-movement follows in many cases from the reason that triggered the movement (morphology), so for those cases we do not need to have in addition something like the Head Movement Constraint (I will return to this point in more detail later); (b) it is easy to find in the literature cases of XP-movement where either the antecedent-government requirement or the head-government requirement is violated. The same does not hold for X-movement. This difference boils down to the fact that most accounts have not taken a head-government requirement to be operative for X-movement. Clearly, if Roberts proves

successful in imposing the head-government requirement for X-movement, he has to try to find selective instances of X-movement which violate only the head-government requirement. Since this second point reduces to the question of head-government, there are in effect two differences between X- and XP-movement that he wants to eliminate.

4. Where α and β are opposite values of (non-)L-relatedness.
5. This would roughly yield (i) and (ii) for heads, duplicating the equivalent definitions of Rizzi (1990b) for maximal projections and thereby making these definitions insensitive to the head versus maximal projection nature of the elements of the chain:
   (i) A head X antecedent-governs Y only if there is no head Z such that
   - Z c-commands Y and does not c-command X
   - Z is a typical potential antecedent-governor for Y.
   (ii) Z is a potential antecedent-governor for Y if both X and Z are α-L-related heads.
6. In Roberts (1991: p. 211), excorporation is defined as "successive cyclic head-to-head movement where one head simply passes through another, first incorporating and then moving on."
7. In a parallel case, imagine that you are doing physics and discover that 29 of your 30 particles have an anti-particle. There is nothing obviously different about the 30th one; so you'll continue searching until you either find that anti-particle or a reason that would differentiate your 30th particle. Intuitively, one could say that such a position is justified. But imagine you have 2 particles and only one has an anti-particle. Would you be equally motivated to impose symmetry there? These examples are all from physics, but if language is like other biological systems, then the point can be made more strongly, since biological systems are notoriously "messy."
8. In principle, his proposal does not dictate that every case of head-movement that does not take place for morphological reasons should be non-L-related (see his discussion around his (20) and (21)). In general comments he explicitly states that though movement for morphological reasons is L-related, there can be other cases of L-related chains as well. But when he discusses the specifics of LHM (and the other constructions) he deduces the non-L-relatedness of the chain from the argument that movement did not happen because of morphological considerations. In other words, he effectively (but not in principle) identifies L-relatedness with m(orphological)-selection and non-L-relatedness with the remaining cases of movement. If this is correct, one wonders why we would need the notion of (non-)L-relatedness at all, given that (non-)m-selection (which he has anyway) will yield the required distinctions. Overall, Roberts' paper does not discuss what L-relatedness does over and above m-selection.
9. Recall that following Rizzi & Roberts (1989), Roberts distinguishes between substitution and adjunction for head-movement.
10. According to Roberts, this movement into C creates an economy of derivation issue, as follows. Since this derivation involves the additional trace in C, it will be longer than just movement of the clitic to the local AGR2. And since the shorter movement is not preferred, i.e. since the longer movement is possible, it must be that it is obligatory, and therefore clitic-climbing will never be

optional. But it seems to me that somewhere in this line of reasoning there must be a gap, since clitic-climbing in Italian is optional, as he himself shows some time later in the text, without addressing, as far as I can tell, the contradiction to the aforementioned conclusion.

(i) Lo voglio fare
(ii) Voglio farlo

But it is unclear whether such an economy of derivation issue really exists here. There is a tension between the Economy requirements of shortest move and fewest steps; Chomsky (1992) attempts to resolve the tension between these contrary requirements in his formulation of "form chain" as a unitary operation.

11. In addition, according to Roberts' discussion just above his (43), infinitival AGR does not m-select T. And since he also believes infinitival AGR can be skipped, it leaves open the possibility that even substitution into infinitival AGR (i.e. not just adjunction to it) may constitute non-L-related movement. I will not pursue this further.
12. In all this discussion, AGR2 is the AGR of the lower clause. If phrase structure is indeed as in (9) of section 2.1, then structure (10) will be more complicated. This will also raise the question of what happens with the AGR of the main verb (i.e. AGR2 of (9)) when the clitic appears before the auxiliary verb as in a simple sentence like (i):
    (i) Je l'ai mangé
    I it have eaten
    'I have eaten it'
    Is the AGR of *manger* skipped over by the clitic on its way to the AGR *avoir*? For this to be possible, AGR should be a head governor, an option not taken in Modern French according to Roberts. Is the AGR moved into and excorporated from? But then this should also be considered a possibility for the Italian cases.
13. Notice that in this imaginable scenario it would be harder to show why negation did not create a problem for clitic-climbing: it could be because Neg was an L-related head and therefore skipped over or it could be because negation is a non-L-related head and therefore forcibly an intermediate landing site for the clitic.
14. This is just an instance of the more general dialectical point that if both a position A and its opposite ~A are compatible with a position B, A cannot be used to argue for B.
15. Obviously one cannot avoid this whole problem by not accepting the relevance of (or even existence of) excorporation in these cases, because it is a crucial component of Roberts' entire system.
16. Unless he wants to make a distinction between non-L-related *heads* and non-L-related *head-positions*, which he does not do explicitly, but which may be desired.
17. I looked at Roberts (1991) for some candidates. The only cases of excorporation discussed there are clitic-climbing and verb second in combination with verb raising. The first one is the actual case under discussion here and the second one is not of much help within the present context because it is unclear how some of the assumptions of the present paper would be extended to the con-

struction discussed there.

18. Notice that he is again not considering non-m-selected L-related movement, an option that he has left open in principle.
19. Excorporation of the infinitive from AGR would leave the clitic behind (i.e. it would cause enclisis), as desired.
20. There is no discussion of the interaction of this non-L-related chain of long movement of the infinitive with non-L-related heads like negation, i.e., it is left open whether/how negation is skipped over or moved to and excorporated from.
21. It seems to me that an additional step must be involved, because otherwise it is unclear why the clitic-auxiliary order of simple sentences is not involved, yielding the unacceptable (i):
    (i) * lo avendo Gianni fatto...
22. As in the section on long movement of infinitives, a more comprehensive treatment would include the interaction of (non-L-related) AUX-to-COMP with (non-L-related heads like) negation, but Roberts has no such discussion.
23. This alternative might leave unsatisfied Roberts' desire to be able to point to selective violations of either the head-government or antecedent-government requirement. Take as an extreme example the case Roberts gives as satisfying the head-government but not antecedent-government (his (52)):
    (i) *Lo voglio che faccia
        it I-want that I-do (subjunctive)
    Here *voglio* does not exercise its property of restructuring which Roberts himself has identified with the ability to head-govern a trace in C. Or maybe it's just that the clitic cannot excorporate from overt C the way it is supposed to be unable to from Neg. In fact, it might be interesting to consider a more general alternative analysis in which there is no requirement for antecedent-government, but only one for head-government (as in Rizzi (1990b) for XPs) in combination with the constraint on excorporation from non-L-related heads (as in (14)). And in case the latter turns out to be reducible to head-government, maybe head-government by itself will prove sufficient.