

# Vector Effects on *Wh*-Extraction

Louis-H. Desouvrey

lhd33 at yahoo dot com, ldesouvrey at hotmail dot com

January 2008

This paper discusses *wh*-elements in embedded clauses. Matrix clauses and complement clauses are generated as stand-alone structures, glued to one another with a relative anaphor, known as complementizer in current generative theory. When a *wh*-operator is the subject in the complement clause, it triggers the violation of some constraints that cannot be repaired through normal movement because *wh*-operators are vectors in the sense they exhibit scope and direction with respect to the matrix anaphor they antecede. I show that the grammar uses an alternative input in which the complement clause is transformed in a relative clause, which obviates the use of the matrix relative. An operator in object position does not control any antecedent in the matrix clause, and therefore may be directly fronted in order to avoid an OCP violation.

**Keywords:** relative pronoun, relative anaphor, OCP, reference, vector, scalar, French dummy matrix clause, complex operator.

## 1. Introduction

Syntactic phenomena can be best explained by taking into account a set of grammatical distinctive features whose interactions are governed by constraints. In Desouvrey (2007), it is shown that extraction of *wh*-elements from root sentences in English and French obey a few constraints (see (1)) acting on feature specifications of specific morphemes.<sup>1</sup> This paper further explores the behavior of *wh*-elements in the context of embedded clauses, i.e., embedded questions (or long *wh*-extraction) and indirect questions. Approaching complex sentences as a series of independent and complementary structures connected by reference, I will show that *wh*-movement across subordinate clauses falls in the realm of constraints (2) and (3), which subsumes the superiority condition (1c). It will be shown that subject *wh*-operators cannot move from complement clauses. Constraint violations that trigger *wh*-movement in root clauses are avoided in subordinate clauses by an alternative input in which the matrix relative (also known as complementizer) becomes useless and disappears altogether. As a result the arguments of the matrix clause and the complement clause are swapped, which makes *wh*-fronting possible.

- (1) a. Obligatory Contour Principle (OCP)  
Two elements with identical Case features are not allowed in the same syntactic domain.
- b. Obligatory Contour Principle-prime (OCP')  
No elements can stay between two other identical feature-bearing elements. (That is, identical feature-bearing elements must be string-adjacent.)
- c. Superiority Condition (SUPC)  
The relative positions of [ $\sigma$ ]-specified elements in the clause must be preserved throughout the derivation.
- (2) Reference Inalterability (RI)  
A referring element must keep the same antecedent throughout the derivation.
- (3) Vector Effects (VE)  
Syntactic elements are either vector or scalar. A vector has a scope and a direction (of spreading):
- i. Scope
- (a) A vector has scope in its structure. If there are two or more vectors in a structure, their scope is related to their position, such as a vector takes scope over every vector to its right.
- (b) Scope is inalterable in the derivation: a lower vector can't move past a higher vector.
- ii. Direction of R-features
- (a) Direction of spreading is one-way through the derivation.
- (b) A vector can only spread a feature to an element within its scope.

Unlike other constraint-based theory, for instance Optimality Theory (cf. Prince and Smolensky 1993), constraints are not absolutely violable. That is, when a constraint is violated in the input, the grammar allows its repairing by movement, which may give rise

to a few derivational steps. If movement cannot solve the problem, a derivation can proceed from an alternative input. In any event, in the process of repairing a constraint violation in the input, a new constraint may be violated and the derivation would pursue until it reaches a perfect state where no constraint is violated. However, some constraints can appear violated in the output. Either they cannot be repaired at all, in which case the sentence is spelled out with a pitch movement, or they can, but the derivation runs out of time. The length of a derivation may not exceed three derivational steps by virtue of what I shall refer to as the Derivation Extension Number, (4) (see also Desouvrey 2007). When the maximal value of  $\varepsilon$  is reached, the derivation is cutoff and spelled out. This value is arbitrary to the extent that one can't say at this point why it is not smaller or greater; nonetheless it makes it possible to constraint further the grammar; for a derivation has to have an end independent of linguist judgement.

(4) Derivation Extension Number ( $\varepsilon$ )

A derivation can give rise up to three steps, including the input:  $\varepsilon \leq 3$ .

The general outline of this article is as follows. In the next section, I briefly present the structure of relative clauses proposed in Desouvrey (1996, 1997), which is crucial for the treatment of *wh*-extractions. In section 3, I turn to account for long-extraction in French, and then in English, showing that fronting of embedded subject operators is realized through a relativization operation. In section 4, it is shown that *do* cannot appear in indirect interrogatives in English, because there is no *wh*-movement across the clause. In section 5, I discuss the fact that interrogative structures in colloquial French is normally embedded under a dummy matrix clause, and then I conclude the paper (section 6).

## 2. The structure of relative clauses

The structure of relative clauses I assume in this paper is significantly different from current assumptions in the literature. It presents two main features: (a) the so-called complementizer is treated as a relative pronoun/anaphor, and (b) clauses in a complex



(8) [Le taxi<sub>i</sub> est venu] [Marie avait insulté que<sub>i</sub>]

Notice that it is not necessary for the whole relative clause to move if the head noun is at an edge of its structure. In such a case, only the relative pronoun adjoins to the head noun, as seen in (9).

(9) a. [Il appelle le taxi] [Marie a insulté que]  
 b. [Il appelle le [taxi=que Marie a insulté t]]  
 He calls the taxi that Marie has insulted.

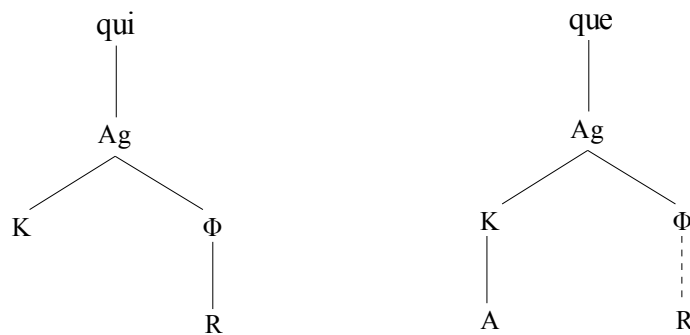
A benefit of this analysis is that it can be extended to complement clauses, which depend on matrix verbs; see (10). Unlike traditional analyses which assume that complement clauses include a complementizer, I claim that the latter is a relative pronoun, complement of the matrix verb, and therefore the matrix clause has to be seen as a relative clause.<sup>3</sup> In (10), the antecedent of the relative is intended to be a clause, but is actually the closest referring element, namely the subject of the complement clause. The complement of matrix verbs has to be mediated by a relative pronoun in order to prevent the subject of the complement clause from receiving accusative Case from the matrix verb and nominative Case from the embedded verb. Thus, since unlike the familiar relatives, the matrix relative modifies a clause (as opposed to a single head noun), it cannot be incorporated to any head by adjunction. We may note that, since relative *que* is specified for accusative Case, as is the verb selecting it, OCP is violated in the matrix clause, a point to which I shall return shortly.

(10) [Tu crois que<sub>i</sub>] [Marie<sub>i</sub> a insulté Paul]<sub>j</sub>

In the theory of coreference proposed in Desouvrey (2002), a referring element without a R-node is an anaphor and a referring element with a bare R-node is a pronoun, whereas full referential expressions have an R-feature which depends on the R-node. In order to capture the referential scheme described in (10) with indexes, I claim that French relative *que* and its allomorph *qui* have the feature tree in (11). Relative *qui* is Caseless, and

therefore its Ag(reement) node dominates a bare K-node, unlike relative *que*. The latter further distinguishes from *qui* by its floating R-node, which is intended to mean that it can function both as a pronoun and an anaphor. Under (11), *qui* is expected to be used either as complement of oblique verbs, which lacks accusative Case, or as subject, in which case it is assigned nominative Case by the verb, as seen above. Relative *que* cannot be used in those contexts, since feature mismatch is not allowed, as discussed in Desouvrey (2000).

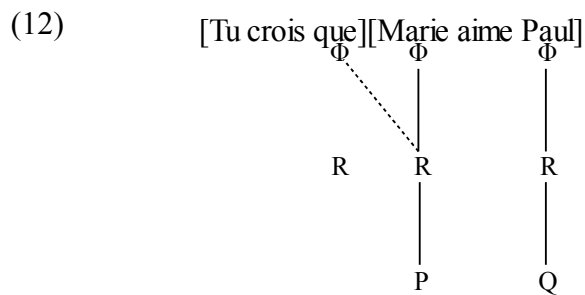
(11)



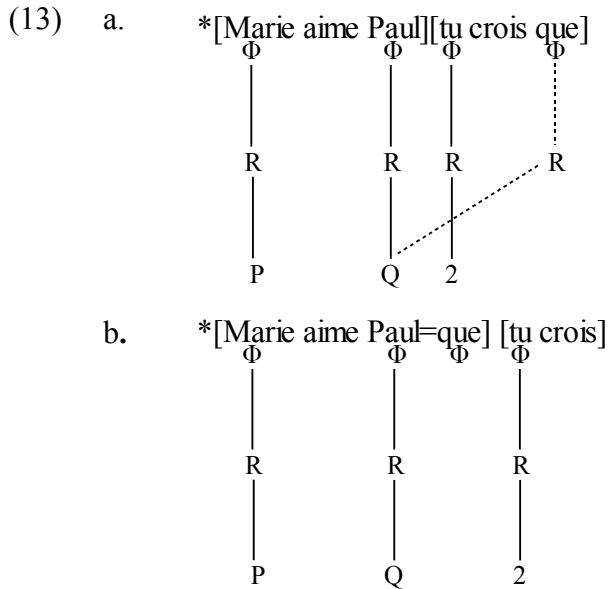
The phi-nodes hold gender, person and number features, but they are not specified with actual features, which make them capable of reflecting any referring element bearing such features. This explains the fact that relative elements must usually be adjacent to their antecedent in order to avoid ambiguity. Relatives *que* and *qui* are only compatible with elements that are specified for phi-features (see below). Elements specified for such features are scalar, and they oppose elements which are specified for omega features, namely  $[\sigma]$  and  $[\pi]$ , as discussed in Desouvrey (2000). Omega-specified elements will be referred to as vector, as defined above, for reasons that will become clear shortly.

Given these feature trees, a structure with a matrix clause and a complement clause can be represented as seen in (12). Notice that the phi-node of NPs like *Mary* and *Paul* expands to terminal features of gender and number (not shown in (12)) that appear in a different plane from the R-node. Also, Case features of the verbs, though important, are conveniently omitted in (12) and throughout. In this structure, the relation of coreference is realized by spreading: matrix relative *que* is assigned an R-node by the subject of the complement clause, *Marie*, and its floating R-node is ignored. Notice that if *Marie* were to

assign its P-feature to the R-node of *que*, the relative would be a pronoun. However, such an alternative is excluded by an instance of the constraint against ambiguity (cf. Desouvrey 2002), which rules out the use of pronoun whenever an anaphor is available. In effect, a phi-node is protected by the no-line-crossing requirement, given that all R-nodes are in the same plane, while terminal features (P and Q) appear in different planes, and therefore can freely spread to whatever available R-node.



Now one can answer the question why both clauses have to be in that order, especially why OCP is allowed to be violated in the matrix clause (recall that both the verb and relative *que* are accusative-specified). Consider structure (13), where the clauses are in the reverse order. The matrix relative can only be antecedented through its R-node, since its phi-node is not accessible, given the NLC requirement. Therefore it is construed as a pronoun. Now OCP can be resolved by movement of the relative outside the verb domain, since a pronoun can be reached wherever it is by an antecedent. Configuration (13a) is analogous to (9a), and the relative pronoun would be interpreted as an adjunct to *Paul*, as seen in (13b). In such a context, the interpretation can only be that the matrix clause is a simple relative clause headed by an NP, not by a clause. To put it differently there is no more structural difference between a matrix relative and an ordinary single-headed relative clause. Thus the interpretation that the whole complement clause is the intended antecedent would no longer be possible.<sup>4</sup>



To conclude, in (12) the relative may not move under the assumption that an anaphor is better than a pronoun. In (13) the relative starts the derivation as a pronoun, and has no reason to oppose the movement triggered by OCP. Thus, it adjoins to the closest antecedent, *Paul*, and the result is ruled out at the output, for interpretive reasons; the relative may not incorporate to an argument, while the whole clause is intended to be its antecedent.

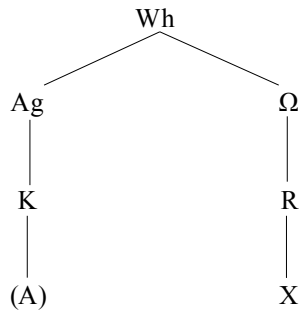
### 3. Vector extraction

First of all, I suggest that the tree structure of *wh*-elements includes an omega-node, which expands to three features, [R], as well as [ $\sigma$ ] and [ $\pi$ ], not shown for simplicity, as seen in (14). I take *wh*-elements to be fully referential, i.e., unlike third person pronouns, their R-node hosts an R-feature, conveniently named X. Feature  $\sigma$  represents the scope of the vector and  $\pi$  prevents it from changing tier by adjunction (or incorporation) to another element (cf. Desouvrey 2000, 2007). In addition, (14a) shows that an Ag node expands to a K node (Case) which may or may not end with an accusative feature, according to the morpheme. Subject *wh*-elements in French and English must have an Ag node mothering two nodes, a bare K-node and a phi-node that holds the phi-features. This is for two reasons: *wh*-subjects are not Case-specified, and they must phi-agree with verbs, as seen in (14b).<sup>5</sup> A transitive

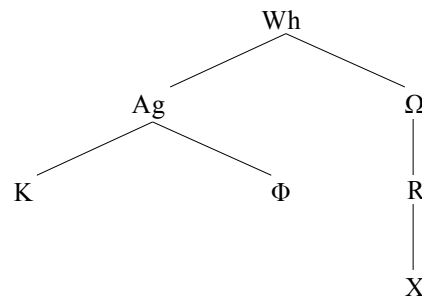


verb has two Ag nodes, one for the subject and one for the object. AgO holds the accusative Case, while AgS holds nominative Case and phi-features, as seen in (14b,c). English differs from French in that verbs are also specified for omega, (14d) (cf. Desouvrey 2007). In both languages, if the verb is infinitive, it loses its AgS node, hence the dependents of the latter.

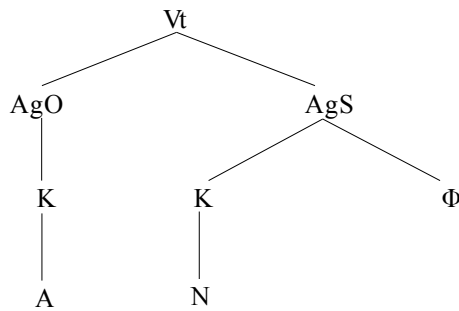
(14) a.



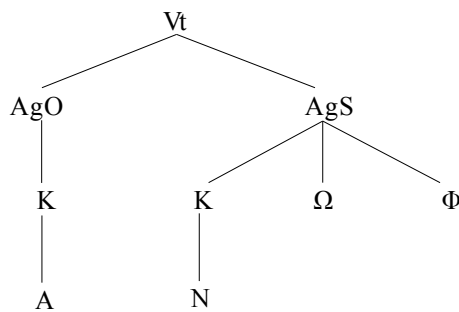
b.



c.

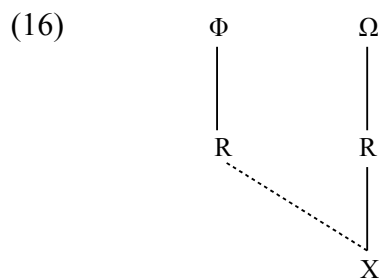
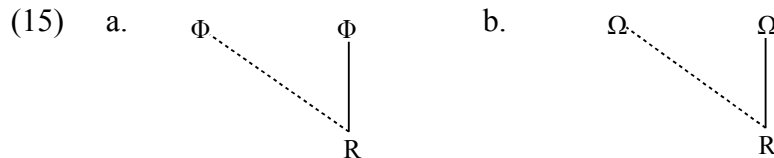


d.



The difficulty with *wh*-extraction is due to the vector effects as well as agreement between the relative anaphor and its antecedent. An anaphor has no R-node and must be assigned one by a nominal or a *wh*-element. If the anaphor is a scalar, i.e., it has a phi-node, it can receive only a R-node from a scalar element. That is to say, an R-node originating

from a phi-node can only spread to a bare phi-node, while an R-node originating from an omega node can only spread to a bare omega node, as seen in (15a,b). An omega-dependent R-node cannot spread to a bare phi-node, and vice versa. Of course, on this view an anaphor that is specified neither for omega nor phi would be compatible with both vector and scalar R-nodes. On the other hand, pronouns can be assigned an R-feature by any type of element, since R-features 'see' only R-nodes, (16), at least in English-like languages.<sup>6</sup> Vector effects in French and English are discussed in turn.

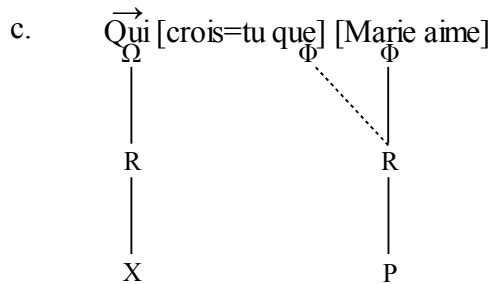
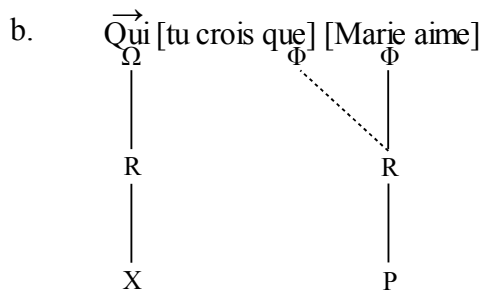
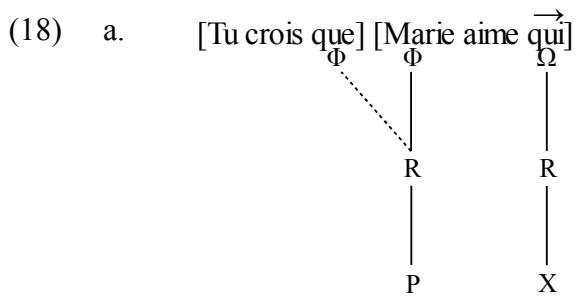


### 3.1 French

As is well known in French, *wh*-extraction of a subject argument from a subordinate clause displays the so-called *que/qui* alternation, while *wh*-object extraction does not show any particular complexity. I claim that the simplicity of *wh*-object extraction is due to the fact that the object operator is not the antecedent of the matrix anaphor. Consider sentence (17). Subject inversion occurs in the matrix clause and is a consequence of the fronting of the embedded operator, as discussed in Desouvrey (2007). Nothing occurs in the embedded clause, except of course the gap caused by the movement of the operator. This state of affairs can be simply captured by taking structure (18a) to be the natural input for this sentence. (Conveniently, irrelevant features are omitted.) Since vector *qui* and the verb are A-specified, OCP is violated in the complement clause, which forces the movement of the

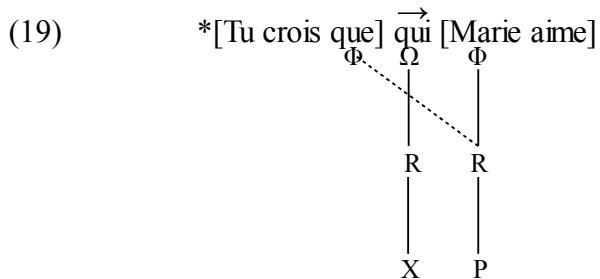
operator.<sup>7</sup> Then under OCP', the subject adjoins to the right edge of the verb, yielding the desired result in (18c). These processes, or OCP cycle, namely movement under OCP, and then under OCP', is discussed in Desouvrey (2007).

- (17) Qui crois-tu que Marie aime?  
 Who think you that Marie likes.  
 'Who do you think Marie likes?'

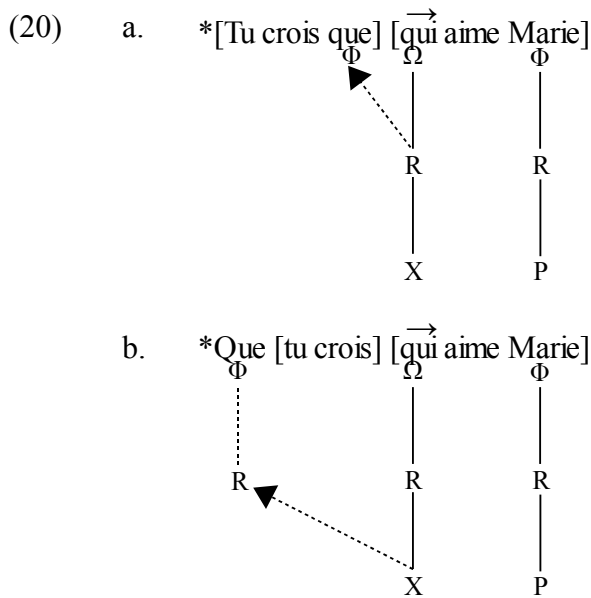


Now, in French and in English (see below), the operator may not end up to the left edge of the complement clause. The present system makes it possible to account for this fact in a principled way. From (18a) movement of the operator in between the clauses

would create a fatal line crossing effect, as seen in (19). The fact is, the *wh*-operator can't supersede the subject of the complement clause as the antecedent of the relative, since coreference relation may not be modified in the course of the derivation, consistent with (2).



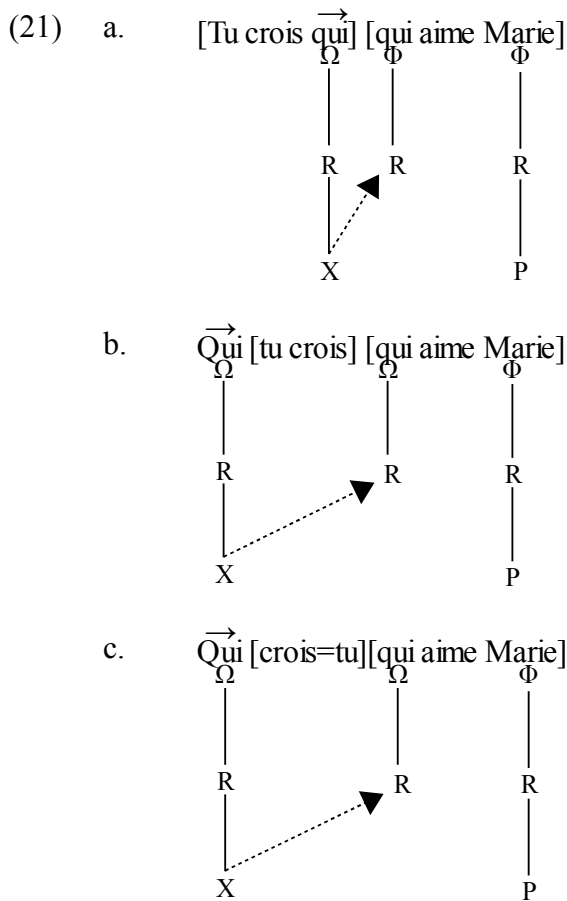
However, when the operator is the subject of the complement clause, it is the antecedent of the matrix relative, as seen in input (20). Since the relative enters the derivation as an anaphor, its phi-node clashes with the R-node of the operator, and therefore the structure is ruled out. Can the operator instead target the R-node of the relative in order to avoid the feature clash? If the matrix relative is used as a pronoun, it need not be adjacent to the operator; it would move under OCP to the front of its clause, without losing its antecedent, as seen in (20b).<sup>8</sup> If this were possible, the relative would then be the most prominent in the structure. Suppose that whenever possible, the operator must target the first position of the clause. Since only a pronoun can move away from its antecedent, the operator has to treat the relative as an anaphor in order to prevent it from having more prominence. Therefore, an alternative input must be used, which will allow the operator to move to the front of the matrix clause, as shown immediately. We will see that this analysis is confirmed by the fact that in another context the operator will target the R-node of the relative in order to free itself from the relative, so that it can be fronted under OCP.



Furthermore, we may note that the operator cannot move directly to the front of the matrix clause, since vector redirection is not allowed. In effect in the input, the operator controls the anaphor from the right, and after movement it would come to control it from the left, which is ruled out by a vector effect. Recall that movement of an object operator is regular, since in such a position an operator does not control anything in the structure.

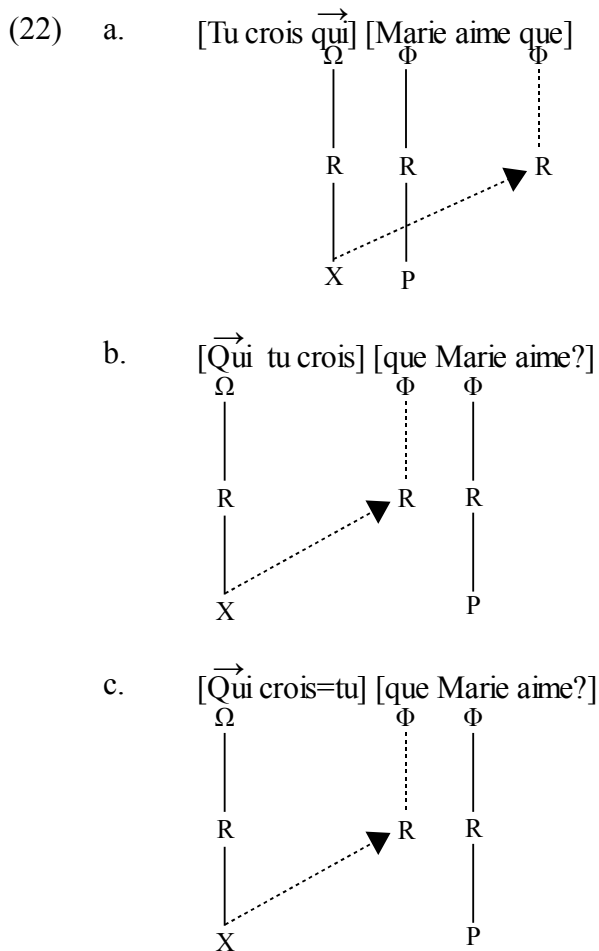
To overcome the difficulties inherent to input (20), namely feature mismatch, vector limitations, I suggest that the grammar uses an alternative structure that consists in relativizing the complement clause to the operator, which renders useless the matrix relative. I take the matrix verb to select a clausal complement, as opposed to an NP; the complement clause cannot be directly merged with the verb, so a relative element is used as a mediator. Now if the complement clause is a part of a complex NP, it is no longer the argument of the matrix verb; only its head, i.e., the operator, counts as an NP argument, and therefore the mediation by the matrix relative can be dispensed with. Thus, the derivation starts with (21a), a *salva veritate* substitution input in which the operator is reanalyzed as the object of the matrix clause, while the complement clause becomes a relative clause modifying the operator.<sup>9</sup> The operation that derives this input is pre-derivational under the assumption that further access to the lexicon is banned once the structure is built. Since the

operator is the object of the matrix verb, it appears under its accusative allomorph, and therefore induces an OCP violation in the matrix clause. Since the relative is the subject of the complement clause, it appears as the Caseless allomorph, as seen above. From (21a), the operator moves to the front of its clause, as seen in (21b), where it triggers leftward movement of the subject (under OCP'), yielding the correct result in (21c). We may note that in (21), the operator moves to the front of the matrix clause, but its R-feature is not redirected.



This process, namely extraction by relativization, if applied to the case of object extraction discussed above, would apparently yield the same result. Thus instead of input (18a), the derivation would proceed from the alternative input in (22a). The operator and

the relative, which takes its accusative allomorph, moves to the front of their respective clause, and further the operator triggers subject inversion in the matrix clause, yielding (22c), the desired result. Notice that in (22), and more generally when the relative is in the complement clause, it enters the derivation as a pronoun in order to avoid a line-crossing effect.



In the present system, a harmonization process should normally favor derivation (22) over (18). By virtue of such a process, the grammar would symmetrically use the same strategy to extract subject and object, unless some other constraints are at stake.<sup>10</sup> Indeed, the relativization strategy would not produce a well-formed sentence in every case of object extraction. Some French verbs take an indirect object complement, introduced by

preposition *à* 'to' in most cases. The preposition bears an oblique Case, allowing the argument to agree with the verb, and therefore it cannot be stranded by the argument. Under the relativization strategy, the *à*-phrase would clash with the accusative-specified matrix verb. With a long movement, however, one obtains the desired result, as seen in (23). Under the assumption that the grammar does not use implicational constraint, the relativization strategy must be ruled out for all case of object extraction. Otherwise, an OCP-triggered movement would have to consider whether the operator is a direct object or an indirect object.

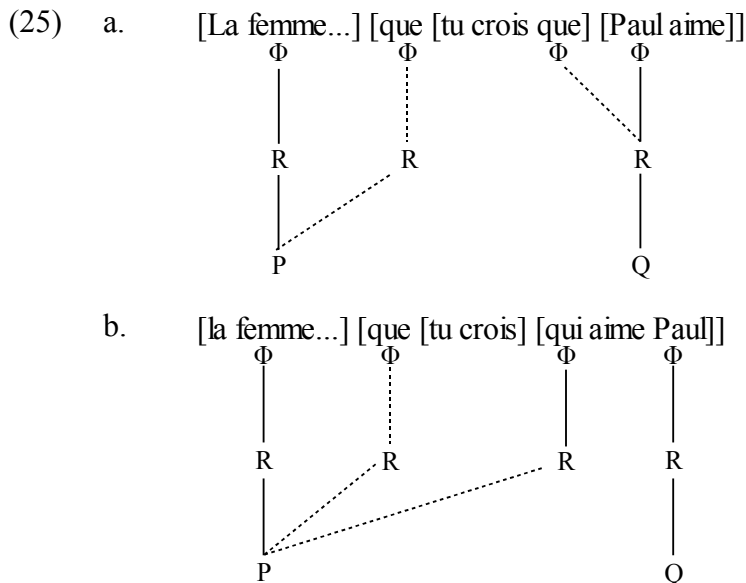
- (23) a.  $\bar{A}$   $\overrightarrow{\text{qui}}$  [crois-tu que] [Marie nuit?]  
 b. \*Tu crois  $\bar{a}$   $\overrightarrow{\text{qui}}$  [Marie nuit  $\bar{a}$  qui?] (failed input)

Now suppose that the complex structure discussed above is itself a relative clause to a scalar head noun, as illustrated in (24). Each structure contains two relative elements: the matrix anaphor and a relative pronoun (*que* in (24a) and *qui* in (24b)), both standing for the scalar head *la femme*. In the case of (24a), the object relative in the complement clause moves directly to the front of the matrix clause, as seen in (25a). As for (24b), the subject of the complement clause is relative pronoun *qui*, which is the antecedent of the matrix relative. Therefore the complement relative clause has to be relativized to the relative pronoun, just as in the case with an operator. Under relativization, the matrix relative comes to refer to the extracted relative pronoun, no longer to the relative clause, and therefore disappears. The extracted subject relative pronoun becomes the object of the matrix clause, and takes its accusative shape, *que*, then moves to the front of the matrix clause. Ultimately, the whole complex relative clause must adjoin to the head noun *la femme*, a step conveniently omitted. In this example, where the complement clause is a relative clause, the relativization process applies vacuously, since the result of the process is no different from the natural input. It might well be possible that the grammar uses a shortcut; since matrix *que* refers to *qui*, which refers to *la femme*, it follows that matrix *que* refers also to the same head noun. Therefore, the matrix relative enters the derivation as a pronoun, getting its R-



feature directly from the head noun, as seen in (25b), the result of either analysis.

- (24) a. La femme [que [tu crois que][Paul aime]]...  
 The woman that you think that Paul likes...  
 b. La femme [[que tu crois][qui aime Paul]]...  
 The woman that you think that likes Paul...



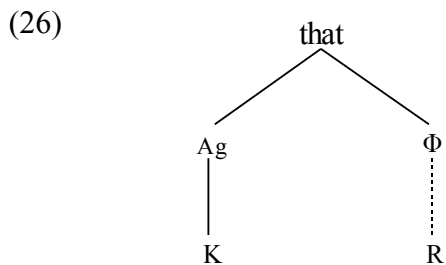
To conclude, an object argument can move directly from the embedded clause to the front of the matrix clause, while a subject argument has to be extracted by relativization, given the vector effect that occurs with operators. Similar phenomena in English will confirm this analysis.

### 3.2 English

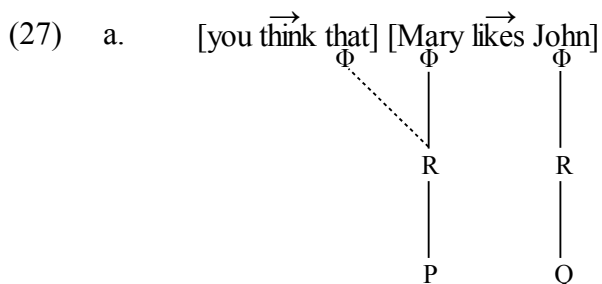
While in French only *wh*-operators are vectors, in English tensed verbs, including auxiliaries and modals, are vectors as well (cf. Desouvrey 2003, 2007). Therefore, *wh*-operators and verbs are expected to interact in the sense of the set of phenomena referred to above as vector effects.

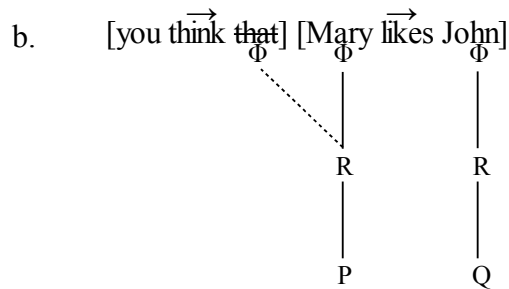
Central to the analysis I am proposing is the assumption that English, unlike French,

has one Caseless matrix relative, namely *that*, and therefore it can be used either as subject or object, given the feature structure in (26). In addition, the English relative can be null in object position. Deletion of subject *that* is presumably ruled out by the constraint against ambiguity (cf. Desouvrey 2002). In effect, the distinction between the main verb and the subordinate verbs would be lost, as clause boundaries fade away.<sup>11</sup>

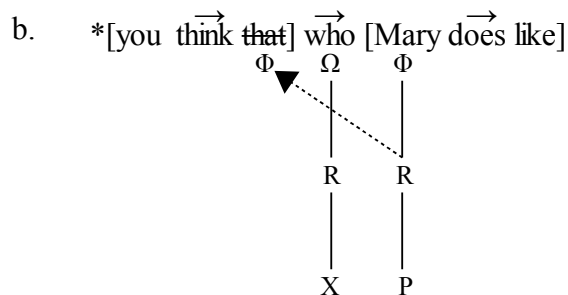
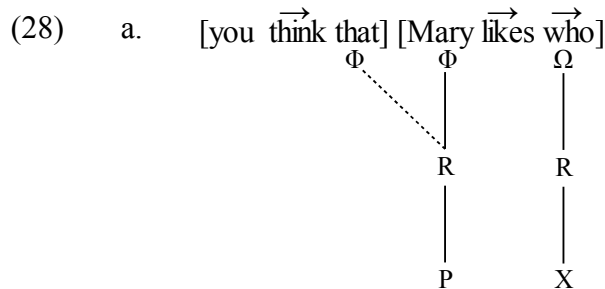


Thus, it appears that in English the matrix relative can be freely deleted, unlike French. Therefore, both (27a) and (27b), with a deleted relative, are equally good.<sup>12</sup> I assume that the deletion operation consists in delinking the segments from the skeletal slot, leaving intact the feature tree of the morpheme. As a result, simple deletion cannot solve a structure with a feature mismatch, rather it signals its existence, just as the pitch movement observed elsewhere (cf. Desouvrey 2000). As shown later on, adjunction (or incorporation) of the relative to another element, mostly its antecedent, makes it possible to resolve the issue.



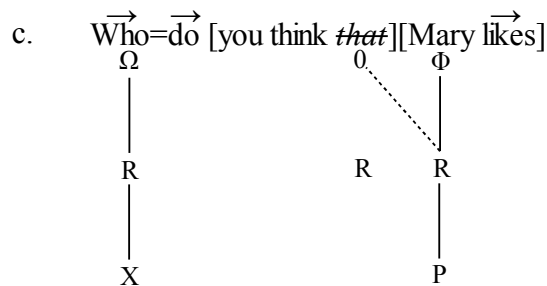
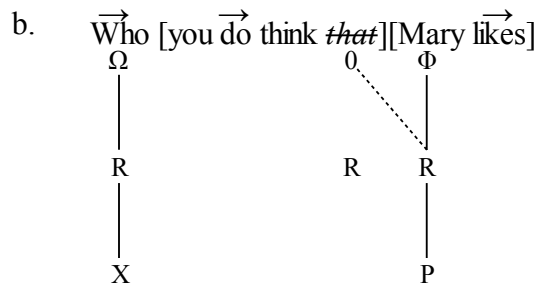
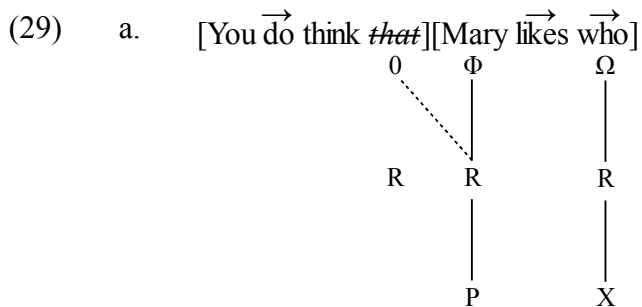


Let's now replace the object of the complement clause by an appropriate *wh*-interrogative, *who*, as seen in (28). In this structure, OCP is violated in the complement clause (the verb is A-specified, so is the operator), and therefore the operator should move to resolve the OCP problem. Just as in French, the operator can't end up to the left edge of its clause, since this move would alter the coreference between the subject and the matrix relative (NCL). We may note that this problem exists whether or not a *do*-support is used in the complement clause.



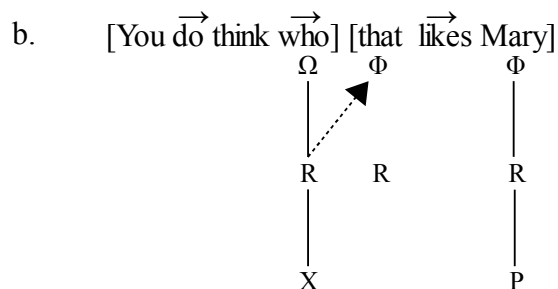
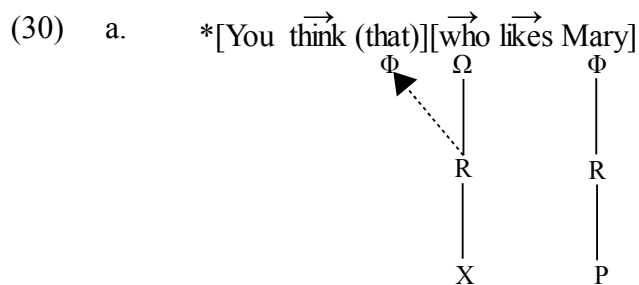
However, the operator can move directly to the front of the matrix clause, without the mediation of *do* in the complement clause, since superiority effect holds only in the clausal

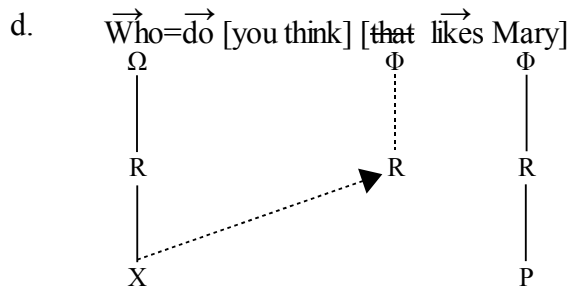
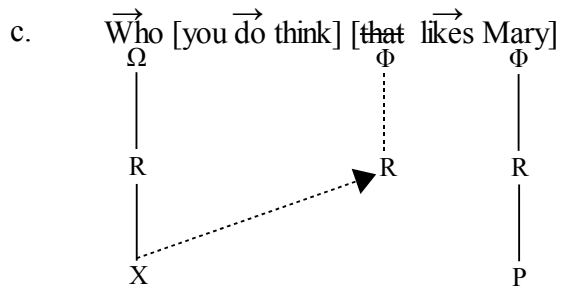
domain (cf. Desouvrey 2000). That is, operators are strictly ordered in this particular domain, and a lower operator can move (non-vacuously) past a higher one, as far as it does not end up in the same domain. In the matrix clause, the verb is a vector, and therefore it may not allow the lower operator to cross it over in its domain. Thus, just as in root clauses, *do* must appear in the matrix clause in order to allow fronting of the lower operator, as seen in (29). Auxiliary *do* is then incorporated to the operator by adjunction, canceling the superiority effect in the matrix clause, as seen in (29c), the desired output.



Now consider the case where the *wh*-operator is the subject of the complement clause, as seen in (30a). The omega-specified operator cannot assign its R-node to the phi-specified anaphor. This problem cannot be solved by deletion, since the feature tree remains intact,

only the phonetic segments being delinked from the skeleton. Moreover, the *wh*-vector comes to assign an R-node within the scope of another vector, namely the matrix verb. A long movement of the operator to the front of the matrix clause is also excluded, since it would come to antecede the relative to the left, a redirection that is banned. Therefore, extraction via relativization appears to be the only solution, as seen in (30b). From this alternative input, the operator and the anaphor clash, just as in the case above, which triggers the deletion of the relative. However, since anyway the operator must be fronted under OCP, and will have to target the R-node of the relative, the feature clash is avoided in the output, (30d). This is the most plausible account of the unfelicitousness of overt *that* in an embedded subject position (for certain speakers), a phenomenon known as the *that*-trace effect in the literature. However, since the feature mismatch in (30b) is only transitory, not permanent, other speakers ignore it altogether, hence does not even signal its existence by deletion.<sup>13</sup>





Notice, however, that the relative is mandatory, presumably not submitted to deletion, in contexts where an adverbial is present in the structure (cf. Culicover 1993, Browning 1996, Grimshaw 1997, etc). In my view, something forces both groups of speakers to agree. This thing is an instance of the constraint against ambiguity (cf. Desouvrey 2002). In effect, if the relative is not overt, it may not be clear whether the adverbial element is a member of the complement clause or instead a modifier of the matrix clause. This is illustrated in (31) (adapted from Grimshaw 1997, ex. (50)).

(31) Who did she swear *that* most of the time accepted this solution?

Before moving on, it should be noted that this analysis sheds a new light on the set of phenomena known as island effects. Traditionally accepted relative clauses, as opposed to matrix clauses (relative clauses in the present system), are islands, since they do not allow *wh*-movement. For instance, in (32a) the object in the relative clause, *Mary*, cannot be *wh*-extracted, as seen in (32b).

(32) a. The man [that likes Mary] came.

- b. \*Who=does the man [that likes *t*] come?

It turns out that (32b) is doomed, because there is nothing that the grammar can do to rescue it. If the *wh*-object of (32b) can't move out of the *that*-clause, one would expect the relativization strategy to circumvent this problem. However, since in the main clause there is no relative anaphor that would yield its argument slot to the operator, extraction by relativization is impossible; and, if applied, the argument structure of the main verb would be violated, creating a fatal flaw. Let's take a closer look to the derivation of (32b). From the corrected input (33a), all the usual operations take place, but the last one, namely adjunction of the relative clause to its head (33d), falls beyond the scope of  $\varepsilon$ , and therefore the structure is ruled out.

- (33) a. [the man<sub>i</sub> does come][that<sub>i</sub> likes who] ( $\varepsilon=1$ )  
 b. Who [the man<sub>i</sub> does come][that<sub>i</sub> likes] ( $\varepsilon=2$ )  
 c. Who=does [the man<sub>i</sub> come][that<sub>i</sub> likes] ( $\varepsilon=3$ )  
 d. \*Who=does the man=that likes] come? ( $\varepsilon=4$ )

As it happens, (33c) can't be an acceptable output either. Why is this the case? The fact is that adjunction (or incorporation) of the relative pronoun to its antecedent is the key operation in (restrictive) relative clauses. This has to be the case in order to be consistent with their semantic function, otherwise the interpretation would be compromised. (Recall that relative elements have a bare phi-node and therefore can refer to any element with phi-features.) In fact the fate of this structure would be the same if incorporation were to occur earlier in the derivation. Adjoining the relative clause to the head noun directly from the input, one would obtain a structure where the head noun and the relative clause are attached with an association line so that the last segment of the head is linked to the first segment of the relative pronoun, as seen in (34a). As a result, each clause comes to be in its own timing tier. Such a nonlinear structure is then conflated by a linearization convention, which cannot process a zigzag (cf. Desouvrey 2000, 2005).<sup>14</sup> Now, if the operator were to move, it could only end up to the left edge of its own tier, as seen in (34b). Obviously, such a

structure is not conform to the linearization convention, and therefore is ruled out. Although derivation (33) correctly predicts the ill-formedness of the structure, it does not crucially rule it out. In effect, this derivation is extended in order to avoid the superiority effect induced by English verbs. French verbs are not vectors, but the result is still the same. Therefore, one can conclude that in languages with relative pronouns, the island effect is due to the adjunction of the relative to its head.

- (34) a.     The man                             came.  
              |  
              that does like who
- b.     \*The man                            came.  
              |  
              who that does like

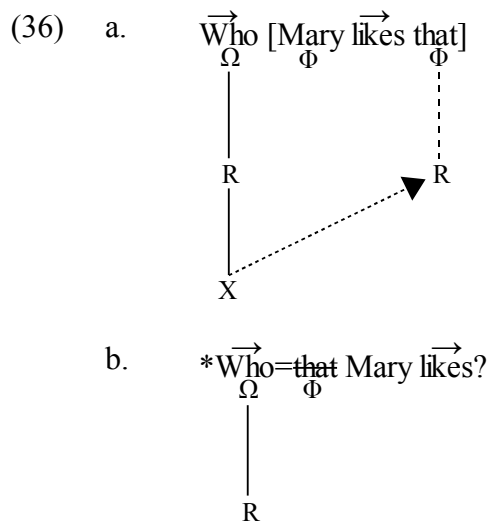
Furthermore, an important fact that can be tackled under this analysis is the difference between embedded clauses and root clauses with respect to *wh*-extraction. In root clauses, an alternative input with *do* is used in order to settle the conflict between OCP and SUPC. For instance, the derivation of *Who does Mary like?* must proceed from the alternative input (35a), not the natural input (35c), where for reason of scope the *wh*-vector cannot move past the vector lexical verb. From (35a), the *wh*-vector moves to the left edge of its clause and then the auxiliary incorporates it by adjunction, which cancels the scope effect, as seen in (35b) (see Desouvrey 2007).

- (35) a.     Mary  $\overrightarrow{\text{does}}$  like  $\overrightarrow{\text{who}}$ ?  
      b.      $\overrightarrow{\text{Who}} = \overrightarrow{\text{does}}$  Mary like?  
      c.     \*Mary  $\overrightarrow{\text{likes}}$   $\overrightarrow{\text{who}}$ ?

Applied to root clauses, the relativization strategy would yield a structure like (36a). The relative pronoun would move to the operator, but the resulting structure would not be any different from that of a derivation which would proceed from the natural input (35c),



since the phi-specified relative pronoun has to be deleted in order to signal a feature clash with the omega-specified operator. In other terms, if such a derivation were used, infants would have hard time to learn that lexical verbs are vectors, and this would have considerable consequences elsewhere in the grammar of English. Actually, derivation (36) yields a string vacuous operation, which is barred for the purpose of language acquisition. In other terms, the grammar must provide learners with clues to identify constraints and features (see Desouvrey 2000).<sup>15</sup>



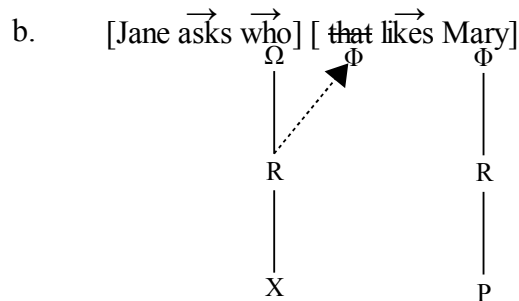
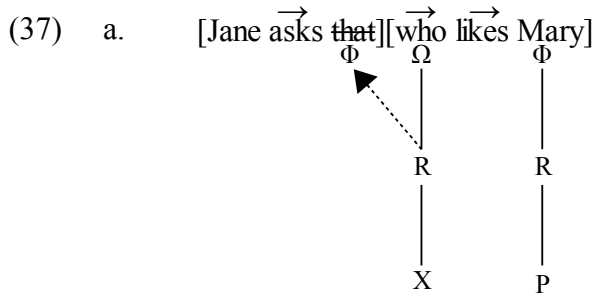
#### 4. Subordinate interrogatives

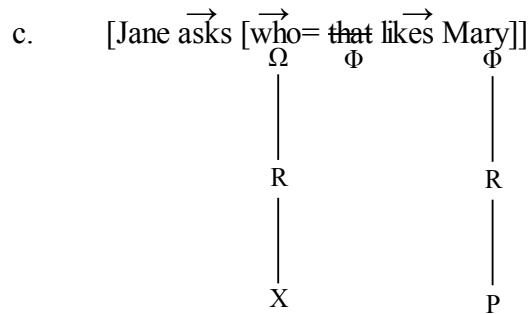
In subordinate interrogatives (SI), *wh*-operators must stay under the scope of the matrix verb, that is, they can never be fronted, which is arguably due to the semantic properties of this construction.

SI constructions are not significantly different from the structures with matrix verbs discussed above. I assume that syntactically SI verbs select an interrogative clause as complement, which has to be mediated by a relative pronoun, just like non-interrogative matrix verbs. The difference between both types of verbs is likely to be found in their LCS. Suppose that the interrogative property of SI verbs is rooted in their LCS, unlike non SI verbs which may only have an omega feature in their tree structure, as seen above in

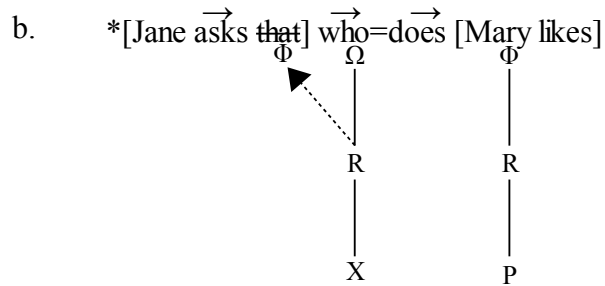
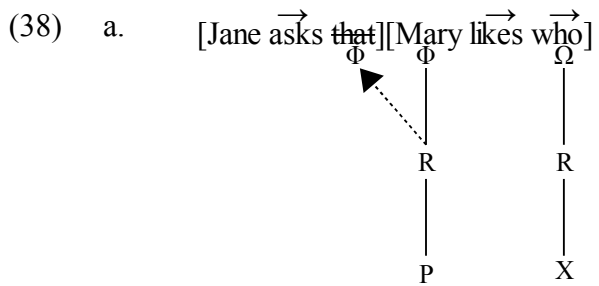
English. As a result, the interrogative feature of SI verbs cannot be bypassed by a morpho-syntactic operation like *do* plus a non-tensed form of a lexical verb. To put it differently, whatever the morphological shape of the SI verbs (infinitive or tensed), the interrogative property is always present, since the LCS of such verbs has a built-in omega feature. Therefore, an interrogative complement will always stay under the scope of the SI verb, consistent with the vector effect.

On this view, the natural input for a sentence like *Jane asks who likes Mary*, must be as shown in (37a), which is ruled out because of a feature mismatch and a vector effect, as discussed earlier. From the alternative input (37b), which is derived by the relativization strategy seen above, the relative is incorporated to the operator, and then deleted, as seen in (37c). Incorporation defuses the feature mismatch under the natural assumption that an element cannot spread a feature to itself.



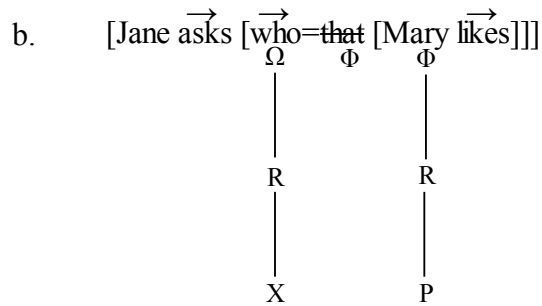
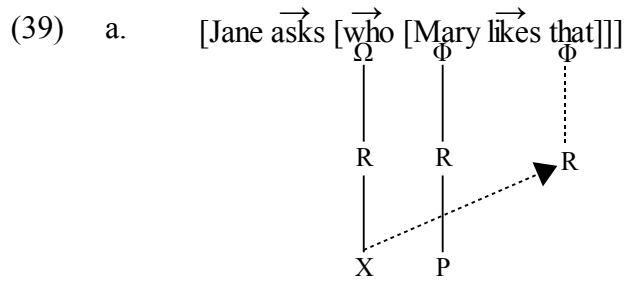


The object-extracted pendent of (37) is the stronger evidence in English grammar for the relativization strategy. If the natural input were correct in either of the structures discussed above, one would expect local movement of the operator in (38a) to be mediated by auxiliary *do*. The operator would move to the left edge of its clause, as seen in (38b), which is incorrect for the reasons discussed above, namely reference alternation (i.e., demotion of the subject as the antecedent of the relative), and rightward spreading in the domain of a higher vector (vector effect).

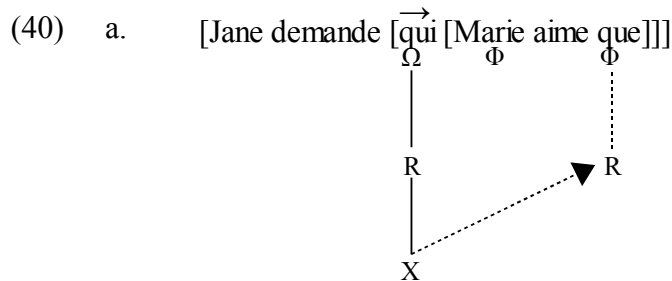


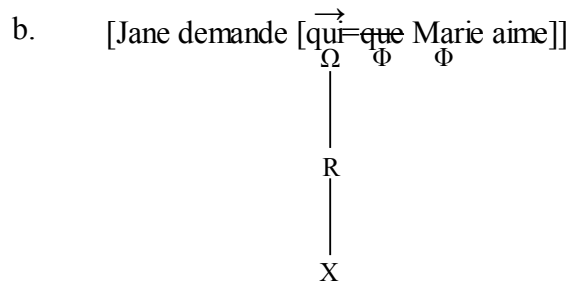
However, under the relativization strategy, the derivation proceeds from input (39a), where the relative pronoun moves to the operator, and then deleted to resolve the feature

mismatch, as seen in (39b), which is the desired output. It is clear that *do* may not appear in subordinate interrogatives, since there is no *wh*-movement in such structures and the relative pronoun is not a vector.<sup>16</sup>

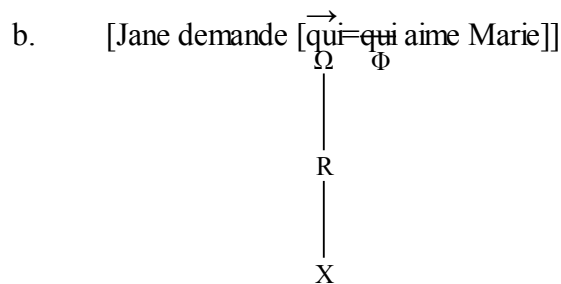
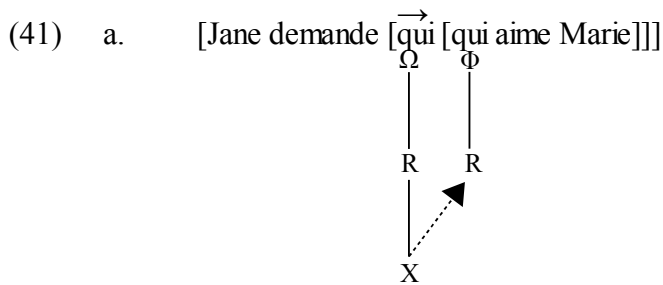


SI constructions are universal, and therefore the same analysis is expected in languages with relative elements. In (standard) French, just as in English, the relative is systematically deleted in order to avoid a feature mismatch. From the alternative input, (40a), the relative pronoun is adjoined to the operator, and is then deleted, as seen in (40b).





If the relative is the subject of its clause, incorporation still takes place, even though there is no feature mismatch with a pronoun, as seen in derivation (41). This is presumably due to the process that harmonizes the subject with the object extraction case.<sup>17</sup>



To complete this analysis, it remains to account for indirect interrogatives introduced with *whether* and *if* in English, and their French counterpart, *si*, as illustrated in (42). It is obvious that these interrogative elements introduce sentential complements, which correspond to a yes-no question in root clauses, as illustrated in (43). The relativization strategy cannot take place, since there is nothing to extract; such elements like *if* and *whether* that introduce this construction are obviously neither referential nor argumental.

- (42) a. He asks whether Paul came.  
 b. Il demande si Paul est venu.
- (43) a. Did Paul come?  
 b. Paul est-il venu? (formal French)

I take those elements, namely *whether* and *if*, to be interrogative adverbs modifying a clause. On this view, a sentential complement of a SI verb must be normally introduced by a matrix relative. Thus the input for (42a) is (44a). As discussed above, the relative is intended to refer to the complement clause, but its actual antecedent is the closest NP, namely the subject. From this input, the matrix relative adjoins to the adverb in order to be closer to its antecedent, as seen in (44b), which is the correct output. Once adjoined, the relative has to be deleted in order to signal a mismatch with the interrogative adverb, which is presumably omega-specified.

- (44) a. [He asks that] [whether [Paul came]]  
 b. [He asks [whether=~~that~~ [Paul came]]]

Similarly, in French the derivation of (42b) proceeds from input (45a), where the matrix relative adjoins to the interrogative adverb and then deleted, as usual, since there is a feature mismatch.

- (45) a. [Il demande que] [si [Paul est venu]]  
 b. [Il demande [si=~~que~~ [Paul est venu]]]

In colloquial French, a structure with a complex operator is used instead of (42b) and (43b), as seen in (46). Such structures, which distinguish the standard dialect from the colloquial ones, are discussed in the next section.

- (46) a. Il demande est-ce que Paul est venu. (Colloquial French)  
 b. Est-ce que Paul est venu?

### 5. The nature of complex *wh*-operators

In the tree structure in (14a), which represents French non-animate *wh*-operators, there is no phi-node at all. Therefore, the Caseless non-animate operator can never be used as subject, which must agree in phi-features with the selecting verb in French. In root clauses, a compound operator, *qu'est-ce qui*, or a variant *c'est quoi*, must be used in all dialects of French (see below), as seen in (47a,b). Notice, however, that operator *que* can be used as the subject of an embedded clause (47c). (I will return to this shortly.) On the other hand, in subordinate interrogatives standard French uses rather the compound *ce qui* 'what', as seen in (48a), unlike colloquial French which still uses the former, as seen in (48b).

- (47) a. *Qu'est-ce qui* intéresse Marie?  
What is this that interests Marie?
- b. \**Quoi*/\**que* intéresse Marie?
- c. *Que* crois-tu *qui* intéresse Marie?  
What do you believe interests Marie?
- (48) a. Jane demande *ce qui* intéresse Marie. (standard French)  
Jane asks this that interests Marie.
- b. Jane demande *qu'est-ce qui* intéresse Marie. (colloquial French)

With respect to (48a), it is important to find out the features of *ce* (or *c'* by phonology). This element is used as a neutral demonstrative pronoun to identify something or someone. Its use is much restricted than *cela/ça*, which have similar meaning. In fact, in modern French *ce* is apparently compatible only with copula *être* 'to be', as seen in (49). Under the present set of assumptions, if *ce* cannot be the subject of any verb but *être*, it must be the case that it is specified for (or lacks) some feature that only *être* possesses (or lacks).<sup>18</sup> Fortunately, such a feature can be tracked down.

- (49) a. *Ce*/\**ça* n'est pas important.  
This not is not important.

- b. Ça/\*ce n'a pas d'importance.  
This not has not importance.
- c. Ça/\*c' intéresse Marie.  
This interests Marie.
- d. Qui est-ce?  
Who is it?

It is clear that (49a-c) are not interrogative sentences, unlike (49d). Now a property of the copula verb in French is that it requires agreement between its subject and its adjective complement, as illustrated in (50). More generally, two elements in a copular relationship must agree in phi-features or in omega features. Thus, if *ce* was a scalar, it could not be used when the copula introduces a vector, (49d), and it would not be ruled out as subject of a lexical verb, (49b,c); if it were a vector, it could not appear with a predicative scalar, (49a). Both cases are possible, however. Therefore, I am led to propose that this element is neither a vector nor a scalar. In other terms, it is neutral with respect to both interrogative and non-interrogative elements. Furthermore, I take *ce* to be a fully referential element, that is, an element with an R-node expanding to a R-feature, which I appropriately refer to as N (from neutral).<sup>19</sup>

- (50) a. Cette fille est heureuse.  
This girl.FEM. is happy.FEM.
- b. \*Cette fille est heureux.  
This girl.FEM. is happy.MASC.

With this discussion in mind, one can posit for (48a) the structure in (51a), where the matrix relative disappears as usual, since it refers only to NP *ce* that heads the relative clause. It turns out that the relative pronoun is used as a buffer between the verb and the intended subject, *ce*.<sup>20</sup> This structure yields (51b) after incorporation of the relative to its head. Interestingly, deletion of the relative need not take place, since there is no feature mismatch with the head.



- (51) a. [Jane demande [<sub>0</sub>ce [<sub>Φ</sub>qui intéresse Marie]]
- |            |
- R            R
- |            /
- N
- b. [Jane demande [<sub>0</sub>ce=<sub>Φ</sub>qui intéresse Marie]]
- |
- R
- |
- N

When the object is questioned, this construction must still be used, as seen in (52). Notice (52) is possible despite the fact that there is an appropriate *wh*-element, *que*, which is used in root clauses (cf. *Que veut Marie?*). This is due to the harmonization effect that takes place in the paradigm (see Desouvrey 2000).

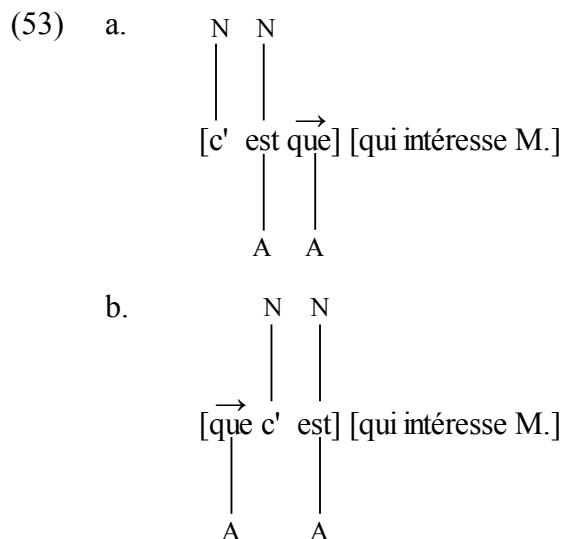
- (52) [Jane demande [<sub>0</sub>ce=<sub>Φ</sub>que Marie veut *t*]]
- |
- R
- |
- N

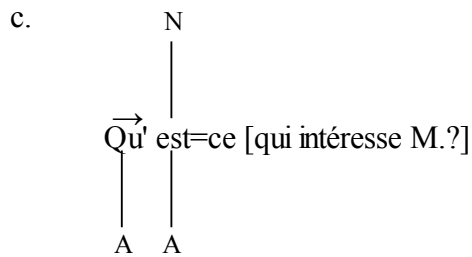
### 5.1 Deriving complex operators

As seen above, French has no *wh*-element that can refer to a non-animate in subject position. However, I argue that the grammar uses a trick to overcome this lexical gap. First of all, recall that a *wh*-element referring to human cannot stay in the embedded subject position of complement clauses, a condition due to the feature clash between the operator and the matrix relative. The relativization strategy is used to extract the operator while

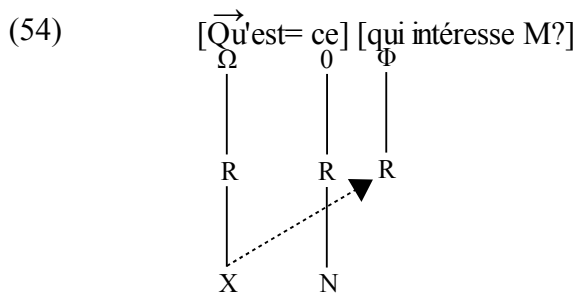
avoiding a vector effect with the matrix relative, which is then deleted. Now this strategy is so efficient that it even applies to compute the faulty accusative operator as the subject of the complement clause, compare (47b) and (47c) above. Clearly, an operator is missing in the paradigm, and it turns out that the accusative operator can fill that gap in embedded clauses (47c), but not in root clauses (47b). Therefore, to correct the latter, what is needed is a matrix clause, a meaningless one that will not introduce further information in the structure. Such a dummy matrix clause is *c'est que* 'it/this is that'.

Indeed, I show that the structure of complex operators of the type seen in (47a) consists of dummy matrix clause *c'est que* and a relativized complement clause. Since operator *que* (or *quoi*) can't be used as subject, a relative pronoun headed by these operators appear in subject position, as a buffer. Since the matrix relative comes to refer only to the operator, not to the clause headed by the operator, it disappears, just as in the cases seen earlier. Thus, the derivation of (47) has to proceed, after reanalysis, from input (53a).<sup>21</sup> The OCP cycle takes place in the matrix clause, fronting the operator, (53b), then moving the subject to the right side of the copula verb, as seen in (53c), which is the desired output. The relative clause is not affected by these processes, since the subject relative pronoun is not specified for Case.



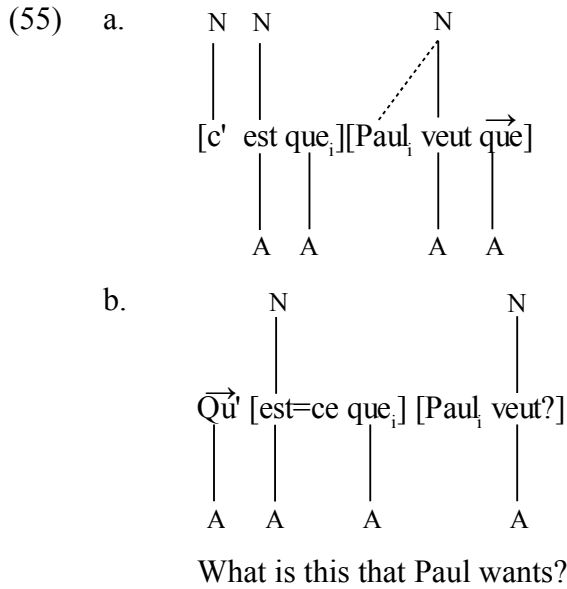


From the referential plane, there is no clash between *qui*, which is always a pronoun, and the operator, since only the R-projection matters with pronouns, as seen in (54). We may note that the relative clause is prevented from adjoining to its *wh*-head, for at the end of the OCP cycle, the derivation reaches its maximal extension, and is spelled out.

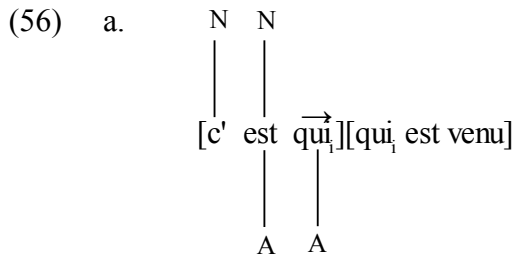


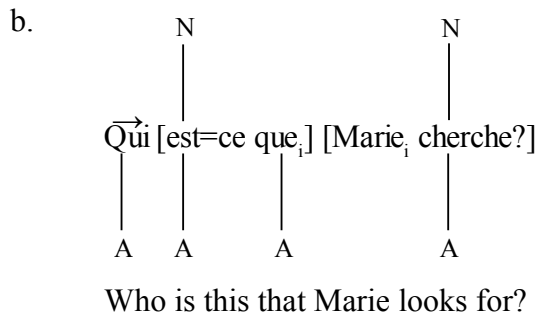
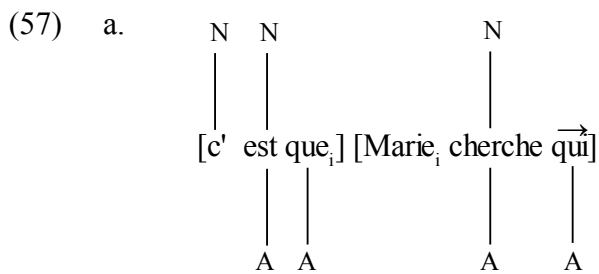
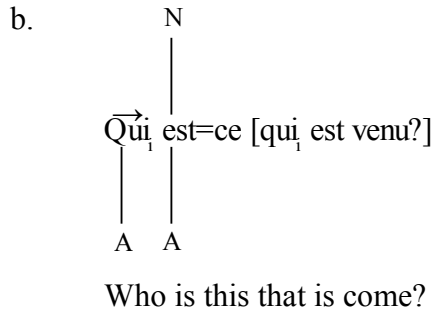
This construction should normally be used only to fill in the gap in the paradigm. However, by virtue of the harmonization process, it comes to plague the whole paradigm of *wh*-operators in French, supplanting simple operators in ordinary speeches, so that in colloquial French interrogative sentences are preferably embedded under the dummy matrix clause.

Thus, to question a non-animate object, operator *que* can normally be used, as discussed in Desouvrey (2007). However, in colloquial French a complex operator is preferred. Structure (55a) is an input where a complex operator is the object of the embedded clause. The relativization process may not take place with object operators; so under OCP the operator moves directly to the front of the dummy matrix clause, where it triggers movement of the subject (OCP'), yielding the desired result, (55b).

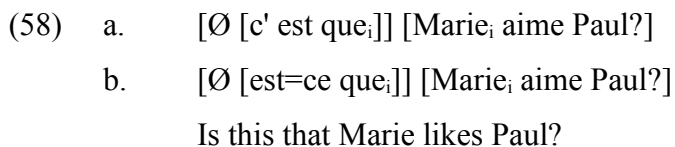


With an animate operator, use of the relativization strategy is an effect of the harmonization process, since there is no mismatch with the verb and the operator (cf. *Qui est venu?*). Thus, from input (56), which is derived by the relativization strategy, the operator is reanalyzed as the object of the matrix verb, while relative *qui*, which is compatible with both animate and non-animate, appears as the subject of the embedded clause. In (57), however, the object operator moves directly to the front of the dummy matrix clause, since it has no reference in the matrix clause.<sup>22</sup>





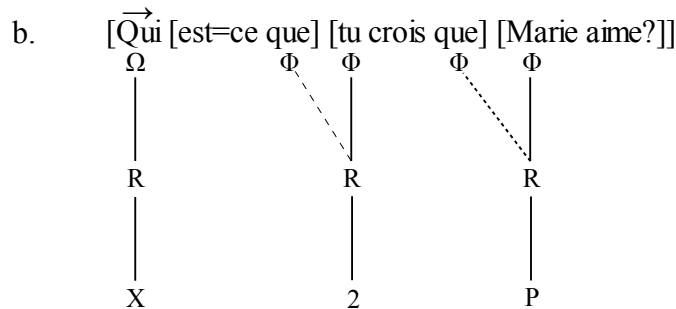
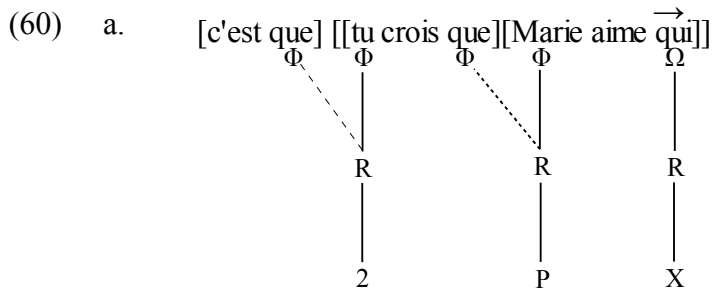
This type of construction is even used in yes-no questions, as is (58a). In such a construction, the matrix relative is not dismissed, since nothing is extracted. The only operator in the structure is the null adverbial operator (see Desouvrey 2007), which triggers movement of the neutral subject to the right-edge of the copula verb, as shown in (58b).



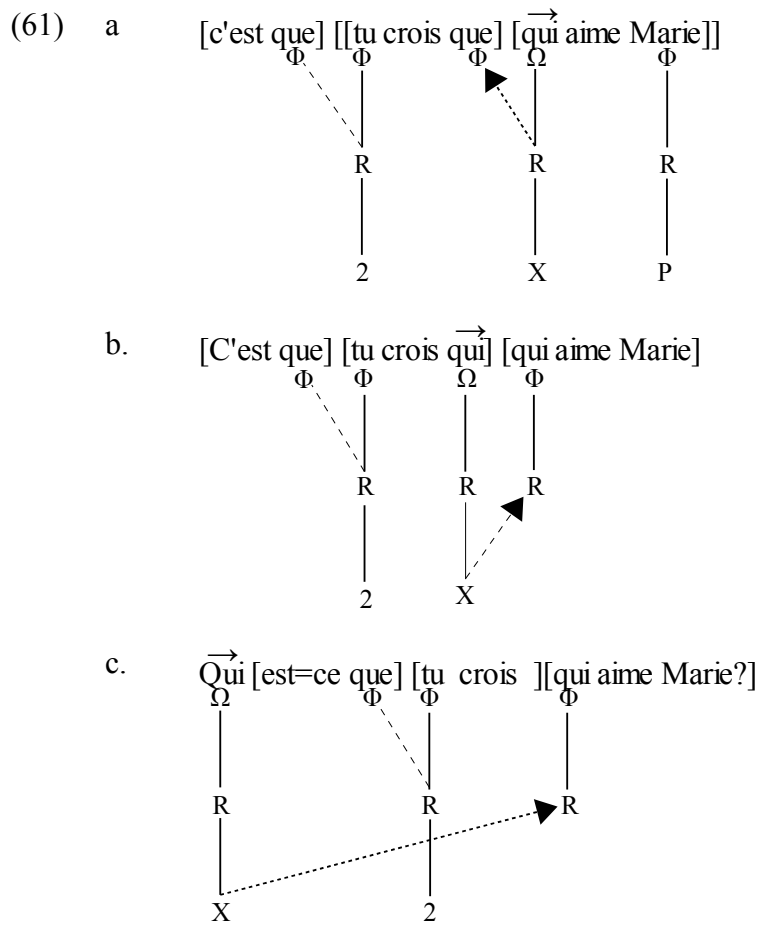
Complex *wh*-operators are even used in structures with matrix clauses. Consider the derivation of the sentences in (59). The input for (59a) is (60a), where the complex

structure is embedded under a dummy matrix clause. The object operator moves to the front of the dummy matrix clause, triggering subject inversion, as seen in (60b). Coreference between the matrix relatives and their antecedents remain intact.

- (59) a. *Qui est-ce que tu crois que Marie aime?*  
 Who is this that you think that Marie likes?  
 b. *Qui est-ce que tu crois qui aime Marie?*  
 Who is this that you think that likes Marie?



As for (59b), its natural input is (61a). Being in subject position, the operator cannot move, as seen above. By the relativization strategy, it comes to head a *qui*-relative clause, and at the same time, the matrix relative, which no longer refers to a clause, disappears, yielding the alternative input (61b). Now, the operator takes the argument slot of the matrix relative, i.e., it becomes the object of the matrix clause, and then moves to the front of the dummy matrix clause, yielding the desired result, as seen in (61c).

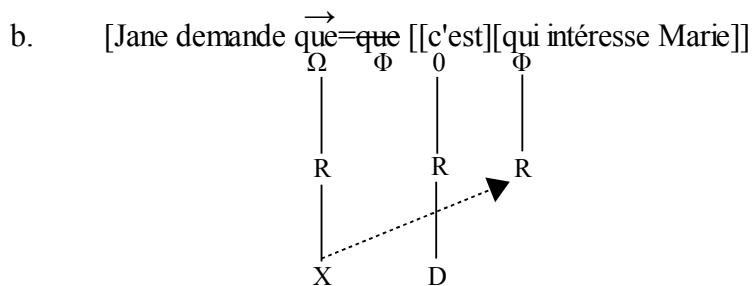
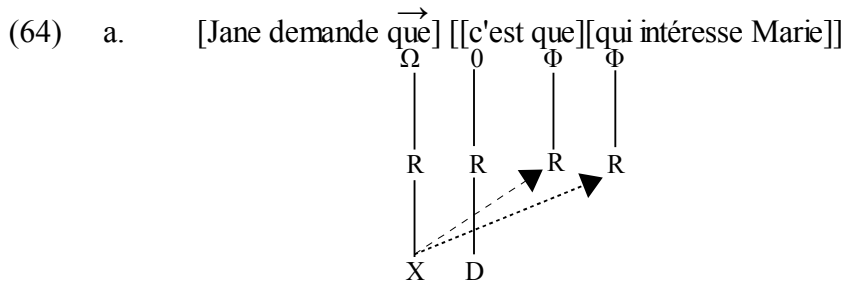
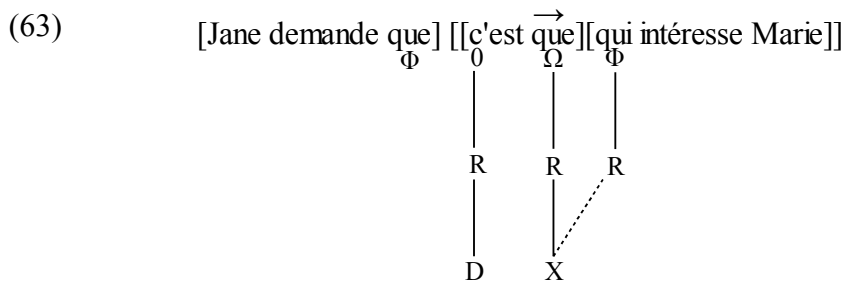


In SI constructions, colloquial French embeds a structure with a dummy matrix clause, as seen in (62a). In this input, the actual antecedent of the higher matrix relative is the neutral element *ce*, which is the subject of a dummy matrix clause. The matrix relative adjoins to the null omega adverb in order to get closer to its antecedent, and is then deleted to avoid the mismatch (the null adverb is omega-specified), while at the same time the neutral pronoun adjoins to the verb under OCP', yielding (62b), the desired result.

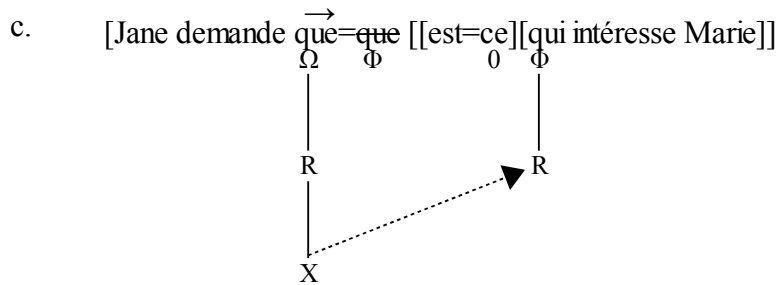
- (62) a. [Il demande que] [ $\emptyset$  [c'est que][Paul est venu]]  
 b. [Il demande [ $\emptyset$ =~~que~~ [est=ce que][Paul est venu]]]

Embedded *wh*-questions proceed the same way. Consider the case where a subject *wh*-question, like (54), is embedded under an SI verb. The relativization strategy must

apply twice, first as a buffer to the operator, as seen in (63), and then as extractor of the operator, since OCP is violated in the dummy matrix clause and direct movement would change the antecedent of the matrix relative. As a result the matrix relative disappears at its turn. This is shown in (64a). The relative in the dummy clause adjoins to the operator and is then deleted, (64b), and further it triggers subject inversion, as seen in (64c), the desired result. A similar analysis would take place with an embedded object *wh*-question.





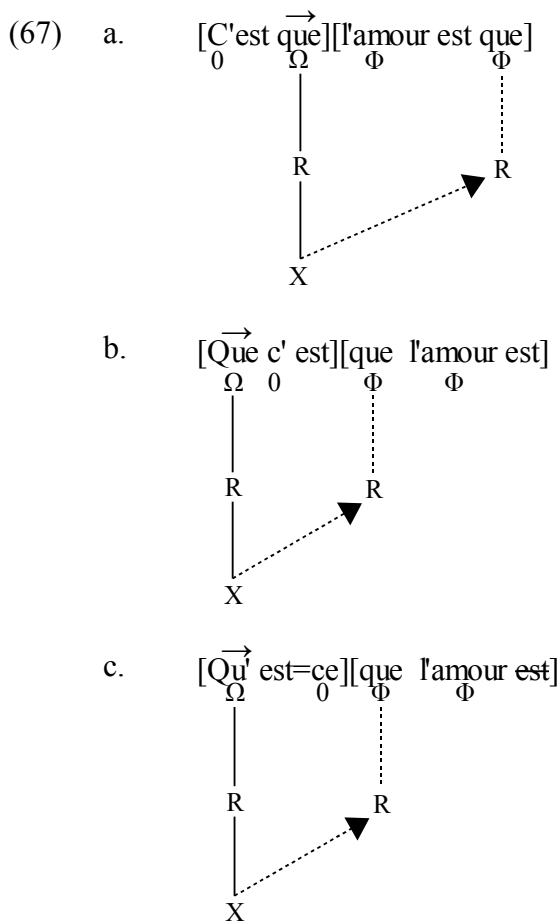


Let us turn now to another type of interrogative in French that crucially uses a dummy matrix clause. A copular NP can only be questioned with a 'complex' operator in standard French, as illustrated in (65).

- (65) a. Qu'est-ce que l'amour?  
 What is it that the love?
- b. \*Qu'est l'amour?  
 What is the love?

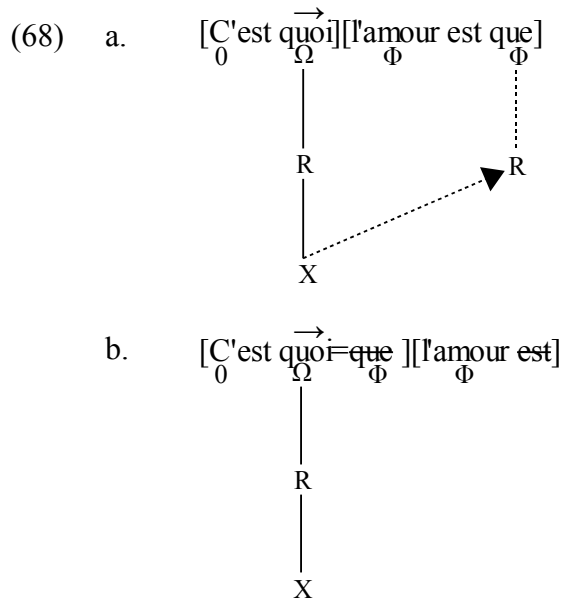
One can posit that this NP is construed as originating from an equative copula construction, such as *l'amour est x*, where *x* stands for some NP. However, *x* cannot be replaced by the relevant operator, *que*, since the latter has no phi-features. Indeed, such inputs like (66a) or (66b) are incorrect, since the scalar NP, *l'amour*, can agree neither with operator *que* nor with operator *quoi*. The grammar can avoid the mismatch by embedding the structure under a dummy matrix clause, just as seen in (66b), while a relative pronoun is buffering for the *wh*-operator. As a result, the matrix relative disappears and therefore the derivation proceeds from (67a). Both the operator and the relative move around their clauses under OCP, (67b). Furthermore, an OCP' effect triggers subject inversion in the dummy matrix clause, while the copula is deleted in the complement clause, presumably because it becomes useless after being stranded by the relative pronoun, as seen in (67c).<sup>23</sup>

- (66) a. \*L'amour est  $\overrightarrow{\text{que}}$ / $\overrightarrow{\text{quoi}}$ .
- b. \*[C'est que][l'amour est  $\overrightarrow{\text{que}}$ / $\overrightarrow{\text{quoi}}$ ]



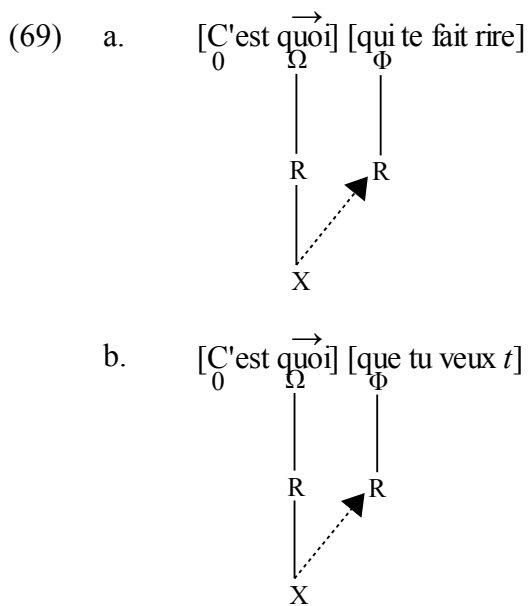
In the previous example, the dummy clause is used in correlation with the relativization of the embedded operator in order to avoid a feature mismatch in the complement clause. Now, seemingly, (67c) could be obtained from direct movement of operator *que* in (66b) to the front of the matrix clause. Such a movement would be triggered by OCP. If this derivation were correct, and was simply a matter of resolving an OCP violation in the complement clause, the Caseless operator, which does not violate OCP, would produce a well-formed sentence in situ. This is not true however, cf. *\*L'amour est quoi?* Moreover, if the relativization strategy is mandatory to derive (67c), it is expected to be used with the Caseless operator as well, given that in most contexts in French Caseless operators are used symmetrically to the Case-specified allomorph. This appears to be correct. Indeed, operator *quoi* must be relativized, so that the derivation proceeds from

input (68a). Since *quoi* is unspecified for accusative Case, there is no OCP cycle in the matrix clause. As a result, the relative can reach it, adjoining to the in situ operator, as seen in (68b), which yields a feature mismatch. Therefore, the relative is deleted, as is the copula, yielding a well-formed sentence.



It appears that the dummy matrix clause and the relativization strategy are used in copula constructions in order to avoid a feature mismatch (vector vs. scalar), just as they are used to supply the lack of a non-animate *wh*-subject in the paradigm. In any event, the operator is the head of a relative pronoun, and the matrix relative disappears, yielding an apparent substitution at the input, as seen above. The OCP cycle only occurs in the matrix clause if the operator is Case-specified. Thus in (69a) the Caseless operator stays in situ, in the place of the matrix relative it supersedes. We may recall that in this example, the subject relative is a pronoun, and therefore it need not be incorporated to the string-adjacent operator, as is usually the case. When the relative is the object of the clause, as in (69b), it harmonizes with the subject case, which means it does not incorporate to the operator, thus avoiding deletion. It is important to notice that (69b) must harmonize with (69a), not the other way. That is, one cannot derive (69b) with the relative adjoined and deleted, and then

harmonize (69a) with (69b), so that the relative is deleted in both structures. In effect, the whole process of embedding and relativization is used in the first place to accommodate (69a), since as we know, there is no subject non-animate operator in French. Moreover, there is an alternative structure for (69b), *tu veux quoi?*, not for (69a), *\*quoi tu veux?* In other words, construction (69b) exists because of the need to derive (69a), and therefore the harmonization goes in the sense of (69a). The harmonization process makes it possible to maintain an opposition between subject and overt object relative in French; if it does not exist, the object relative would have to be deleted for the same reason as in (68b), which does not have a subject pendent.



To wrap up, consider the interrogative sentence in (70), categorized as familiar or popular by normative grammarians. Its peculiarity comes from the reduplication of the dummy clause. Recall that originally the dummy clause appears concomitantly with the relativization strategy to compute the faulty *que* as subject. Then the harmonization process overgeneralizes the use of the dummy matrix clause in colloquial French, so that every interrogative sentences come to be embedded. As a result, the original purpose of the dummy clause is completely masked, and nothing prevents speakers from seeing two

distinct processes: (a) every interrogatives must be embedded, and (b) relativization under embedding must be used to avoid any feature mismatch. Thus, the derivation of (70) starts with two occurrences of the dummy matrix clause; one is for the general embedding pattern, and the second one is for especially addressing the mismatch between scalar *ça* and operator *que*. Then the matrix relative in the second dummy clause is dismissed by the operator heading the relative pronoun, as usual. The operator moves to the front of the structure, while the relative pronoun it is the antecedent of moves to the front of the lower clause, triggering deletion of the now useless copula, as seen in (71b), the desired result.

(70) Qu'est-ce que c'est que ça?  
 What is it that it is that that?  
 'What's that?'

(71) a. [c'est que<sub>i</sub>] [c'est  $\overrightarrow{\text{que}}_j$ ] [ça est que<sub>j</sub>]  
 b.  $\overrightarrow{\text{Qu}}'_j$  [est=ce que<sub>i</sub>] [c'est] [que<sub>j</sub> ça est]

This structure is similar to (67c), where a single dummy matrix clause is doing the same job. Indeed, use of two distinct matrix clauses in (67c) is correct, but it belongs to the same register as (70). Thus:

(72) a. [c'est que<sub>i</sub>] [c'est  $\overrightarrow{\text{que}}_j$ ] [l'amour est que<sub>j</sub>]  
 b.  $\overrightarrow{\text{Qu}}'_j$  [est=ce que<sub>i</sub>] [c'est] [que<sub>j</sub> l'amour est]

Why is this reduplication limited to sentences with copula verbs? The copula verb is unique, perhaps belonging to a different paradigm from lexical verbs, in that it is the only case where the relativization strategy is required for a complement, hence the relativization by *que*. With lexical verbs, however, a complement never undergoes relativization; they move directly to the front of the matrix clause.

## 6. Conclusion

It is suggested that matrix verbs select as complement a relative pronoun/anaphor whose antecedent is intended to be the complement clause, but is actually the subject of the latter. Constraint violation triggered by embedded *wh*-elements, namely vector effects and feature mismatch, are resolved, in French and English, through an alternative input in which the *wh*-operator becomes the head of its clause by relativization. Under this condition, only the *wh*-operator is the complement of the matrix verb, which renders useless the matrix relative (complementizer). The consequence is that the derivation proceeds from an input in which the operator and the matrix relative are seemingly swapped; the embedded *wh*-operator becomes a member of the matrix clause where it is fronted, just as in root clauses via the mediation of *do*. In subordinate interrogatives the matrix verb being intrinsically interrogative its scope cannot be bypassed by a *do*-support and therefore *wh*-elements cannot be fronted.

Under the present analysis the difference between French and English with respect to subordinate interrogatives is reduced to a difference in morpheme features. English has one relative morpheme for complementation, Caseless *that*, while French has two different morphemes, *qui* and *que*, one being the Case allomorph of the other, just like *he* and *him* in English. Accusative-specified *que* is normally used in complementation, and *qui* is used elsewhere. In both languages, the matrix relative can be used either as a pronoun or an anaphor. In the latter case, their antecedent has to be a scalar. If the antecedent is a vector, incorporation and/or deletion occur to avoid a feature clash.

Although the analysis presented above may shock many linguists due to decades of analyzing *that*, and *que* and *qui* as complementizers and constituents of the embedded clause, it is grounded in that it better describes and accounts for the facts than any variant of the complementizer-based analysis. Moreover, this analysis can be improved or challenged only by using the same tools, namely features and constraints acting upon them. After all, features are the smallest grammatical unit.

## References

- Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge: MIT Press.
- Bošković, Ž., and H. Lasnik (2003). On the Distribution of Null Complementizers. *Linguistic Inquiry* 34, 527-546.
- Browning, M.A. (1996). CP Recursion and *That-t* Effects. *Linguistic Inquiry* 27, 237-255.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge MA: MIT Press.
- Cole, P., and G. Hermon (2000). Partial Wh-Movement: Evidence from Malay. In Lutz, U., et al. (eds), *Wh-Scope Marking*. Amsterdam: John Benjamins.
- Culicover, P. (1993). Evidence Against ECP Account of the *That-Trace* Effect. *Linguistic Inquiry* 24, 557-561.
- Desouvrey, L.-H. (1996). Case Tier, Clause Structure and the Nature of the Complementizer Trace Effect. Ms.
- Desouvrey, L.-H. (1997). Relativization in French without Complementizer. Proceedings of CLA 1996. Calgary Working Papers in Linguistics.
- Desouvrey, L.-H. (2000). Romance Clitics and Feature Asymmetry: An Autosegmental Based-Approach. Doctoral dissertation, UQAM.
- Desouvrey, L.-H. (2002). Adverbs, Negation and OCP Effects. [www.semanticsarchive.net](http://www.semanticsarchive.net)
- Desouvrey, L.-H. (2003). The Proper Treatment of Coreference Relations. [www.semanticsarchive.net](http://www.semanticsarchive.net).
- Desouvrey, L.-H. (2005). Romance Clitic Clusters: The Case Connection. In Heggie, L. and F. Ordóñez (eds), *Clitic and Affix Combinations*. Amsterdam: John Benjamins.
- Desouvrey, L.-H. (2006). Underspecification and Long-Distance Antecedent: The Case of Chinese Ziji. [www.semanticsarchive.net](http://www.semanticsarchive.net)
- Desouvrey, L.-H. (2007). Wh-Interrogatives: The OCP Cycle.
- Grimshaw, J. (1997). Projection, Head and Optimality, *Linguistic Inquiry* 28, 373-422.
- Grimshaw, J. (2006). Location Specific Constraints in Matrix and Subordinate Clauses. Rutgers Optimality Archive.
- Prince, A. and P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. ROA. Rutgers University.
- Rubach, J. (2000). Positional Neutralization and the Expression of Contrast: A DOT Analysis. *Linguistic Inquiry* 31, 271-317.

## Endnotes

<sup>1</sup> This paper builds on Desouvrey (2007), and therefore the latter should be read first. As well, the reader may need to take a look at Desouvrey (2003).

<sup>2</sup> The difference between *que* and *qui* resides in their Case specification. It appears that *qui* is underspecified for Case, while *que* is accusative specified, hence its limited use as direct object (see Desouvrey 1997 and below).

<sup>3</sup> From this perspective, the statement expressed by the independent clause is restricted (or weakened) by the matrix clause.

<sup>4</sup> First and second persons pronouns are always known to the speaker, and therefore I assume that they have an R-feature, conveniently represented by a digit for the relevant person.

<sup>5</sup> The trees in (14) are simplified and it must be seen as incomplete, pending further research on feature dependency. Nevertheless, our results are grounded on real terminal features, descriptively adequate, and therefore are less likely to be ruled out by subsequent knowledge of tree organization. The proposed features for verbs and operators are consistent with the constraints on merge (cf. Desouvrey 2000, 2007). Basically in the building process of a syntactic structure, two elements A and B must have the same feature, or at least the argument element must be underspecified for the feature of the selecting head.

<sup>6</sup> This may slightly vary from language to language, just as feature inventory and feature dependency (Desouvrey 2006).

<sup>7</sup> In Desouvrey (2007), it is suggested that the key element in French *wh*-paradigm is *que* [A] and its Caseless allomorph *quoi*, both being used as non-animate arguments. The impossibility of *quoi* as a non-animate subject is assumed to follow from its specification for the grammatical relation of object, which the K-node depends on. This view must be slightly revised, however. It turns out that Caseless *quoi*, just like its allomorph *que*, lacks a phi-feature node, and therefore cannot agree with the tensed verb. The tree structure for *que/quoi* is (14a). Supporting facts and the strategy to circumvent these limitations are discussed below in the text. As for the animate counterpart of *quoi*, Caseless *qui*, it has a phi-node, (14b), and it can merge with the verb as a subject. In English (see below), the master key, so to speak, is *whom* [A] and *who*, which is Caseless. In colloquial English, *whom* is eroded into *who*, which yields an opposition between *who* [A] and *who* [ ] (no Case), a pair that rules the whole paradigm. Thus, non-animate operators follow this pattern, so that *what* [A] comes with a Caseless allomorph that is used as subject. English *wh* can all be represented with a single tree, (14b).

<sup>8</sup> A structure like (20b) is possible in certain languages, including German and Malay (see endnote 13).

<sup>9</sup> To clarify further the alternative input, let V be a matrix verb, N a NP, and S a sentence. By some constraint C, structure (ia) is acceptable, while (ib) is not. However, C can be satisfied in a structure such as (ic), where A is a referring element whose antecedent is intendedly S, but actually is some N in S.

- (i)
- a. [V N]
  - b. \*[V S]
  - c. [V A<sub>i</sub>] [S<sub>i</sub>]

Now, as it happens N<sub>i</sub> can be extracted from S such as one obtains (iia), where S is now a subcomponent of N. Therefore, (iia) is equivalent to (ia), as seen in (iib). By reanalysis, (iib) is *salva veritate* (iic). This is the logic behind input (21a).

- (ii)
- a. [V A<sub>i</sub>] [N<sub>ij</sub> [S<sub>j</sub>]]
  - b. V [N<sub>j</sub> [S<sub>j</sub>]]
  - c. [V N<sub>j</sub>] [S<sub>j</sub>]

<sup>10</sup> In other words, in a paradigm or any subpart of the grammar where at least two syntactic elements, say A and B, are symmetric, if an operation affects A, there will be a trend towards affecting B as well. Using this process, the grammar is dispensed with implicational constraints, as there is no reason to include contingencies like the nature of A (for instance a subject or not) in the constraint. A full study of the effects of this constraint will be tackled in future research.

<sup>11</sup> It seems that the contexts where *that* deletion is impossible can be explained as follows. Since there is no Case opposition as in French, *that* deletion may allow to avoid ambiguity in the course of processing the sentence. In effect, it gives rise to a clear opposition between subject morpheme and object morpheme, which would not exist otherwise. If this view is correct, deletion must be limited to the subset of contexts where there is such an opposition. That is, it may not take place in constructions which have no subject pendent, as is for instance the case where the relative is a complement to an



## Endnotes

NP, as in *the fact that* (see endnote 12).

<sup>12</sup> The structure of relative clause, discussed above with French data, is universal and is presumably used in languages with relative pronouns. A possible objection for this analysis may be motivated by the sentences in (i) (Bošković and Lasnik, 2003), where apparently the complement clause is extraposed with the 'complementizer,' suggesting the latter is not a constituent of the matrix clause. This type of structures does not rule out the relative clause analysis. I claim that sentence (ib) rather derives from (ii), where the relative is a complement of an NP, *the fact*. In other words, the relative is used for the complementation of *he liked linguistics* to NP *the fact*. This view is supported by the fact that this construction is ungrammatical if a subject is present, like *it*.

- (i) a. It was widely believed that he liked linguistics.  
b. That he liked linguistics (\*it) was widely believed.
- (ii) [[The fact that] [he liked linguistics]] was widely believed.

<sup>13</sup> Notice that the relative element is an anaphor in the input, but a pronoun in the output. Unlike the case discussed in (20b), movement gives prominence to the operator (the spreader), not to the anaphor, the receiver. While the operator can move and then targets the R-node of the relative, the latter cannot move on its own in order to force the operator to treat it as a pronoun, given the assumption that the operator may not allow another element to have prominence at its expense.

Furthermore, in the present system if in a language the matrix relative is only a pronoun (unlike English and French), it can move under OCP, since a pronoun need not be adjacent to its antecedent. As a result, *wh*-operators would not move to the front of the matrix clause. As far as I can tell, German and Malay are suchlike languages (see Cole and Hermon, 2000, and references therein for some data). In fact, German has a relative element, *dass*, similar to English *that*, and a distinct pronoun, *was*, perhaps similar to French *ce* (see below) but specified for accusative, which are analyzed respectively as a complementizer and a scope marker in the literature. When *dass* is used, fronting is mandatory, presumably under an analysis similar to that discussed in the text. However, if *was* is the matrix relative, it moves to the front of the matrix clause under OCP, while the *wh*-operator moves to the left edge of the embedded clause, where at a distance it assigns an R-node to the fronted relative pronoun, just as in (20b).

<sup>14</sup> The linearization convention runs like this. From left to right and from the main tier, embed the secondary tier without zigzag, in a continuous flow up to the last morpheme. Notice that adjunction takes place at the skeletal level, the x-slot hosting the segment 'th' is attached to the x-slot of 'n' (see Desouvrey 2000).

<sup>15</sup> Haitian Creole apparently uses the relativization strategy in root clauses to extract *wh*-operators. Interestingly, the subject relative pronoun is overt while the object one is abstract, and therefore the string-vacuous structure is avoided, as seen in (i). Certain French dialects display this type of constructions as well (see endnote 22).

- (i) a. Kies [ki reme Marie?]  
who that love Marie?
- b. Kies [Ø Marie reme?]  
Who -- Marie love?

<sup>16</sup> Grimshaw (2006) attempts a new account of the absence of *do* in indirect interrogatives by proposing that constraints are differently ranked in matrix and subordinate clauses.

Moreover, relying on work by McCloskey, Grimshaw cites some dialects of English where inversion is allowed in SI. From the present perspective, it is difficult to discuss isolated examples, without a thorough knowledge of feature and morpheme distribution in the paradigm. However, at first glance, it seems that, all other things being equal, this dialect deletes the relative prior to movement. For instance, sentence *I wonder what is he like at all* (Grimshaw's (30c), referring to McCloskey (2006)) would be derived as follows. From the derived input (ia), the auxiliary is attracted to the operator by OCP, while the relative is perhaps simply deleted in situ to prevent a feature clash (ib). Speakers of this dialect may see the process of movement and then always deletion of the relative as a string vacuous and ignore it altogether, which leaves room for the the auxiliary to be incorporated to the operator.

- (i) a. I wonder [  $\overrightarrow{\text{what}}$  [  $\overrightarrow{\text{he is like that at all}}$  ] ]  
b. I wonder [  $\overrightarrow{\text{what}} = \overrightarrow{\text{is}}$  he like -- at all ]

<sup>17</sup> The harmonization process follows from the assumption that the grammar does not include complex constraints of the implicational type. So in a closed paradigm, a structural change A that affects one element is likely to affect other elements unless another existing constraint is at stake and prevents A from occurring. It makes it possible to avoid a movement rule like the following: "Incorporate the relative if it is the object of its clause." The harmonization process may

## Endnotes

not appear instantly at the rise of a feature that triggers some operation; it may take a long period of time in the history of the language to be completed, and some speakers may even reject it. It causes French to have two interrogative systems (see below in the text).

<sup>18</sup> In effect, two elements, say A and B, which can have features F, T or 0-feature, agree and can merge if both are specified for the same feature, either F or T. If A is specified for F, while B is specified for T, there is a fatal mismatch and they cannot merge. If A is specified while the other is underspecified, there is no feature clash, and the resulting structure can be more or less acceptable, or can even be completely ruled out by a blocking effect (cf. Aronoff 1976). Usually underspecified elements, if not ruled out, are spelled out with a pitch movement, as observed in Desouvrey (2000).

<sup>19</sup> Neutral *ce* appears to be a clitic when used as a subject, which suggests that it is specified for nominative Case in this context (cf. Desouvrey 2000). Thus, it must be the case that there are two morpheme *ce*, one nominative and one Caseless. Alternatively, it might be given relevant features during the building of the structure. This aspect of *ce*, which can be clarified in further research, is not crucial for this article.

<sup>20</sup> Let's precise. The natural input for (51a) is (ia), which is ruled out, since the verb can't agree with *ce*, an element that has no feature, as discussed in the text. Moreover, recall that SI verbs must have an interrogative complement; so (ib) is also ruled out, since both arguments of the embedded verb are phi-specified. The alternative input resolves the agreement issue in (ia), while eliminating the matrix relative, (ic).

- (i) a. \*[Jane demande que][c'intéresse Marie]
- b. \*[Jane demande que][ça intéresse Marie]
- c. [Jane demande ~~que~~] [ce [qui intéresse Marie]]

<sup>21</sup> Keep in mind that (53a) is really the first step in the derivation for which the natural input can't even exist. It is the cleanup version of [c'est ~~que~~][que [qui intéresse Marie]]. The rearrangement of the brackets by reanalysis is not a step and nothing hinges on that matter.

<sup>22</sup> This type of interrogative constructions has many variants in popular and dialectal French. In Montreal popular French, the *wh*-pronoun does not appear in a dummy clause, as seen in (i), a fact traditionally referred to as doubly-filled complementizer.

- (i) a. [Qui [qui aime Marie]]
- b. [Qui [que Marie aime]]

<sup>23</sup> In French, there are the so-called attributive adjectives which require a copula and epithet adjectives which do not. It seems that the adjacency of the relative and the NP is interpreted as an epithet, which obviates the need to parse the copula.

Notice that if the subject is a clitic, that is, a Case-specified argument, it cannot exist without its supporting verb, and therefore deletion of the auxiliary is not possible. Compare: *qu'est-ce qu'il est?* and *\*qu'est-ce qu'il?* 'what is he?'