1. Introduction
Since Chomsky (1993), Copy as an operation to derive the displacement effect has been introduced into the syntactic theory. Bobaljik (1995) notes that the so-called Copy is an illusion. Gärtner (1998) points out a technical problem of the Copy Theory (CT) (see our 2.2), Lasnik (1999) argues that there is no copy for A-movement, and Epstein et al. (1998) straightforwardly deny the existence of the copy operation. Nevertheless, the use of this assumed operation is abundant in the current syntactic literature (e.g., Fox & Nissenbaum 1999, Hornstein 2001, Nunes 2001). Some claim that only CT can account for certain facts satisfactorily (e.g., Donati 2000, Fanselow & Cavar 2001b). To copy or not to copy has thus become one of the basic issues of syntactic theory. If CT is indeed problematic, one needs to be shown its conceptual and empirical inadequacies. The goal of this paper is to argue against CT. We will first present some problems of CT, then we will show how the Remerge Theory of movement (e.g., Epstein et al. 1998) is superior to CT, and finally, we will discuss problems or alternative approaches of some analyses where CT is claimed to be crucial.

Six problems of CT are discussed in section 2. In section 3 I advocate a remerge theory of movement. Then I argue in section 4 that constructions where identical elements occur are not derived by a copy operation in syntax. In section 5 I show that Frazier & Clifton's (2001) copy $\alpha$ is not a syntactic operation. In section 6, I discuss some constructions where CT has been claimed to play a crucial role. Section 7 concludes the paper.

2. Problems of the Copy Theory
CT assumes that movement contains two more steps than Merge: copy and PF-deletion of the silent copy, and that the occurrence of a tail copy in a movement chain captures the reconstruction effect. I present six problems of CT.

2.1 The Assumed PF-Deletion
Both the motivation and the operation of the assumed PF-deletion of CT is problematic. Chomsky (2000: 114) denies the deletion without an explanation: “copy theory is the simplest version of transformational grammar, making use of Merge, not Merge followed by an operation that deletes the original.” However, nearly all copy approaches use this deletion operation. Let us examine various possibilities. If the assumed copy operation created two copies that were equal with respect to phonological features, and the two copies were sent to PF, deletion of either of them at PF should be motivated phonologically, rather than by the assumptions such as only one copy can be interpreted at LF (Kitahara 1997, among others). Instead, one might claim that deletion were required since only one copy could be interpreted at PF. Let us consider two cases. On the one hand, if we assume that two identical elements cannot co-occur, how can we explain reduplication? On the other hand, if we assume that deletion of one of the two copies is required by linearization, then both copies should have equal chances of deletion, or the deletion should be sensitive to phonological conditions.

* An earlier version of this paper received comments both from ZAS colleagues Hans-Martin Gärtner, Paul Law, and André Meinunger, and from the participants of the conference Tools in Linguistic Theory (Utrecht, July 6-8, 2001), especially Klause Abels, Michael Brody, Barbara Citko, Chris Collins, Sam Epstein, and Andrea Moro. I am grateful to all of them. Remaining inadequacies are my own.
Although deletion of the upper copy instead of the lower one has been proposed for some constructions, the relevant licensing conditions are not phonological. Similarly, there is no phonological reason for deletion of the lower copy to be the default case. Alternatively, if the assumed copy operation had the effect that only the copy rather than the original could have phonological features, the operation would violate “the Inclusiveness Condition, which bars introduction of new elements (features) in the course of computation” (Chomsky 2001: 2).

In approaches such as Nunes (1995), the assumed PF-deletion is claimed to be required by Kayne’s (1994) LCA. Specifically, since the two copies of a movement chain occur in different positions, they fail to achieve a unified asymmetry c-command relation with another element. Therefore, deletion of one copy becomes necessary. It is clear that the presupposition of this LCA-driven PF-deletion is the existence of two copies of x. The viewpoint in which he seeks the unification of the c-command relation is the one where all derivational steps are arranged in a flat representation. If x does not have a copy, the whole LCA argument becomes irrelevant. We assume that “X C-commands all and only the terms of the category Y with which X was paired/concatenated by Merge or by Move in the course of the derivation.” (Epstein et. al 1998: 32) If Move of x is Remerge of x (see section 3), x c-commands a set of terms after its initial Merge, and it c-commands a different set of terms after the Remerge. At each step of the derivation, a new relation is established, the c-command relation is clear, and there is no reason to unify anything.

2.2 The Locality of Feature-Checking
CT contains a paradox with respect to the locality of feature-checking. Chomsky (1995) stipulates that the foot of a chain differs from the head in not being visible to the computational system. Hornstein (2001: 67) however assumes that all copies are grammatically equal. If x reaches the checking domain of y, a copy of x is left in situ. As noted by Gärtner (1998, 1999), if the locality condition on checking is taken in its strictest sense, the lower copy of x must not be affected when the higher one checks against y. In such a case, however, the computation will never exhaust the resources driving movement/feature-valuation operation. If we adopt Chomsky’s (2001) mutual feature-valuation theory of checking between x and y, how can we explain this double-checking of the same features of y by the two links of a single chain: a local checking against one copy of x and a remote checking against another copy of x? Chomsky (1995:381) states “Technical questions arise about the identity of α and its trace t(α) after a feature of α has been checked. The simplest assumption is that the features of a chain are considered a unit: if one is affected by an operation, all are.” Instead of stipulating this magic "one-equals-to-two" effect, one wonders whether what is involved in the operation of "movement" is in fact one element (x that undergoes the movement/remerge) rather than two (x and its trace/lower copy).

2.3 The Theta-Role Receptivity
CT creates a problem for the theta-role assignment. It has been stipulated that an argument cannot move to a theta-position. As noted by Lóbez (1999: 36), however, “it is not clear why the head and the foot of chain, being identical copies, are different with respect to theta-role receptivity.” This can be illustrated by the cases where object raising to vP is overt: it is not clear why v assigns a theta-role only to a subject rather than a raised object, although both are equally local (both are at SpecvP in Chomsky 1995). If the two copies of the object are grammatically identical, they should have equal chances to receive a theta-role. Note that in some CT approaches (e.g., Hornstein 2001), an argument can receive two theta-roles, thus logically, the raised object can receive another theta-role from v. Even we stick to the principle that one argument receives only one theta-role, we still face the problem why it is
the lower copy rather than the higher copy of the object that receives a theta-role, if the two copies are identical, and each of them is local to a theta-role assignor, v and V, respectively.

2.4 Reconstruction of Adjuncts
CT claims that reconstruction effects are the result of selective deletion of copies (Chomsky 1993, 1995: 202), however, as noted by Kim (1999: 14), CT cannot explain reconstruction effects of adjuncts.

Consider the contrast between (1) and (2), discussed by Heycock (1995):

(1) a. ?[How many stories about Diana] was she really upset by t?
   b. [How many lies aimed at exonerating Clifford] did he claim that he had no knowledge of t?

(2) a. *[How many stories about Diana] was she likely to invent t?
   b. *[How many lies aimed at exonerating Clifford] is he planning to come up with t?

Both (1) and (2) have R-expressions contained in an adjunct, which are inserted noncyclically after wh-movement (Riemsdijk 1981, Freidin 1986, Lebeaux 1988, 1991, Stepanov 2001). Reconstruction effects of Binding Condition C are present in (2), but not in (1). Unlike those in (1), the verbs which select the wh-phrases are “creation verbs” in (2).

On the other hand, due to semantics of the creation verbs invent and come up with, the quantificational n many phrases in (2) only have reconstructed scope reading, since the wide scope reading is pragmatically out. The wide scope reading for (2a), as spelled out in (3a), is impossible. (2a) has only the narrow scope reading in (3b).

(3) a. #For what number n, such that there is a particular set of n many stories x about Diana such that Diana is likely to invent x
   b. For what number n, such that Diana is likely to invent n many stories x about herself

Since (2a) is ungrammatical with the coreference between Diana and she, we should conclude that there is a correlation between binding and scope relation at LF. Given that (1) is good, (2) involves no Condition C violation either before the wh-movement of how many stories, or after both the wh-movement and the insertion of the adjunct about Diana. The only possible level to get such a violation is LF. The reconstruction of the quantifier phrase under the scope of the embedded verb yields the LF-representations (4), where the Binding Condition C is violated.

(4) a. *How_n [is she likely to invent [n many stories about Diana]]
   b. *How_n [is he planning to come up with [n many lies aimed at exonerating Clifford]]

However, this reconstruction effect cannot be restated as the effect of Copy. The intermediate structure of (2a) would be (5) in CT:

(5) [how many stories about Diana] is she likely to invent [how many stories]

Note that the adjunct phrase about Diana does not show up in the position of the trace. This is why there is no Condition C violation in the parallel construction (1). Without any
copy of the adjunct phrase in the trace position, CT cannot account for the reconstruction effect of the Binding Condition C.

CT claims that Copy can substitute the operation reconstruction, however we now see that it cannot. We thus wonder why we need the operation Copy.¹

2.5 Movement Paradoxes
The two instances of "movement paradox" listed by Breul (2001) challenge the CT, which claims that a displacement effect of a sentence is the result of the Copy operation and that the silent part of the sentence comes from the PF-deletion of one of the two identical copies. The first paradox is shown in (6), (7), and (8) (Bresnan (1991, 2001: 16ff.):²

(6) a. *He didn't think of that he might be wrong.
b. That he might be wrong he didn't think of.
(7) a. *We talked about that he was sick for days.
b. That he was sick we talked about for days.
(8) a. *This theory captures that languages are learnable.
b. That languages are learnable is captured by this theory.

In (6) and (7), if the that-clause in the b-sentences had an identical copy in a position where it is complement of the preposition of or about, we would expect the a-sentences to be grammatical, contrary to the fact. Similarly, if the that-clause in (8b) had an identical copy in a position where it is complement of the verb capture, we would expect (8a) to be grammatical, contrary to the fact.

The second paradox is shown in (9) (similar data can be found in Ward 1986/1988: 192ff.):

(9) They said we should stand firm, and {stand/?stood} firm we have

In (9), if the uninflected VP had an identical copy in the post-auxiliary position, the representation would not be well-formed.

The two cases look like a kind of anti-reconstruction effect of c-selection (in (6), (7) and (8)) and the inflectional feature-checking (in (9)).

2.6 The Compulsory Clone Operation
The obligatory copy operation of CT is ad hoc. No other case in our cognitive system has been proved to require a copy or clone operation to achieve a displacement effect. Moreover, this assumed operation is not related to any special property of human language.

All of these problems lead us to cast a reasonable doubt on the existence of grammatical identical copies of elements in the syntactic computation.

3. A Remerge Theory of Movement
It needs to clarify that there is no operation of copy in syntax. The only overt syntactic operation is Merge. Displacement effect of x is derived when x is remerged with y after it has been merged with z. Thus Move is “`Remerge,’ which formally unites both Merge and Move

¹ We have not seen how the Remerge Theory of movement solves the problem of adjunct reconstruction. A similar issue is discussed in Gajewski (2000). If QR and RL are post-cyclic operations, as suggested in Chomsky (2000), they share the Shortest principle with cyclic operations (Bruening 2001).
² But Bresnan also notes (2001: 24, n. 2) that "in some dialects of English examples like [7a] do occur."
as identical instances of pair/Concatenate.” (Epstein et. al. 1998: 45) I call this remerge theory of movement RM.

3.1 Derivation: Phrase Markers vs. Sets of Terms
Bobaljik (1995a) clearly shows that if syntactic derivation is viewed in terms of sets of terms, instead of phrase markers, which are only convenient notations to represent individual terms of a derivation rather than individual stages of the derivation, movement is not copying at all, and copying is simply an illusion or a derived effect. Let us see the details.

To derive the sentence A monster ate Fido, CT assumes the following steps:

(10) Array: a, monster, [+past], eat, Fido

a. \[\text{eat} \quad \text{Fido}\]
   \text{Merge}

b. \[\text{eat} \quad \text{Fido} \quad \text{a} \quad \text{monster}\]
   \text{Merge}

c. \[\text{a} \quad \text{monster} \quad \text{eat} \quad \text{Fido}\]
   \text{Merge}

d. \[\text{[+past]} \quad \text{a} \quad \text{monster} \quad \text{eat} \quad \text{Fido}\]
   \text{Merge}

e. \[\text{[+past]} \quad \text{a} \quad \text{monster} \quad \text{VP}\]
   \text{Copy}
   \text{a} \quad \text{monster}

f. \[\text{[+past]} \quad \text{a} \quad \text{monster} \quad \text{VP}\]
   \text{Merge}
   \text{a} \quad \text{monster}

Now let us see how to derive the same construction in terms of sets of terms. Merge is defined as in (11) (Chomsky 1995: 396):

(11) Apply to two objects \(\alpha\) and \(\beta\), Merge forms the new object \(\gamma\).

The terms of (10d), for instance, are listed in (12) (the numerals in front of each set are used for exposition, not for an order):

(12) 1. \{\text{Infl}\}
2. \{a\}
3. \{\text{monster}\}
4. \{\text{eat}\}
5. \{\text{Fido}\}
6. \{\{\text{eat}\},\{\text{Fido}\}\}
7. \{\{a\},\{\text{monster}\}\}
8. \{\{\{a\},\{\text{monster}\}\},\{\{\text{eat}\},\{\text{Fido}\}\}\}
9. \{\{\text{Infl}\},\{\{\{a\},\{\text{monster}\}\},\{\{\text{eat}\},\{\text{Fido}\}\}\}\}
From this list, we select two terms, say the 7th and 9th. We then apply the operation Merge to these two terms, that is, these terms from (12) should be the input terms (i.e., $\alpha$, $\beta$) to the operation (11). The output of this operation is the complex term $\gamma$ in (13). This expands the list of terms in the derivation by exactly one term, and the resulting list is given in (14). For convenience, the new term is added at the bottom of the list, though the order of the list is again arbitrary for presentation and irrelevant for the computation.

\[(13) \quad \alpha = \{{\{a\}},{{\{\text{monster}\}}}\} \]
\[\beta = \{{\{\text{Infl}\}}, \{{\{a\}},{{\{\text{monster}\}}}\}, \{{\text{eat}\}},{{\{\text{Fido}\}}}\}\} \]
\[\gamma = \{{\{a\}},{{\{\text{monster}\}}}\}, \{{\{\text{Infl}\}}, \{{\{a\}},{{\{\text{monster}\}}}\}, \{{\text{eat}\}},{{\{\text{Fido}\}}}\}\} \}
\]

\[(14) \quad 1. \quad \{\text{Infl}\} \]
\[2. \quad \{a\} \]
\[3. \quad \{\text{monster}\} \]
\[4. \quad \{\text{eat}\} \]
\[5. \quad \{\text{Fido}\} \]
\[6. \quad \{\{\text{eat}\},\{\text{Fido}\}\} \]
\[7. \quad \{\{a\},\{\text{monster}\}\} \]
\[8. \quad \{\{\{a\},\{\text{monster}\}\},\{\{\text{eat}\},\{\text{Fido}\}\}\} \]
\[9. \quad \{\{\text{Infl}\}, \{\{a\}},\{\{\text{monster}\}\}, \{\{\text{eat}\}},\{\{\text{Fido}\}\}\}\} \]
\[10. \quad \{\{a\},\{\text{monster}\}\}, \{\{\text{Infl}\}, \{\{a\}},\{\{\text{monster}\}\}, \{\{\text{eat}\}},\{\{\text{Fido}\}\}\}\}\} \]

Bobaljik (p. 51) states that Merge does not relate two phrase markers (e.g., (10a) to (10f)). Rather, the operation relates two subsequent stages of the derivation, e.g. (12) to (14). He also states that there is no restriction prohibiting any given term from serving as the input to more than one operation of Merge (p. 54). This is exactly what RM means. For instance, there are two operations which have the term $\{\{a\},\{\text{monster}\}\}$ as one of their input terms. In one step, this term is merged with the term $\{\{\text{eat}\},\{\text{Fido}\}\}$, and the 8th term in the list in (14) is formed. In another step, this term is merged with the term $\{\{\text{Infl}\}, \{\{a\}},\{\{\text{monster}\}\}, \{\{\text{eat}\}},\{\{\text{Fido}\}\}\}$, and thus the final term in the list in (14) is formed. These two steps are represented below:

\[(15) \quad \text{a. Merge } \alpha = \{\{a\},\{\text{monster}\}\} \]
\[\beta = \{\{\text{eat}\},\{\text{Fido}\}\} \]
\[\text{b. Merge } \alpha = \{\{a\},\{\text{monster}\}\} \]
\[\beta = \{\{\text{Infl}\}, \{\{a\}},\{\{\text{monster}\}\}, \{\{\text{eat}\}},\{\{\text{Fido}\}\}\}\} \]

We can see that the "copying" illusion comes from the remerge of $\{\{a\},\{\text{monster}\}\}$, not from the assumed copying operation.

3.2 The Differences between RM and CT
Conceptually, RM differs from CT in a straightforward way: the former assumes that to move x is just to merge x again, whereas the latter assumes that to move x needs to clone x first. Regarding the relation to the other overt operation, Merge, CT has to stipulate that a copy operation must be followed by a Merge operation. RM, however, does not need such an stipulation. The fact that Remerge must be preceded by an initial Merge is not a stipulation. Empirically, RM avoids the problems of CT (section 2), as shown below.
3.2.1 Merge and Move can be unified

RM captures the common property of Merge and Move. Kitahara (1994, 1995, 1997) makes an effort to unify the concept Merge (generalized transformation) and the concept Move (singulary transformation). The unification is advocated in Epstein et al. (1998: 19, 26, 37) and Epstein (1999: 319, 324, 329). Specifically, Epstein (1999: 324) claims that Merge and Move can at least partly be unified in that each pairs (concatenates) exactly two categories, A and B, rendering them sisters immediately dominated by the same (projected) mother, C (where C = the head of A or B). “[T]here is a fundamental operation, common to or shared by Merge and Move: ‘Concatenate A and B, forming C (C = the head of A or of B).’”

In the operation of Move, as claimed by Chomsky (1994: 19, also 1995: section 4.4.2, 2000, section 5) and argued by Nunes (1998), it must be the target that projects. Thus unlike the elements that undergo initial Merge, those undergoing Remerge never project.3

3.2.2 The issue of trace and reconstruction

CT eliminates a type of element, trace, at the cost of adding a new type of operation, Copy. The absence of trace in RM, however, follows directly from Merge, which simply pairs two available elements together, without leaving a trace anywhere. Traces are at best a notational device employed for encoding previous stages of the derivation in an output representation and therefore should be eliminated. Here we need to distinguish the occurrence of an operation from the occurrence of an assumed element (trace) to signal the occurrence of the operation. That we cannot undo an operation and that the derivation history has an effect on the interpretation do not mean that we need a trace. The notion of trace has been assumed to capture the fact that some operations may have an effect on the later operations or interpretation. It is this memory effect of derivations rather than the assumed notion trace or lower copy that creates the so-called reconstruction effects.

It has been widely discussed that certain types of operations and/or in some constructions there is no reconstruction effect, as we see in the A-movement in (16).

(16)  Everyone seems not to have arrived yet.   (∀ >>< ¬; * ¬ >> ∀)

The absence of a certain type of reconstruction effect does not mean the absence of a Move operation (contra Manzini & Roussou 2000, Sauerland 2000, Aoun & Li 2001. See Boeckx 2000 for arguments against Manzini & Roussou 2000). There are different accounts for the presence and absence of reconstruction effects. Boeckx (2001), for instance, links scope-reconstruction effects with Case-checking: "as long as an NP has an unchecked Case feature, its feature set is uninterpretable. Once Case is checked, the element is free for interpretation." (p. 318) In (16), for instance, everyone is not interpretable in its initial merge with the verb arrive. It is interpretable only after it is remerged with the matrix T projection and gets its Case-checked. This explains why there is no scope reconstruction in A-movement. This approach, however, cannot be extended to the various binding reconstruction asymmetries (see Munn 1994, Kim 1999, Barsb 2001, etc. for the asymmetries). Nevertheless, generally speaking, it is possible that a memory effect shows up in only certain types of the operations of language computation, as in other operation systems. It is possible that some interpretation rules are not sensitive to the history of a certain type of derivation. The

---

3 Iatridou et al. (2001: 224) claims that relativization is a case where the moved element projects, assuming that a modifier of a nominal must also be nominal and that “the relative pronoun moves to change the category of the relative clause” into a nominal. However, we are not convinced by the first assumption. In addition, relative clauses are syntactically different from other types of modifiers of nominals (Zhang 2001). The differences may be accounted for by the assumption that the clausal projection of relative clauses are selected by D (Kayne 1994, Bianchi 2000).
occurrence of a certain type of reconstruction effect is determined by the type of the syntactic operations itself, rather than by the choice between the deletion of the upper copy and the deletion of the lower copy at LF. The choice between the two ways of deletion has to be stipulated to reflect the type of the syntactic operations anyway.

Accordingly, since there is no counterpart of copy in RM, there is nothing to delete. Thus the above PF-deletion problem does not occur. Moreover, since there are no identical copies at all, neither the dubious “one-equals-to-two” effect of feature-checking occurs, nor does the issue of the equal chance of getting a theta-role exist.

3.2.3 RM and the Movement Paradoxes
CT is a representational approach, whereas RM is a derivational approach. Accordingly, CT requires the assumed two links of a movement chain to be identical. In contrast, there is no notion of chain in RM, and what matters in RM is the syntactic constraints of the target, the current sister, not assuming that the remerged element is identical to the silent part of the sentence. Consider the Movement Paradoxes introduced in 2.5 in the view point of RM. We repeat some of the relevant data below:

(17) a. *We talked about that he was sick for days.
   b. That he was sick we talked about for days.
(18) a. *We have stand firm.
   b. They said we should stand firm, and stand firm we have.

In RM, the ungrammatical (17a) is the result of the wrong c-selection of the Merge between about and the that-clause, and the ungrammatical (18a) is the result of the wrong Merge between have and stand (for whatever reason). In contrast, the b-sentences are grammatical because no such wrong Merge is seen. Specifically, our grammar allows a clause-initial topic to be a that-clause, as in (17b), and a bare VP, as in (18b).

3.2.4 RM avoids redundancy
CM brings a series of redundancy to the computation theory. RM avoids them. First, since there is no copy operation, we do not need the constraints proposed on copy and deletion. For instance, in order to avoid the case where x is copied but it is not immediately merged, Hornstein (2001: 100) stipulates that “A copy C made at step N of a derivation must be grammatically integrated at step N+1.” Such a constraint on Copy is not necessary in RM. Moreover, no effort such as Nunes (1995, 1999) is needed to find out why sometimes it is the lower copy that is deleted, and sometimes it is not. If x has a choice between a PF-realization and PF-null form (e.g., pro and PRO) in the initial Merge, so does it in Remerge.4

Second, since movement is remerge, it shares constraints with initial merge on both target and source.

With respect to the issue of the target of an operation, the Extension Condition or the Cyclicity Condition rules out not only downward movement, but also upward movement and Merge which do not extend the structure on the top.5 As pointed out by Bobaljik (1995b: 57), "the Extension Requirement is an axiomatic part of the definition of Merge: the operation simply and solely derives new terms". Thus one constraint on the target of an operation, the Extension Condition, is applied to both Move and Merge.

As for the issue of the source of an operation, Merge cannot combine elements which belong to another Array/phase, nor can movement start from a position internal to another

---

4 One general constraint is needed that a theta-relation is set up in the initial Merge only and thus no new theta-relation occurs in Remerge, if theta-criterion is true (but see Fanselow 2001 and Hornstein 2001).

5 See Epstein (2001) for a new account for syntactic head movement.
spelled-out phase (Chomsky 2000, Nunes & Uriagereka 2000). Thus one constraint on the source of an operation, the phase constraint, is applied to both Move and Merge.

Third, since Move is simply remerge, the so-called "chain" of movement cannot be a syntactic entity. It is just a convenient way to mention the abstract derivation history (just like the abstract transition from x’s previous marriage to the current marriage). As noted by Hornstein (1998), if chain is created in the computation, the derivation violates the Inclusiveness Condition.

3.2.5 The Cost of Remerge
RM captures the expensive nature of movement without a global consideration. If there is no available element in the Array to satisfy the relevant local feature-checking requirement, the generalized EPP (Chomsky 2000), remerging of a qualified element is the only choice. This is a last resort. If there is an available element in the Array to satisfy the relevant requirement, remerging of an element is expensive/illegal, since syntax needs another (unavailable) operation to deal with the unused element in the Array. The choice between Merge and remerge is thus decided within the syntactic computation of each phase. In contrast, if one claims that movement is more expensive than Merge because the former requires a post-copying PF-deletion while the latter does not (as in Kitahara 1995), one evaluates the two choices beyond syntax, since PF-processes count.6

4. On the occurrence of identical elements
We have argued that movement involves no copy operation. We now show that apparent copy configurations are not derived by a copy operation in syntax either. We consider four cases: predicate clefting in some Creole languages, the predicate cleft construction in Russian, the Verb-Copying Construction in Chinese, and the wh-copying construction in German.

4.1 Predicate clefting in some Creole languages
In some Creole languages such as Haitian and Papiamentu, predicate clefting is represented by reoccurrence of the verb and some other elements:

    'It is just buy he just buy flower PL
    ‘It is just buy flowers that he did (not, e.g. just sell).'

b. Se manje (*pen an) Jan manje pen an.
    It-is eat bread DET John eat bread DET
    ‘It is eat that John did to the bread (not, e.g., baking).’

(20) a. ta duna m’a dunabo e buki. (Papiamentu. Muysken & Law 2001)
    be give 1SG-PAST give-2SG DET book
    ‘I have really given you the book.’

b. *ta a duna m’a dunabo e buki.
    be PAST give 1SG-PAST give-2SG DET book

As shown in (19b), no complement can be clefted together with a verb in Haitian (Larson & Lefebvre’s (5b)). However, adverbs can, as in (19a) (Larson & Lefebvre’s (5a)). Since the reoccurring part such as fek achte ‘just buy’ in (1a) is not a constituent, the construction cannot be derived by a copy operation in syntax. One might claim that there is a

---

6 If a movement is optional, it is possible that the operation has a semantic effect (Chomsky 2001).
remnant-type of copy-operation, however, we need evidence for the claim. In (20), on the other hand, the left/higher copy of the verb cannot have a past tense marker while the right/lower one can. Assume that the tense marker has a feature-checking relation with T. Such data show that the higher copy does not, whereas the lower copy does, have a feature-checking relation with T. If the two identical verbs were derived by a copy operation, the contrast would be unexpected either from Hornstein’s claim that two copies are grammatically equal or from Chomsky’s stipulation that the lower copy is invisible in the computation. Thus it is unlikely that such constructions are derived by a copy operation in syntax. Larson & Lefebvre propose that the clefted part is pleonastic, and is base-generated in Spec of CP.7

4.2 The Predicate Cleft Construction in Russian
The Predicate Cleft Construction in Russian also contains reoccurrence verbs:

(21) Citat’ (-to) Ivan eë citaet, no nicego ne ponimaet. Readinf (TO) Ivan itfem.acc reads but nothing not understands

'Ivan does read it, but he doesn't understand a thing.'

There are at least three arguments against the copy-operation in the construction. First, since the upper copy of the verb must be in the infinitive form and the lower copy must be in the finite inflected form, the two copies are not grammatically identical. In Abels (2001) it is claimed that what is copied is the whole VP, and that in the lower copy, only the verb is pronounced and morphologically related to Infl. It is further claimed that the function of the verb in the lower VP-copy is like that of the do-support in English. It is not clear to us why the verbs in the assumed two VP-copies have different feature-checking relations with Infl, and as a general question to the whole CT (section 2.1 above), and what phonological arguments support the selective PF-deletion of the lower-copy. Second, the verb in the upper copy is in the infinitive form rather than the bare-verb form, this is not expected if it is the VP that is copied, even though infinitive forms are default forms of verbs in the language (Abels 2001 sec 4). Third, if the position of the upper copy of the verb is between CP and IP, as claimed by Abels, there is no way to rule out the possibility that the Array for this phase contains elements which are identical to that of the vP phase.

4.3 The Verb-Copying Construction in Chinese
Like the predicate clefting construction discussed above, the Verb-Copying Construction in Chinese (Tsao 1987, among others) also contains identical verbs and some other elements:

(22) a. Akiu mai (*-le) yinliao mai-le san ping Kele. Akiu buy -PRF beverage buy-PRF three bottle Coca-Cola

‘Buying beverage, Akiu bought three bottles of Coca-Cola.’

b. ni kan dianying kan-mei-kan-gou? you see movie see-not-see-enough

‘Have you seen enough movies?’

c. *ni kan-mei-kan dianying kan-gou?

7 It has been mentioned to me that a copy operation might not copy all features of an element, and then the lower copy of V in (20) may have some feature that has a checking relation with T of the embedded clause, but the upper copy lacks such a feature. I do not think this is possible if we accept Chomsky's (2001) mutual valuation theory of checking. If the matrix T fails have a checking relation with a verb, whereas the embedded T does have such a checking relation, the derivation should crash at the step of the matrix TP. However (20a) is acceptable.
(22a) shows that the left copy does not allow an aspect suffix such as the perfective le, whereas the right one does (Tsao p. 17). The sequence of verb-ASP should have a feature-checking relation with the head of AspP, which is projected somewhere above VP. Data like (22a) indicate that the higher copy does not, whereas the lower copy does, have a syntactic relation with Asp in computation.

The contrast between (22b) and (22c) shows that the left copy cannot have an A-not-A interrogative form, while the right one can (Shi 1996). Assume that the [Q] feature of an A-not-A predicate is licensed by, or has an Agree relation with, a [Q]-related feature of a c-commanding Infl/C, (Huang 1991). Such data suggest that the higher copy does not, whereas the lower copy does, have a syntactic relation with Infl/C in computation. If the two identical verbs were derived by a copy operation, the two contrasts in aspect markers and the A-not-A formation would be unexpected either from Hornstein’s claim that two copies are grammatically equal or from Chomsky’s claim that the lower copy is invisible in the computation. Thus it is unlikely that the verb-copying construction is derived by a copy operation in syntax.

I claim that the derivation of data like (22) simply involves identical elements in different Arrays. The VP containing the upper verb in the Verb-Copying Construction, such as mai yinliao ‘buy beverage’ in (22a), is argued to be base-generated at the surface-position, either as a VP-topic in Tsao (1987), or as an adverbial in Shi (1996). In either analysis, the VP is out of the vP where the lower verb of the construction is base-generated. Thus the two identical verbal phrases are not derived in the same phase. If every phase has its own Array, there is no way to avoid the case where identical elements occur in different Arrays.

4.4 The wh-copying construction in German

The wh-copying construction in German contains identical wh-words, as in (23a) (McDaniel 1989, Fanselow & Cavar 2001a, among others).

(23) a. Wen denkst du wen sie (*wen) liebt?
   who think you who she who loves
   ‘Who do you think that she loves?’

   b. Was denkst du wen sie (*wen/*was) liebt?
   what think you who she who/what loves
   ‘Who do you think that she loves?’

   c. {Wen/Was} glaub(s)t {du/*keiner/*jeder} wen sie liebt?
   who/what believe you/nobody/everybody who she loves

As shown in (23a), identical wh-forms cannot occur in situ. This suggests that if Copy is a syntactic operation, it could only apply to the second rather than the initial link of a chain. It is hard to explain this constraint syntactically. Note that German has wh-expletives, such as was in (23b). The wh-expletive construction and the wh-copying constructions share syntactic properties such as no wh-forms may occur in situ (23a-b), and no operator may intervene between the two wh-forms (23c). It is possible that the latter construction is a subcase of the former, and there is no syntactic Copy operation.\(^8\)

---

\(^8\) Kim (2001: 94) treats the upper copy of the wh-copying construction as a wh-expletive that is homophonous with its wh-associate (i.e., the lower copy). He also argues that there is no LF-movement of the wh-associate to the clause-initial wh-expletive in the wh-expletive construction.
To recapitulate, the apparent copy configurations are not derived by a Copy operation in syntax.

5. On the preference of identical structures in conjuncts

The question whether copy is an operation in syntax becomes more serious when the CT claims that it is cheaper to Merge an element than to Move an element via Copy (Chomsky 1993, Hornstein 2001, etc.), whereas some copy theory of conjunct-similarity claims that it is cheaper to copy a structure than to build a structure accumulatively (Frazier & Clifton 2001). This section is concerned with the latter claim.

Frazier & Clifton (2001) find that a conjunct is read faster if it is structurally parallel to its preceding conjunct than if it is not. (24a) was read more quickly than (24b):

(24)  a. Hilda noticed a strange man and a tall woman when she entered the house.
    b. Hilda noticed a man and a tall woman when she entered the house.

They claim that a copy α operation is present in (24a) and absent in (24b). However, the assumed operation has three properties, which are not shared by a regular syntactic operation. First, it is applied to conjuncts only. There is no parallelism effect if two elements are linked by a transitive verb rather than the conjunction and (Frazier & Clifton: 3). The following two sentences do not show the contrast in (24).

(25)  a. A strange man noticed a tall woman yesterday at Judi’s.
    b. A man noticed a tall woman yesterday at Judi’s.

This property runs against the general property of a syntactic operation: it cannot be construction-specific. Second, it copies a structure only rather than the elements which realize the structure. This second property is not found in syntactic operations either: no syntactic operation can affect a structure without affecting the elements which realize the structure. Third, in the absence of the assumed cheap structure-copying, the derivation never crashes. For instance, although (24b) takes longer time to process than (24a), both sentences are grammatical. In contrast, choosing an expensive operation instead of an available cheap operation in computation must cause a crash, and thus no grammatical sentence can be derived. These three properties indicate that the assumed copy α looks more like an assimilation tendency in human parsing mechanism than an operation in syntax. Frazier & Clifton state in their footnote 11: “In fact we suspect that Copy α is just the linguistic reflex of a more general cognitive ability that we dub ‘ditto’.” It is possible that what they reveal in their study is a reflection of the general law of inertia or assimilation in language processing. We conclude that the so-called Copy α is not a syntactic operation.

6. On Some Analyses Where the Copy Theory Plays a Crucial Role

6.1 Feature Copying?

One hypothesis of covert copying is motivated for feature-checking. Watanabe (2000) proposes a ϕ-feature copying approach to complementizer agreement and switch reference. In fact, as noted by himself (p. 167, fn 5), the assumed feature copying operation is hard to distinguish from Chomsky’s (2000) covert operation of Agree in situ, empirically. However, he claims that the value assignment of the Agree operation should not be applicable to the ϕ features of verbs, and he thus still insists on feature copying rather than covert Agree. I fail to find an argument in his paper against an Agree operation on the ϕ features of verbs. It is
possible that the $\phi$ feature checking of verbs by a nominal is directly related to $\theta$-role assignment (Fanselow 2001).9

6.2 VP-Ellipsis by Copying?

VP-Ellipsis has been claimed to undergo copying. There are two versions of this copy hypothesis. Wasow (1972) and Williams (1977) assume that the elided part is generated as a phonological null element and a copying operation copies the semantic features of the antecedent to the elided part at LF. Donati (2000), on the other hand, assumes that the elided part is originally fully copied from its antecedent in syntax, and then gets deleted at PF. This copy hypothesis faces the following problems. First, the presence of a sloppy reading indicates that the interpretation of the elided part can be different from that of the antecedent and thus semantic features are not copied, as shown in (26).

(26) Mary has always loved her job and John has too.
   a. Mary has always loved Mary's job and John has always loved Mary's job. (strict)
   b. Mary has always loved Mary's job and John has always loved John's job. (sloppy)

Second, phrases can be elided under non-identity with its antecedent, suggesting that there cannot be a copying operation. The following data are cited from Potsdam (1997: 358) and Johnson (2001: 468):

(27) a. I didn't touch the TV, but Percy might have touched the TV.
   b. John said that he would never take money on the side but I knew he was taking money on the side.
   c. We haven't decided to blacklist any firms. But there's a chance we might blacklist some firms.
   d. David Begelman is a good laugh and when he does laugh, his eyes crinkle at you the way Lady Brett's did in The Sun Also Rises.

Moreover, sentences potentially containing ellipsis can be pronounced without any deletion:

(28) Mary will leave tomorrow, and Paul will (leave) the day after.

The possibility of PF-realization of the elided part makes one wonder why the elided part cannot be generated independently with identical or near identical categories and structures to its antecedent, regardless of what mechanism makes it silent.10

In order to argue that "the deleted constituent and its antecedent are not two lexical items taken from the numeration, but two copies generated in the syntax" (Donati 2000: sec 3.2), Donati uses the so-called "parallelism effects" to show a dependency constraining the

9 In Watawabe's (2000) data of complementizer agreement, both complementizers and verbs have subject-agreement markers. Thus both contain $\phi$ features, which are uninterpretable to them. In this case the interpretable $\phi$ features of the subject have multiple uninterpretable counterparts on the head elements. The reverie situation can be seen in the multiple occurrences of the Nominative Case feature in languages such as Japanese and Korean. According to Pesetsky & Torrego (2000), Nominative case of an argument is an uninterpretable counterpart of the interpretable tense on T. In this case, the interpretable tense feature of a head has multiple uninterpretable counterparts on the nominals. In both cases, an interpretable feature is multiply "misplaced" in some elements and thus becomes uninterpretable. It is possible that the two cases may have similar feature-checking operations.

10 Considering the optionality fact, the copy approach does not want to rule out this possibility of independent Merge rather than copy (Donati 2000: sec 6).
interpretation of the assumed two copies. For instance, the VP in the first conjunct of (29) is ambiguous between two comically opposed readings. If the elided part in the second conjunct received its interpretation independently from the one of its antecedent, she argues, we would expect it to be ambiguous as well. However, in fact, the elided part inherits its interpretation from the first conjunct. As a result, (29) is only two-ways ambiguous rather than four-ways ambiguous.

(29)  The children are ready to eat and so are the chickens.

This "parallelism effect", however, is similar to what we discussed in section 5. It is a principle of economy in processing that plays a role here. Thus this effect cannot be used to argue for a copy operation in syntax.

6.3 XP-Splits by Copying?
Fanselow & Cavar (2001b) analyze the XP-Splits in languages such as German and Croatian.

(30)  a. knjige mi je Marija zanimljive preporucila (Croatian)
books me has Mary interesting recommended
‘She has recommended interesting books to me.’
b. Mit was hast du für Frauen gesprochen? (German)
with what have you for women spoken
‘With what kind of women did you speak?’

In (30) the underlined two parts should be read as a continuous single constituent in each sentence. Fanselow & Cavar list some properties of XP-splits. For instance, they arise in the context of operator movement only, and they can retain or invert the order of the elements found in continuous counterpart. Considering the island effects and the phonetic shape of the parts of the split phrase, they argue against a movement approach, which assumes that the part of the XP appearing in clause-initial position is moved out of the XP, stranding the material left behind. They also argue against a base-generated approach, which assumes that the two split parts are each base-generated at their surface positions. They conclude that the discontinuous parts must be derived form a continuous category. They propose a complete-copying plus distributed deletion account. For instance, (30a) is derived as in (31).

(31)  mi je Marija zanimljive knjige preporucila =>
me has Mary interesting books recommended
zanimljive knjige mi je Marija zanimljive knjige preporucila =>
complete copying
partial deletion in upper copy
zanimljive knjige mi je Marija zanimljive knjige preporucila =>
complementary deletion in lower copy
zanimljive knjige mi je Marija zanimljive knjige preporucila

However, there is an alternative analysis to this complete-copying & distributed-deletion approach: complete base-generation of a topic/focus & distributed-deletion. (32) (Fanselow & Cavar’s (11), cited from Frey 2000) shows that if there is a sentential adverbial, such as wahrscheinlich 'probably,' the left part of the XP-splits must precede the adverbial in German.
(32) a. dass er teure Bücher wahrscheinlich der Frau keine schenken wollte

that he expensive books probably the.DAT woman no give

wanted

‘...that he probably did not want to give the woman expensive books as a gift.

b. ?*dass er wahrscheinlich teure Bücher der Frau keine schenken wollte

This contrast indicates that teure Bücher is out of the vP-phase. If movement approach is wrong, it is possible that keine teure Bücher is base-generated in both vP, as a direct object, and in next phase, CP, as a topic, and then distributed deletion occurs to derive the surface split configuration. If we cannot rule out this alternative analysis, we cannot use the XP-splits as an independent argument for the copy operation in syntax.

6.4 Overt vs. Covert Operations

In the current syntactic literature, quite a few authors argue against the division between overt and covert (i.e., post-spellout) component of the grammar (Kayne 1998, Fox & Nissenbaum 1999, among many others). However, as pointed out by Fox & Nissenbaum (1999: 132), the result of this is that there must be some alternative means for distinguishing “overt” from “covert” operations. They mention two ways. One is to adopt CT. They cite Bobaljik (1995b), Pesetsky (1998), and Grout & O’Neil (1994). In this CT approach, “overt” movements are the result of phonology targeting the upper copy of a chain for pronunciation, while “covert” movements result from phonology targeting the lower copy of a chain. This approach is also seen in Hornstein’s (1995) analysis of QR. Since we are arguing against the CT approach, let us consider the other way offered by Fox & Nissenbaum. This is the multiple Spellout approach: “there are many instances of spellout, each one updating a previously computed PF. Under this assumption, a principled account must be local, updating only information that was introduced by the most recent operations. Covert operations would be ones that are not immediately followed by a spellout operation.” (p. 135) In any case, there are alternative ways to substitute CT. We leave the technical details for future research.

7. Conclusion

I have argued in this paper that while Merge is an available operation in syntax, Copy is not, and that overt movement is simply Remerger of a given term. RM avoids many problems of CT. In processing of conjunction constructions, the parallelism effects are the result of the general law of inertia. No syntactic copy operation is involved. In syntax, the assumed copy machine is expensive. It's time to sell it.
References
Bobaljik, J. 1995a In Terms of Merge: Copy and Head Movement. MITWPL 27, 41-64.
Fox, D. & J. Nissenbaum 1999 Exposition and Scope: A case for overt QR. WCCFL 18, 132-144.


Stepanov, A. Late adjunction and Minimalist Phrase Structure. *Syntax* 4:2, 94-125.
Zhang, N. 2001 The Deals Made Among Asps and Ds in Relativization. Ms. ZAS-Berlin.