On Wh-head-movement*

Caterina Donati

Università di Urbino
c.donati@uniurb.it


*Acknowledgments: Different versions of this paper have been presented at Going Romance 2002, the Workshop on “On WH Movement” and GLOW 2003. Many thanks to the audiences for important comments, questions and suggestions. The paper has also benefited from the observations of an anonymous reviewer, whom I thank.
ON WH-HEAD-MOVEMENT

0. Introduction

The merit of “On WH movement” was that of explaining in a principled way and of unifying under a theory – movement – a number of properties previously unobserved, or unrelated, or unaccounted for. I shall start my paper by considering with a critical and “minimalist” eye another property of wh-movement that was simply assumed as a primitive in the theory, and left unexplained, namely the fact that wh-movement operates on a phrase. More generally, the theory of movement as it developed in the seventies and eighties all took for granted the existence of two configurationally different movements: head movement and phrase movement, which would never intersect and never compete in their domain of application, in what can really be seen as a remnant of a “construction driven” approach to syntax.

This approach can no longer be maintained if the minimalist attitude is to be taken seriously. More precisely, an approach assuming as a primitive the existence of two kinds of movement is incompatible with a a feature approach to movement. As is well known, in Chomsky (1995, 2000, 2001), movement is reanalyzed as a complex operation crucially triggered and governed by features: an uninterpretable feature $\alpha$, the probe, searches locally for a matching feature $\beta$, the goal (agree), a copy of which is merged in a local configuration for checking (merge copy). What is really seen by the computational procedure is a feature, and the phrasal properties of wh-movement need to be derived.
The paper is organized as follows: after a preliminary section (1) devoted to exclude simple features-movement as a syntactic option, the issue of the status of head-movement is addressed (section 2), reviewing with a critical eye Chomsky’s recent tentative to exclude it from narrow syntax. The conclusion is that head movement does exist as a syntactic option and that it is indeed the unmarked pied-piping operation, being minimal. The following section (section 3) reviews systematically the conditions on movement spelled out in the theory starting from “On WH-movement”, searching for a property of head movement that might make it unavailable in standard wh-movement contexts, hence triggering the less minimal, more costly operation phrasal movement. Finding no such property, I propose in section 4 that the choice between phrase and head as the target of movement operation is simply due to the phrase structure status of heads and phrases, together with the Condition on uniformity of chains (Chomsky 1995). Two case studies are provided in the following section 5, i.e. free relatives and comparatives, showing that wh-movement does not necessarily involve phrases and that the fact it does in classical case is just a consequence of the definition of what a head/phrase is together with the interpretation of these constructions.

Finally, section 6 examines some cases which look like residuals of specific constraints on head movement, showing that they are effects of a phonological requirement which can interfere with any movement operation. Section 7 provides some conclusions.

1. No features movement

A preliminary question, explicitly addressed in Chomsky 1995, is why the complex operation described above as movement (agree+merge copy) cannot simply occur at the level of features, and some kind of (generalized) pied-piping is necessarily involved. In
other words, why features cannot move. The first tentative answer given by Chomsky was that features cannot move for a phonological constraint:

For the most part – perhaps completely – it is properties of the phonological component that require pied-piping. Isolated features and other scattered parts of words may not be subject to its rules. (Chomsky 1995:262-3)

The problem with this proposal is that it carries the implicit implication that post spell-out movement, being free from any phonological constraint, should display no pied-piping. There is good evidence, however, of the existence of pied-piping, i.e. of phrasal movement, at LF.

A simple case providing evidence for covert phrasal movement is that of Antecedent Contained Deletion: in (1), where the deleted VP is contained within its antecedent, the only way to recover ellipsis without bumping into infinite regress is to allow QR of the phrase “every city Bill does”.

(1) John will visit [every city Bill does]
    [every city Bill does] John will visit [t]

A more convincing alternative is to identify the constraint banning features movement with a property of Merge. When it applies to the numeration, merge indeed is restricted to operate on lexical items, on bundle of features. It is thus natural and minimal to assume that this restriction holds as well when merge is part of a movement operation. Let us then assume that there is a single operation merge in the syntax, which applies freely and is constrained to operate on words (Matushansky 2002).
Taken for granted that features cannot move (or better, be merged), some sort of pied-piping is required. The minimal amount of pied-piping a feature should take along to form a mergeable object is a word, i.e. a head. This derives head movement. What about XP-movement? Being anti-minimal, why should it exist at all?

2. The status of head-movement

The easiest way to solve this puzzle is to proceed as we did for features-movement, banning it from syntax. This is what does Chomsky (2001): he assumes that head movement is not a syntactic operation but part of the “phonological branch of the computation”. If this is true, then the minimal amount of material a feature can pied-pipe becomes a phrase, and the phrasal properties of (standard) wh-movement happen to be derived for free.

This move however is untenable for several reasons: it holds on wrong assumptions, both theoretical and empirical; it is incompatible with other aspects of Chomsky’s theory (e.g. phases); it brings undesirable consequences.

First, this proposal holds on a wrong theoretical assumption, namely that head movement violates crucial syntactic requirements. And in particular, on the assumption that head movement does not extend the target (hence being countercyclical). This is true for the standard head-adjunction configuration given in (2).

(2)
The traditional reason for this configuration is related to the standard cases of head movement usually considered in the literature: in V-to-T movement, or in V-to-C movement, the two heads conflate and behave as a single constituent, hence the assumption that they form a sort of a “derived lexical item” represented in the head adjunction configuration above.

But this is not the only configuration head movement can in principle produce. Suppose we have a head X endowed with an uninterpretable feature α which needs to be deleted. Nothing prevents in principle to Merge a copy of a head Y endowed with the matching feature β to the root of the structure, as in (3).

\[
\begin{array}{c}
Y^\circ \\
\quad \text{XP} \\
X^\circ \quad \text{.....} \\
\quad \text{YP} \\
\end{array}
\]

The configuration in (3) is a structure obtained by merging a new item to the root of the tree, hence complying with the extension condition. Given (3), the head conflation effect correlated in many cases with this configuration can be the result of an independent process, perhaps phonological, call it affixation, which has nothing to do in principle with head movement: affixation is something that can happen to two adjacent heads independently from how and why they ended up being adjacent.

Separating head movement from affixation is rather a good idea at the light of a number of empirical facts: first, there are cases of affixation which do not correlate with head movement, but with phrasal movement (see below); second, there are cases, as we
shall see in Section 5, where head movement does not yield any effect of this kind; third, there are cases, to which we shall return at the end of the paper (Section 6), where affixation does not involve moved heads, but merged ones.

To illustrate the first kind of cases, consider (4) as a possible derivation for a clitic construction in Italian.

(4) \[ \text{TP} \]

\[ \text{DP} \rightarrow \text{TP} \]

\[ \text{D} \rightarrow \text{T} \rightarrow \text{VP} \]

\[ \text{lo voglio} \]

\[ \rightarrow \text{DP} \rightarrow \text{D} \]

Clitic pronouns in Romance have always been a problem given standard assumptions on phrase and head movement: clitics are clear maximal projections in the position where they are generated, but they end up being affixed to an inflectional head. This tension is at the root of standard and influential analyses such as Kayne 1989’s, where the clitic starts up as a phrase but moves and adjoins as a head, changing its phrase structure status in the course of the derivation (violating a condition like that of the uniformity of chains, Chomsky 1995). The derivation in (4) is by far more minimal: here the phrase structure status of the clitic never changes during the derivation, and the clitic moves as a phrase. The D to T incorporation is an independent process akin to phonology.

Concluding so far, nothing in head movement prevents it from being cyclic, and assuming the existence of a cyclic head movement appears to be empirically convenient.
But banning head movement from syntax also runs in troubles with one of its implications which appears to be too strong: being phonological, head movement should never feed LF, and hence never have any interpretive effect. This may well be true if we consider movement of predicative heads, which need to reconstruct for interpretability (Heim & Kratzer 1998). But even in these cases, the issue depends crucially from the status we assign to such a notion as extended projection (Grimshaw 1991): the extended projections of V, or of N, are typically defined as head movement domains, and they do look like interpretable objects as a whole. On the other hand, the LF effects of head movement become simply unquestionable once we turn to consider movement of quantificational heads, which do acquire their scope through this mechanism: we shall consider a case of this sort in section 5.

Returning once again to Chomsky’s proposal, there is also a technical problem with its implementation if combined to some other aspects of the more recent framework: the idea of head movement being phonological clashes with the details of the derivation by phases approach to computation. Suppose a simple structure like (5) is being built by the system.

\[
\begin{array}{c}
\text{TP} \\
\text{T°} \\
\text{vP} \rightarrow \text{phase} \\
\text{v°} \\
\text{VP} \rightarrow \text{sent to Spell Out} \\
\text{V°} \\
\end{array}
\]

Suppose moreover that V° must move to T° to check some uninterpretable feature. At the level of the first strong phase vP, the sister of v (VP) is sent to Spell out. But the edge of the
current phase, $v^\circ$, is not yet accessible to phonology. In other words, the boarding site and the landing site of this head movement operation are not accessible to phonology in the same phase. A solution could be that V to T happens when the next phase (CP) is sent to PF, but then the derivation must deal with two phases at once, hence rising the computation burden, contrary to the whole spirit of the phase theory (Matushansky 2002).

Finally, pushing head movement on a phonological branch of the computation has a broader undesirable consequence for the theory, in that it complicates the grammar positing a ‘second syntax’ in the PF branch, and raises a number of questions: if both phrasal movement and head movement are driven by features, it seems at least strange to locate them on two different computations. Consider also that Chomsky (2000) assumes that PF also involves some kind of phrasal movement (cf. extrapositions: see also Lebeaux 1990), so that the problem of the duplicity of movement is reduplicated on this second syntax.

The conclusion is that there is no principled way to exclude head movement from narrow syntax, and a simple stipulation opens a number of problems. Head movement is indeed an operation available to narrow syntax and it is the default movement option being the minimal pied-piping strategy. The next step is then to find some property of head movement that makes it unavailable in standard wh-movement contexts, hence triggering the less minimal, more costly, operation: phrasal movement.

3. **Head movement vs. phrase movement**

We are now looking for the property that triggers phrase movement instead of head movement in the classical wh-constructions. The first place to look for such a property is in the definition of movement, among the conditions that constrain its various components. Let us review what these constraints on movement given the minimalist
assumptions and the feature approach look like, to see if there is any that might be responsible for the choice between the two movements.

The traditional account relies on conditions concerning Agree, the search procedure, call them locality or closeness: head movement, the default option, is constrained by the HMC (Travis 1984). Whenever the goal is not (a feature of) the head of the complement of the probe, the non default option is to be chosen and the goal pied-pipes its phrase: the HMC has been recently reframed in these terms in Pesetsky and Torrego 2001 in their head-movement generalization.

(6) Head-Movement Generalization

Suppose a head H attracts a feature of XP as part of a movement operation.

(i) if XP is the complement of H, copy the head of XP into the local domain of H;

(ii) otherwise, copy XP into the local domain of H.

This is not a viable solution for both theoretical and empirical reasons. On the theoretical side, there is no room in a feature approach for such a constraint. Notice, interestingly, that Pesetsky and Torrego arrive at (6) assuming that phrase movement is the default option, blocked in case (i) because it results in remerging. But the system cannot work this way. If movement is triggered by a search procedure of a feature (agree), locality is to be defined in terms of features with no reference to the amount of material which gets moved with the feature. This is the case of the Minimal Link Condition of Chomsky 1995 (7), or its more recent version in (8).

(7) Minimal Link Condition

K attracts α only if there is no β, β closer to K than α, such that K attracts β.

(Chomsky, 1995:311)
(8) Locality conditions yield an intervention effect it probe $\alpha$ matches inactive $\beta$ which is closer to $\alpha$ than matching $\gamma$, barring Agree $(\alpha, \gamma)$.

(Chomsky 2001)

In other words: locality is a condition on the search procedure which is a component of movement defined at the features level. At that level, the distinction between phrase and head plays no role and indeed has no room whatsoever. Therefore, the criterion for choosing between head movement and phrase movement we are looking for cannot be defined in the locality conditions$^5$.

On empirical grounds, I shall refer to the well known though controversial "long head movement" phenomena recorded in the literature (Lema & Rivero 1990; Borsley et al 1996; Carnie 1995; Roberts 1994; Manzini 1994 a. o.). To illustrate, the simple Breton sentence in (9) appears to instantiate the derivation in (10), where the verb raises to a left periphery head skipping the head I, hence violating the HMC, or the generalization (6).

(9) Lennet en deus Yann al levr  
    (Breton, Roberts 1998)  

    Read 3sgm has Yann the book  

    “Yann has read the book”
The same derivation yields an ungrammatical result if we posit a different kind of intervener, such as negation: (11).

(11) *Lennet n’en deus ket Tom al levr

Read neg 3sgm have neg Tom the book

The explanation proposed in Roberts 1998 and others is that Relativized Minimality-kind of constraints apply “across the board”, to any movement dependency, no matter whether it involves a head or a phrase. In other words, what counts for locality is the class of interveners and the class of attractees, not the amount of moved material. More data of this sort are given in (12) through (14).
The interpretation in terms of long head movement of the data briefly presented above is controversial and still discussed, but the point is that these are exactly the data we predict given a feature approach to movement: if locality is a condition on the search procedure component of movement, then it is to be defined in terms of features, or class of features. We thus predict the same kind of constraints in both head movement and phrase movement. But if Locality is not responsible for the choice between head and phrase movement, we need to look at other conditions on movement.

There are conditions on the goal, which must to be active (Chomsky 2000) and not embedded in an island. But these conditions do not seem to be relevant for our sake, since they apply to any movement operation.

Finally, we can think of a third class of constraints, concerning the other operation involved in move, namely merge copy. The condition of the uniformity of chains can be defined as such a constraint.
The uniformity condition on chains (UCC)

A chain is uniform with regard to the phrase structure status

(Chomsky 1995:253)

We shall show in the next section that this constraint is only apparently irrelevant for the head/phrase distinction we are looking for. Before turning to this, let us finish with our review of the conditions on movement. We can also think of an economy condition on Merge Copy, for which what is to be copied is the minimal amount of material necessary for convergence: (16).

Copy just enough material for convergence

This is indeed the notion of minimality I have referred to so far, when I started saying that the minimal amount of pied-piping is a head, giving head movement as the unmarked option.

The only conclusion we can draw from what we have seen so far is that there is no difference whatsoever between head movement and phrase movement, except that the former is more minimal than the latter. If we want to find a way to derive phrase movement as an option and its selection in classical wh-constructions, we need to turn to some other component of the grammar. This is what we shall do in next section.

4. **Head vs. phrase**

If the difference between head and phrase movement does not reside in the movement module, the only alternative is to rely on phrase structure theory. There is
indeed a primitive difference between heads and phrases, and we shall capitalize on this to derive the complementarity we need: in a nutshell, a head projects; a phrase is a projection. Together with the independent conditions on merge copy spelled out above (UCC in 15 and the economy condition 16), this primitive difference is enough to derive in a principled way (a) the complementarity between head movement and phrase movement; (b) the choice of phrase movement in standard wh-construction; (c) the existence of minimally different wh-constructions involving head movement.

But let us proceed step by step, trying to be really minimal: suppose a probe α on a head A at the root attracts a goal β. Given that a feature cannot be merged, some extra-material needs to be merged, call it B, and the offending feature is deleted (17).

\[
\begin{align*}
\text{(17)} & \quad \text{\begin{tikzpicture}[baseline={([yshift=-.5ex]current bounding box.center)},scale=0.8]}
  \node (a) at (0,0) {A};
  \node (b) at (-1,1) {B};
  \draw (a) -- (b);
\end{tikzpicture}} \\
\text{(18)} & \quad \text{\begin{tikzpicture}[baseline={([yshift=-.5ex]current bounding box.center)},scale=0.8]}
  \node (b) at (-1,1) {B};
  \node (a) at (0,0) {A};
\end{tikzpicture}} \\
\text{(19)} & \quad \text{\begin{tikzpicture}[baseline={([yshift=-.5ex]current bounding box.center)},scale=0.8]}
  \node (a) at (0,0) {A};
  \node (b) at (-1,1) {B};
\end{tikzpicture}}
\end{align*}
\]

The operation merge (A,B) is asymmetric in essence, so that given the configuration in (17) two things may happen: either A or B must project, yielding the two configurations (18) and (19). The element which projects is the head of the resulting phrase.

Let us combine these trivial assumptions with what we know about movement. Suppose a moved item must be identical to its copy as to its phrase structure status (Uniformity Condition on Chains 15): this means that when a head, which by definition projects, gets moved, it must project, while this does not happen when a phrase moves, which by definition is a projection. As a result, whenever a feature moves as a head all the features it is associated with (and notably the categorial feature) project. Therefore head
movement changes the feature composition of the target. When the very same feature moves as a phrase, this does not happen, and the target gets unchanged.

The economy condition on merge copy in (16) ensures that the two movements never overlap: if the grammar always copies enough material for convergence, it will select head movement unless convergence at LF (interpretation) chooses differently.

Let us now see some illustration of this account turning to (standard) Wh-constructions. In interrogatives, the complementizer head selecting the structure contains a feature, call it WH, which needs to be checked acting as a probe. The goal corresponds to a wh-element embedded in the clause, (a copy of) which needs to be merged in a local configuration with the probe. Since features cannot be merged (see section 1), the minimal option is for the wh-feature attracted by C to pied-pipe the wh-word alone (head movement). But this minimal option does not yield a convergent derivation: moving the wh-feature as a head means projecting all the features associated with it, and notably its categorial feature (D). This would turn the interrogative clause into a complex DP as in (20), which is not interpretable as an interrogative clause at the interface. This is why the most costly derivation (21) is selected.

(20) \[ DP \]
    \[ \overset{\text{wh}}{D} \]
    \[ \overset{\text{wh}}{C} \]

(21) \[ CP \]
    \[ DP \]
    \[ \overset{\text{wh}}{C} \]

Concluding so far, there is only one operation Move which is triggered by a feature and defined and that level, and which copies just enough material for convergence. Then, convergence at LF decides whether it is a head (which retains its projection property
throughout the derivation given the UCC), or a phrase (which again remains a projection throughout the derivation). In standard wh-constructions, such as interrogatives, LF convergence selects the less minimal option, that of moving the entire phrase, preserving the simple CP categorial status of the clause.

5. Free relatives and comparatives

A good evidence that the phrasal status of wh-movement is only a byproduct of the environments where its occurs (interrogative CP’s) would then be an instance of wh-movement embedded in a different environment, notably in a context compatible with a DP-selection, and showing no phrasal pied-piping.

This is exactly what we observe in free relatives. Consider the minimal pairs in (22a-b) and (23a-b), respectively in Italian and in English.

(22)  a. *Mangerò [quanto pane] vorrai[t]
         I-will-eat how-much bread you-will-want

               b. Mi chiedo [quanto pane] vorrai [t]
                  I wonder how-much bread you-will-want

               c. Mangerò [quanto] vorrai [t]
                  I-will-eat what you-will-want

(23)  a. *I shall visit [what town] you will visit [t]

               b. I wonder [what town] you will visit [t]

               c. I shall visit [what] you will visit [t]
At a very first sight, free relatives (a,c) and interrogatives (b) differ in that the wh-element needs to be bare in the former but not in the latter. This anti pied-piping restriction correlates with a different interpretation of the two constructions: as is well know, free relatives are complex nominal structures, while interrogatives are simple clauses. This is shown in a straightforward way testing the selection compatibility of the two structures: free relatives are compatible with a DP-selecting verb (24a, 25a); interrogatives are not (24b,25b).

(24)  
   a. Mangerò i biscotti  
       I-will-eat the cookies  
   b. *Mi chiedo i biscotti  
       I wonder the cookies

(25)  
   a. I shall visit Rome  
   b. *I wonder Rome

This correlation can be expressed through a generalization like (26).

(26) Generalization: a simple wh-structure excludes pied-piping exactly in those cases in which it occurs in a nominal position.

The generalization (26) receives a principled explanation given what we know about heads and phrases. What happens in free relatives is that the wh-element moves as a head. This means that, besides checking as usual the wh-feature on C, it endows the clause with the D-feature required for interpretation, as in (20).
Consider (27) for a more detailed comparison of free relatives, relatives, and interrogatives. The basic difference is minimal, but correlates with important interpretative and syntactic differences.

(27) a. \[
\begin{array}{c}
\text{DP} \\
\text{DP}
\end{array}
\]

b. \[
\begin{array}{c}
\text{DP} \\
\text{CP}
\end{array}
\]

c. \[
\begin{array}{c}
\text{CP}
\end{array}
\]

Free relatives (a), full relatives (b) and interrogatives (c) display minimal differences in their structure: relative clauses are characterized by the fact they are complex nominal clauses (a DP embedding a CP: Kayne 1994, Bianchi 1999), the difference being that D is moved in free relatives and merged in full relatives; interrogatives and free relatives involve the same uninterpretable feature WH on C, which is checked through phrase movement in the former and through head movement in the latter. The crossed arrows in (a) and (c) correspond to instances of movement violating the UCC: a phrase projecting (a)°; a head not projecting (c).

A second piece of evidence that we are on the right track, is given by a different instance of the generalization in (26), represented by comparative clauses. As we shall see, comparative clauses are wh-constructions displaying all the typical properties associated with wh-movement, together with clear nominal properties. We thus predict that comparatives should disallow phrasal pied-piping and involve wh-head movement. Let
us review one by one these properties of the construction and see how this prediction gets confirmed.

First of all, comparatives are wh-constructions, displaying all the typical properties associated with movement since Chomsky 1977. In (28) and (29), it is shown that they are sensitive to standard islands effects in English and Italian. In Italian, the comparative is moreover introduced by an overt wh-element.

(28)   a. *I ate more cookies than I met a man who ate [t]  
       b. *I ate more cookies than I wonder who ate [t]

       CNPC  
       wh-island

(29)   a. *Ho mangiato più biscotti di quanti ho incontrato un uomo che ha mangiato [t]  
       b. *Ho mangiato più biscotti di quanti mi chiedo chi ne abbia mangiati [t]

       have eaten more cookies of WH  have met a man that has eaten  
       me wonder who of-them has eaten

Second, comparatives display clear nominal properties, as has been observed primarily on the semantic side. In particular, they give rise to the kind of scope ambiguities which is typically associated with nominals. This is illustrated with a very classical example taken from Russell 1905 in (30).

(30)   I thought your yacht was larger than it is  
       a. [[than it is] [I thought your yacht was larger]]  
       b. I thought [[than it is] [your yacht was larger]]

       (consistent)  
       (inconsistent)
The easiest and most standard way to account for the ambiguity of the sentence (30) is to derive it from the scope possibilities of the comparative clause, so as to include or exclude the belief verb from the comparison, as informally represented in 30a-b (von Stechow 1984).

The nominal character of comparatives is also visible in their interpretation, which is that of amount or degree *descriptions*, unlike corresponding wh-interrogatives, which are simple amount or degree *predicates* (Heims 1985, Donati 2000 a.o.). Again this is represented informally in (31).

(31)  Mary ate more cookies than Paul did

        \[ \exists y \left[ (y \succ x \left[ \text{Paul ate } x\text{-many cookies} \right] \land \left[ \text{Mary ate } y\text{-many cookies} \right] \right] \]

On the syntactic side, they are strong islands for extraction, just like all complex nominal clauses, but unlike standard wh-constructions: (32).

(32)  a. *When do you eat more cookies in the morning than Paul does [e] [e]?* 

        b. *What do you eat the soup more quickly than Paul does [e] [e] ?*

Given that comparative clauses are wh-structures and they display these nominal properties, the prediction is clear: comparatives should involve WH-head movement. This is exactly what is shown by data in (33) and (34), respectively in Romanian and Bulgarian.

(33)  a. *Maria e cu mult mai desteapta [decît de frumoasa] e Zamfira  \[\text{Grosu 1994}\] \[\text{Maria is with much more clever of-WH of beautiful is Zamfira} \]

        *Mary is much more clever than beautiful Zamfira is*
b. Maria cu mult mai desteapta [decît] e Zamfira de frumoasa

Maria is with much more clever of-WH is Zamfira of beautiful

Mary is much more clever than Zamfira is beautiful

c. [Cît de frumoasa] e Zamfira ?

WH of beautiful is Zamfira

How beautiful is Zamfira?

(34) a. *Ivan izpi povece vino ot-[kolkoto bira] Maria izpi (Izvorski 1996)

Ivan drank more wine of-how-much bear Maria drank

b. Ivan izpi povece vino ot-[kolkoto] Maria izpi bira

Ivan drank more wine of-how-much Maria drank bear

c. [Kolko bira] izpi Maria?

How-much bear drank Maria

d. *[Kolko] izpi Maria bira?

How much drank Maria bear

(33) and (34) show that the very same wh-element (cit in Romanian; kolko in Bulgarian) moves as a head in comparatives (a-b), but as a phrase in interrogatives. This provides good evidence that the wh-element involved in the embedded comparative is a Q-head, which endows the clause with the Q feature required for interpretation. The same evidence holds also English and in Italian. In English however, the data are partially obscured since the wh-element is null and the contrast with interrogatives in not minimal:

(35)  a. *Mary ate more cookies than candies she ate
b. Mary ate more cookies than she ate candies

c. How many candies did she eat?

d. *How many did she eat candies

In Italian, as we shall see in the next section, the movement pattern illustrated so clearly by the Romanian and Bulgarian data is complicated by some gap in the paradigm which needs an explanation.

All the data discussed in the present section provide strong evidence for the following: comparatives and free relatives on the one hand, interrogatives on the other hand display a minimal syntactic contrast which correlate with a semantic difference. While all involve wh-movement, triggered by the same feature and constrained by the same restrictions, free relatives and comparatives move a bare head which changes the CP configuration domain in a DP/QP category, turning the clause into a complex nominal; interrogatives move a phrase, which leaves the CP configuration domain unchanged as a simple (interrogative) clause. This is what we expect it movement is an operation copying just enough material for convergence, and heads and phrases differ only for their projection ability.

5. The freezing effect of agreement

As I mentioned, Italian comparatives require a last notice. In Italian, while there is clear evidence that the wh-element quanto moves as a head (see 36), the situation is complicated by some gaps in the paradigm: in particular, quanto never strands its nominal complement (37a), except when it is cliticized with ne (37c) or when it shows no agreement with its adjectival complement (37b).
(36) a. *Maria ha mangiato più biscotti di **quante caramelle** Paolo abbia mangiato
   Maria has eaten more cookies of WH-pl.fem. candies Paolo has eaten

b. Maria è più intelligente di **quanto bella** sia
   Maria is more intelligent than WH beautiful is

c. Maria ha mangiato più biscotti di **quanti ne** abbia mangiati Paolo
   Maria has eaten more cookies of WH of-them has eaten Paolo

(37) a. *Maria ha mangiato più biscotti di **quante** Paolo abbia mangiato **caramelle**
   Maria has eaten more cookies oh WH-pl.fem. Paolo has eaten

b. Maria è più intelligente di **quanto** sia **bella**
   Maria is more intelligent than WH is beautiful

c. Maria ha mangiato più biscotti di **quanti ne** abbia mangiati Paolo
   Maria has eaten more cookies of WH of-them has eaten Paolo

If we compare this data with those of Bulgarian, the generalization seems to be that agreement within the QP has a sort of a freezing effect, blocking head movement of the Q alone and the corresponding stranding of the remnant of the phrase. The same effect seems at work in the free relative cases discussed above: both in English and in Italian (38,) the D-head can never strand its nominal complement".

(38) a. *Mangerò **quanto** vorrai **pane**
   I-will-eat how-much you-will-want bread

b. *I shall visit **what** you will visit **town**

This kind of gap in the paradigm is very important for the hypothesis under discussion here, because it appears like a residual of a peculiarity for head movement, which would contradict our tentative of unification. We shall see however that this constraint has
nothing to do with head movement *per se* and that we find the same kind of effect with phrase movement. We shall therefore end up confirming our claim that movement is a unitary operation. But let us proceed step by step.

Consider an agreeing QP like *quante caramelle* in the comparative clause (37a).

(39)   

Suppose Q has an uninterpretable feature that needs to be checked by a feature of N. Head movement here is not triggered because it is useless: the head-head checking configuration is already there, merged in the base. Since agreement feeds pronunciation, this configuration must be conserved until spell out. This means that at the level of the first strong phase (vP) Q and NP must be adjacent. But at that point, VP, the sister of v, becomes inaccessible to the computation (PIC\textsuperscript{12}). There is therefore no way of moving Q without N: if it moves before spell out agreement cannot feed pronunciation; after that, it
violates the PIC. The restriction on movement at play in (37) is not due to a peculiarity of head-movement, but to the effect of a phonological adjacency requirement.

Crucially, a similar effect is attested with phrase movement, confirming that it has nothing to do with head movement per se. As is well known, preposition stranding is prohibited in those languages, like French and Italian, where P and D tend to conflate in a P+D suppletive form (Law 1998, Salles 1997). In (40) this restriction against preposition stranding is shown at play in an Italian interrogative.

(40) *Cosa hai parlato [di t]?

What you-have talked of

Consider the corresponding simplified structure in (41).

(41) CP

```
CP
  C° ...
  vP -----> phase
  VP ----> sent to Spell Out
  V° ...... PP
  P      DP
  di D  cosa
```
If P and D need to conflate, this means they need to be adjacent at Spell out. But at the first strong phase, when vP is sent to spell out, VP becomes no longer accessible to the computation: the DP *cosa* cannot make a long movement alone, and preposition stranding is barred.

6. **Conclusions**

The theory of movement as it emerged from *On WH movement* is still the core of the current theory of movement. This paper focuses on what appears like a stipulation of such a theory at the light of the minimalist framework, namely the limitation of wh-movement to phrases, and aims at reviewing critically the more general division of movement into two distinct subtypes: head movement and phrase movement.

Starting from a critique of Chomsky’s recent tentative to ban head movement from syntax, and stepping through a review of the different constraints movement is restricted by and their legitimacy in a minimalist framework, I conclude that there is only one movement, triggered by features and defined at that level. The different amount of pied-piping which happens to hold in different environments is due to an economy principle on movement which states that you merge just enough material for convergence. Given that principle, the effect of two distinct movements can be derived from the irreducible difference which holds by definition between heads and phrases: heads project, phrases are projections. A natural implication of this conclusion is that there is no principled reason for wh-movement to be restricted to phrases in any context: two cases of wh-head movements are discussed as an illustration.
NOTES

1. There is a highly detailed and motivated recent literature dedicated to covert phrasal movement: see Sauerland 1998, Fox 1999, Nissenbaum 2000, Cecchetto 2000, Pesetsky 2000, among others. See Brody 1995 for a different view also discussing ACD cases.

2. See Pesetsky 2000 for a different view, where both features movement and phrasal movement are available to syntax as different operations.

3. The phenomena related to Romance cliticization are outrageously simplified in the text, which is meant only to show that the standard analysis may be complicated by a wrong theoretical assumption (i.e. that affixation is head movement) which we might want to get rid of. (4) is not meant to be a full-fledged alternative analysis of cliticization, of course, an analysis which should for example take into account the differences among clitics standardly reduced to the distinction between a class of syntactic clitics and one of phonological ones.

4. See also Zwart 2001 for a discussion of some clear interpretive effects of instances of verb movement.

5. In fact there is a sense in which a head is indeed always more remote than its phrase, concerning the number of nodes: fewer nodes separate the probe from the phrase bearing the goal than separate it from the corresponding head. But this notion of locality is too restrictive: it amounts to exclude head movement as a whole, which is not what we want.

6. There is an alternative which I shall not explore here being less minimal than the one I chose to pursue in the paper. Suppose head movement and phrase movement are systematically triggered by two intrinsically different classes of features, then it is
conceivable that they are sensitive to different interveners and display different locality constraints. See Matushansky 2002 for a proposal along these lines.

7. The economy condition in (16) also makes an interesting prediction concerning movement of a feature associated with a given category to a target associated with the same feature. Since in this case the projection property of heads does not affect the categorial status of the target, the minimal option should be compulsive, and phrasal movement systematically excluded. More precisely, we predict derivations moving a DP into Spec, DP, or a PP into Spec, PP, or a CP into Spec, CP to be ruled out as violations of the economy condition (16). These means that standard analyses of constructions like the following need to be revised.

(i) Possessives:

(ii) Exclamatives:
    Into the dungeon with the traitor!  (PP in Spec, DP: cf. Barbiers 1995)

(iii) Clitic Left Dislocations:
    Che fosse un assassino, l’ho sempre saputo (CP in Spec, CP: cf. Cecchetto 1999 a.o.)
    that was an assassin, it (cl)-have-1s always known

Addressing the details of such a revision goes far beyond the limits of this paper. Notice that some idiosyncratic properties of these constructions would benefit from an analysis where the dislocated constituent is adjoined to some extra functional head: in particular this would explain the impossibility of recursion in both exclamatives PP’s and CILD’s, and the exclamative mood of examples like (ii). Thanks to Peter Svenonious for pointing these facts to me.
8. There is an exception to this generalization that easily comes to mind: both in English and in Italian, there is a class of free relatives which appear to allow some amount of pied-piping:

(i) Mangerò [**qualunque** pane] vorrai[t]

   I-will-eat whatever bread you-will-want

(ii) I shall visit [**whatever town**] you will visit [t]

   The fact that this movement pattern correlates with the presence of some extra-material in the head of the clause, the suffix *ever-/unque*, suggests an obvious solution, namely that these relatives are only apparently “free”, i.e. defective, but rather correspond to full relative clauses. As such, they are generated as the complement of an external determiner (Kayne 1994, Bianchi 1999), the universal quantifier *-ever-/unque*.

(iii) I shall visit [**[DP[[D[ever] [CP[DP[what town]]]]]]**]

   For some reason perhaps due to its universal value (Larson 1987), *unque/ever* triggers the raising and head adjunction of the wh-determiner, yielding (iv).

(iv) [**[DP[[D[what] [D[ever]] [CP[DP[t] town]]]]]]**

   See Battye 1989 for a series of empirical arguments demonstrating that these relatives ought to be treated as “pseudo free relatives”. See also Kayne 1994:154n for a similar analysis proposed on totally different grounds.

9. There is more than what is said in the text, of course: while the definition of heads as projecting features is certainly correct, that of phrases as non projecting features leaves adjunction out of the picture. This could be a problem for the analysis proposed here if a derivation such as (i) were allowed.
In (i) a phrase XP is moved and it projects in its landing site, turning the root in a right adjunct. There are however good reasons to exclude such a derivation on independent grounds. One concerns the legitimacy of the chain formed in (i), which violates a strict version of the UCC, since XP is maximal in its base position, but non-maximal in its landing site (see Chomsky 1995:257-259 for a discussion). A different kind of reason concerns the structure in (i), which is a case of right-adjunction and questionable as such (Kayne 1994). Thanks to Howard Lasnik to point this potential problem to me.

10. Notice that Grosu 1994, where these data come from, propose a different analysis of the movement pattern illustrated.

11. I have been reported of Greek free relatives where stranding is indeed attested, and no agreement crucially is involved (Agbayani, p.c.).

12. Phase Impenetrability Condition
   In a phase $\alpha$ with head H, the domain of H is not accessible to operations outside $\alpha$, only H and its edge are accessible to such operations.
   (Chomsky 2001).
REFERENCES


