TOWARDS A MINIMALIST TREATMENT OF CERTAIN ISLAND PHENOMENA AND THEIR CIRCUMVENTION

Florian Zellmayer (9500166, 327295)
Dept. of Linguistics, UVienna

Vienna, 8th of June 1997

Abstract:

This paper argues that the ban on extraction out of constituents entering a chain relation can be derived within the Minimalist Program without any additional stipulations if certain modifications by Nunes (1995) are adopted. Crucially, chain formation required for convergence reasons is seperated from the mechanism that merges the chain members, created by a copy rule, in their respective launching sites. Depending on the ordering of particular applications of the copy rule, the relevant contrasts can be derived from the impossibility of chain formation with certain orderings of operations. This claim gives an indirect argument against representational views of grammar which cannot easily make the relevant distinctions. Another goal of the approach defended here is that it neutralizes part of the A/A-bar distinction, clearly a desirable result on minimalist assumptions.

Zusammenfassung:

Dieses Papier argumentiert, daß das Verbot der Extraktion aus bewgeten Konstituenten innerhalb des Minimalistischen Programmes ohne zusätzliche Stipulation abgeleitet werden kann, wenn gewisse Modifikationen durch Nunes (1995) akzeptiert werden. Relevanterweise wird die Kettenkonstruktion wie sie aus Konvergenzgründen notwendig ist, von dem Strukturbildungsmecha-nismus, der die durch eine Kopierregel erzeugten Kettenglieder in ihrer jeweiligen strukturellen Position insertiert, getrennt. In Abhängigkeit von der relativen temporalen Anordnung der einzelnen Applikationen der Kopierregel in der Derivation können die relevanten Kontraste aus der Unmöglichkeit der Kettenbildung bei den jeweiligen Aufrufanordnungen der Operationen abgeleitet werden. Dies bildet auch ein indirektes Argument gegen repräsentationelle Grammatikmodelle, in denen die relevanten Distinktionen nicht ganz einfach gemacht werden können. Ein anderer Vorteil des hier vertretenen Ansatzes ist, daß er die A/A-Quer-Distinktion teilweise neutralisiert, wobei es sich vom Standpunkt des Minimalisten her gesehen zweifelsohne um ein wünschenswertes Resultat handelt.

Acknowledgenments:

The author is indebted to Werner Abraham (Rijksuniversiteit Groningen, Germanistisch Instituut) and to Martin Prinzhorn (University of Vienna, Department of Linguistics) for their encouraging help during writing the manuscript. The material was discussed in classes given at the University of Vienna in SS 1997 at the Departments of Linguistics (Nunes 1995) and German Philology (Uriagereka 1997). Many thanks to Susanne Habla for doing editorial work before publication.

0. Introduction

The paper is organized as follows. Section 1 summarizes the relevant data and briefly comments on empirical generalizations that may come to mind.

Section 2 summarizes the Minimalist Program as outlined in Chomsky (1995) and presents the system of Nunes (1995). Some Minimalist accounts of the data are discussed. Section 3 presents two basic assumptions: Nondistinctness marking applies only to full items but not to terms and COPY is modified such as to immediately feed MERGE.

Section 4 justifies these assumptions to a sufficient degree to allow us to build our account on them. Section 5 presents the basic claim. Chain formation into moved constituents is only partially possible due to the lack of nondistinctness marking on particular instances of the extracted item if movement of the extraction domain PRECECEDES extraction, but not if ordering is reversed. Section 6 shows that ordering of operations is crucial to our proposal and discusses some implementations of covert movement. Section 7 explains how the problems for chain formation deduced in section 5 can be circumvented by multiple extraction such as in parasitic gaps. Ordering and getting nondistinctness marking onto all instances of the extracted item is crucial once again.

Section 8 does away with some apparent objections to our proposal, both technical and empirical. Section 9 briefly mentions true empirical counterevidence but comments on it only very briefly. Section 10 concludes.

1. Movement Islands

In this section I summarize certain configurations that do not allow extractions and can be dubbed "movement islands" for convenience. Certain configurations that have standardly been referred to as Strong Islands can be summarized under this label and be assimilated to other configurations with a similar behaviour. In what follows I will somewhat idealize judgements for reasons of exposition, but this does not affect the proposal since the contrasts are still sharp.

There are five structural configurations that have been referred to as Strong Islands with the major difference between Strong Islands and Weak Islands being that the latter allow certain types of objects to escape from them whereas the former posit a nonselective ban on extraction but they allow circumvention by parasitic gaps, which facts are well known. Consider now Strong Islands.

1 *who$_i$ did [friends of t_i] kiss the girl
*who$_i$ were [pictures of t_i] sold by John
*who$_i$ were [pictures of t_i] on sale
2 *who$_i$ would [for the girl to kiss t_i] displease John
3 *who$_i$ did Bill believe [friends of t_i] to have kissed the girl
*who$_i$ did Bill believe [pictures of t_i] to have been made by John

*who$_i$ did Bill believe [pictures of t_i] to be on sale
4 *who$_i$ did John leave the party [after having kissed t_i]
*who$_i$ did John leave the party [after he kissed t_i]
5 *who$_i$ did [a girl that kissed t_i] visit Bill
*who$_i$ did Bill visit [a girl that kissed t_i]
*who$_i$ did [a girl that kissed t_i] fall asleep

The examples in (1) are Nominative Islands, those in (2) are Sentential Subject Islands, (3) have no standard name and may be dubbed Exceptionally Case Marked DP Islands, (4) illustrates Adjunct Islands, (5) represents Complex NP Islands.

One immediately notices that (2) radically improves if replaced by a string with an expletive filling the subject position and the true subject staying down.

6 who$_i$ would it$_j$ displease John [for the girl to kiss t_i]$_j$

This suggests that the relevant difference in (2) and (6) is either adjacency to the C° (proximity of movement) or movement of the subject. The former may be dismissed for well known reasons.

The second option seems much more promising, as even a superficial look at some of the other examples above shows. Nominative Islands and Exceptionally Case Marked DP Islands both involve movement in English. Furthermore, as seen in (6) subjects do not act as barriers for extraction if staying down in their base position.

Lasnik & Saito (1992) present examples showing that consitutents sitting in A-bar-positions similarily do not allow extraction. These examples are degraded to a varying degree but the contrasts seem sharp in any case. Among the relevant data are extractions out of WH-moved items and topicalized constituents. One might also add Heay NP Shifed DPs, as pointed out by Rochemont (1997).

7 *who$_i$ did John ask [which pictures of t_i]$_j$ Mary sold t_j
8 *vowel harmony$_i$, I think that [books about t_i]$_j$, you should read t_j
9 *who$_i$ did John give t_j to Mary [some recent descriptions of t_i]$_j$

The considerable marginality of these examples suggest that movement rather than A-bar-properties of the launching site may be relevant - as speculated above. Nevertheless, this claim can easily be related to the A-bar-position account, since such positions are generally reached by movement rather than filled by base merger. This follows from reconstruction effects that indicate some position further down in the tree that is relevant for interpretation. By FI, one would then assume that movement is involved.

By picking out the upper examples in (1) to (3) we loose the account of Adjunct Islands and Complex NP Islands. Despite their being genrally subsumed under the same descriptive label we are not forced to posit a natural class for (1) to (5). The movement generalization, which is by no means new, has at least the single advantage of freeing us from taking into account the A-bar-properties of the examples in (7) to (9) since these are obviously difficult to formulate in the Minimalist Program and even more difficult to relate to extraction.

We will leave open the question of how to assimilate Adjunct and Complex NP Islands if such an assimilation is indeed desirable. The analysis of Heavy NP Shift cases by Rochemont (1997) may also extend to adjunct cases due to the notorious heavyness of these constituents tough Rochemont's (1997) account involves a focus head.

Similar suggestions as far as movement is concerned have been made by Branigan (1992) although he formulated them in terms of "sitting in SpecAgrP" in such a way that this particular postion renders a constituent that occupies it a barrier. Nevertheless, as with A-bar-positions, SpecAgrP is in general reached by movement. We thus might view this generalization, which is drawn from the inavailability of WH-extraction out of Nominative and exceptionally Case marked subjects, as basically an instance of the more general ban on extracting out of constituents that have been moved.

Nevertheless, note that there is at least a single treatment that dispenses with A-movement and hence with the availability of this generalization, namely that of Manzini & Roussou (1996). This proposal crucially relies on Chomsky's (1995) assertion that A-movement does generally not involve reconstruction. Unfortunately, this is not true:

10 [pictures of himself$_i$]$_j$ seem to John$_i$ [t_j to be [t_j flattering]]
11 [pictures of himself$_i$] bother John$_i$
12 John showed [each other]$_i$'s friends to [Bill and Mary]$_i$

We might thus safely assume that there is indeed A-movement which has been used by Hornstein (1995) to account for quantifier scope and some of its characteristics such as interaction with binding, bound variable readings, or clause boundedness. The success of such a unification may lend additional support to our rejection of Manzini & Roussou's (1996) claim.

The unification of Nominative Islands and Exceptionally Case Marked DP Islands in terms of movement may be further strengthened by the behaviour of indirect objects, which also do not allow extraction.

13 *who$_i$ did John give [only few children of t_i] some candy

The marginal acceptability of adverbs intervening between the indirect and the direct object in (13) indicates that the indirect object has been moved out of its base position (Koizumi 1995). Similarily the indirect object may be associated with a floating quantifier which, given the standard treatment of this construction as involving a sranded rather than a floated quantifier (Sportiche 1988), is also indicative of movement. Additionally, indirect objects seem to pattern with subjects in that they may contain "obligatory" parasitic gaps (Engdahl 1983).

We thus have a rather robust basis of constituents sitting in SpecAgrP (A-movement), and others sitting in SpecCP, SpecTopP, or else have been HNPSed (A-bar-movement) that justifies subsuming all these constructions under the term "movement island". Furthermore, giving them a unified account may be justified even if we loose the standard Strong Island generalization.

2. Theoretical Framework

2.1. The Minimalist Program (Chomsky 1995)

The Minimalist Program reduces the model of competence to an unavoidable minimum forced by conceptual necessity. This is because language faculty is embedded to Performance Systems determining its design by a need of being supplied with interpretable pieces of information. Since there are two such Performance Systems, the Conceptual-Intentional System and the Articulatory-Perceptual System, only two levels of representation (LF and PF) are conceptually necessary and no other should be postulated in addition. For information must somehow be stored and needs being associated with instructions for its use, a Lexicon is expected by conceptual necessity too. The overall use of language involves recurrent patterns of structured information to allow processing and one might thus expect the Performance Systems to require them too in order to have a base for interpretation, which is the purpose of the Computational System arranging lexical information in a particular way. In the optimal case, this should be all that is needed.

Language is a generative procedure forming derivations, each a sequence of subsymbolic structures $<S_1,...,S_n>$ where S_n is a pair of LF and PF representations. These are sent to the respective interfaces for evaluation where they converge if they satisfy Full Interpretation at the level in question and crash if not. In addition, Full Interpretation must be satisfied in an optimal way where optimality is evaluated by Economy Conditions. The subset of convergent derivations is sorted out of the set of possible derivations, and the convergent derivations are compared for Economy with the most economical one being selected as the only admissible one. The reference set for Economy is defined as the initial Numeration, a set of lexical items each with an indication of how many times it has been selected. Thus, only derivations from the same numeration are compared.

Five operations are available in the Computational System: SELECT takes an item from the numeration, introduces it into the derivation, and reduces its index by one. MERGE concatenates two items as an unordered set $\{\alpha,\beta\}$, embeds this set into a further set $\{L,\{\alpha,\beta\}\}$, and determines the label L. ATTRACT targets a complex item, copies one of its terms, and merges this at the root of the targetted item, if a Checking Relation is established and the Minimal Link Condition is respected. SPELL-OUT splits the derivation into a branch leading to the LF representation and another giving the PF representation. DELETE deletes feature sets. Conceptual necessity leads us to expect the computation from the initial numeration to the representation at LF to be uniform with respect to the operations available (Uniformity). Furthermore it may require Inclusiveness, which says that derivations should be tautological, that is only dealing with information already present in the lexicon without adding any new information.

Local Economy conditions which select the most economical next step at any stage in a given derivation are incorporated to the definition of ATTRACT (Minimal Link Condition). This may not be violated since otherwise the derivation would be canceled by an illegitimate operation being carried out. Global Economy conditions which select only among convergent derivations can be violated for convergence.

"Procrastinate" prefers covert movement to overt movement if possible, i.e. not forced by feature strength. "Have an Effect on the Output" says that an item enters the numeration only if somehow affecting the output and the "Smallest Derivation" Principle prefers shorter derivation to longer ones.

There are several types of features that constitute a lexical item. Semantic features and phonological features are relevant only to interpretation at LF and PF, respectively. Formal features are accessible to the Computational System where noninterpretable formal features have a need to be eliminated in the derivation since they induce LF crash by not being interpretable at this level. Elimination is by checking, i.e. sitting in a local domain with a matching feature. Interpretable formal features are not affected by checking and can enter multiple checking configurations. Strong features are noninterpretable formal features that have to be eliminated before a larger category not headed by what bears them is created. Language variation with respect to linearization is captured by strong features.

2.2. The Modifications by Nunes (1995)

In what follows we will only sketch the most relevant innovations, for further details see the highly revealing Nunes (1995). These mainly concern the operations available which will finally be illustrated by a sample derivation.

Crucially, Nunes (1995) splits the operation MOVE into three constitutive suboperations which are not necessarily associated with each other, copy creation, merger, and chain formation. MERGE is defined approximately as standard (see Chomsky 1994, 1995). COPY may informally be defined as follows (my wording):

14 COPY: Target α, create a copy α', if α already bears nondistinctness marking, add it to α' too, otherwise add nondistinctness marking to both α and α'.

FORM CHAIN mimics the behaviour of MOVE with respect to chain construction in a representational way by forming a chain if two items qualify as such and are in the appropriate structural configuration. There are some minor differences which are not relevant here. Last Resort, Minimal Link Condition (plus Closeness), and c-Command Condition are incorporated. It may be defined as follows (my wording).

15 FORM CHAIN: Target α, β and form a chain CH=(α,β) iff
- (i) α c-commands β.
- (ii) α and β are nondistinct.
- (iii) At least one given feature F of α enters into a checking relation with a sublabel of K, where K is the head of the projection with which α was merged, and the corresponding feature of β could enter into a checking relation with a sublabel of K. (Last Resort part)
- (iv) There is no γ with a feature F' which is of the same type as the feature F of α and is closer to α than β. (Minimal Link Condition part)

Nondistinctness marking is introduced by COPY only, all terms with no such marking are interpreted as distinct by default (see section 4 for some justifying discussion of a strict interpretation of this claim). We will present a revised definition of COPY as we proceed.

Crucially, heads and foots of chains behave alike with respect to feature checking - only the head may enter a checking relation - beacause chains can only be read off by (15) after the head entered a checking relation. That is, unchecked features of foots of chains must somehow be eliminated for convergence with elimination forced by a condition that requires chains to be uniform with respect to feature constitution at LF (Feature Uniformity Condition). There are basically two ways of doing this:

16 CHAIN UNIFORMIZATION: Delete the minimal number of features of a nontrivial chain CH in order to allow its links to satisfy the Feature Uniformity Condition. (Nunes 1995)

Similarily, unchecked formal features may be deleted in the phonological component by the following operation (CHAIN UNIFORMIZATION may apply either overtly or covertly, but in general it will do so covertly) - LINEARIZE is the rule converting asymmetric c-command relations into a particular linear ordering in the phonological component:

17 FORMAL FEATURE ELIMINATION: Given the sequence of pairs σ = <$(F,P)_1$, $(F,P)_2$, ..., $(F,P)_n$)> such that σ is the output of Linearize, F is a set of formal features and P is a set of phonological features, delete the minimal number of features of each set of formal features in order for σ to satisfy Full Interpretation at PF. (Nunes 1995)

Use will be made of this condition to determine by Economy which member of a nontrivial chain is deleted in the phonological component, usually it will be the tail rather than the head.

Since nondistinct terms cannot be properly linearized in the phonological component (Linearize), deletion must reduce chains in such a way that this becomes possible.

18 CHAIN REDUCTION: Delete the minimal number of terms of a nontrivial chain which suffices for CH to be mapped into a linear ordering in accordance with the LCA. (My wording)

Many details have been omitted but we will nevertheless illustrate the workings of this system in an example.

Suppose, we want to derive the sentence "John likes Mary" where the numeration comprises (<u>John</u>, <u>Mary</u>, <u>likes</u>, υ, T). The derivation goes as follows.

19 SELECT (Mary)
SELECT (likes)
MERGE (Mary, likes) → $[_{VP}$ likes Mary]
SELECT (υ)
MERGE (υ,VP) → $[_{υP}$ υ° $[_{VP}$ likes Mary]]
SELECT (John)
MERGE (John, υP) → $[_{υP}$ John υ° $[_{VP}$ likes Mary]]
SELECT (T)
MERGE (T, υP) → $[_{TP}$ T° $[_{υP}$ John υ° $[_{VP}$ likes Mary]]]
COPY (John)

MERGE (John) → [$_{TP}$ John T° [$_{\upsilon P}$ John υ° [$_{VP}$ likes Mary]]]
FORM CHAIN (John, John)
SPELL-OUT ([$_{TP}$ John T° [$_{\upsilon P}$ John υ° [$_{VP}$ likes Mary]]])

Obviously, every step in (19) is as explained above and not in need of further comment. SPELL-OUT sends to the phonological component a structure with semantic features stripped away and phonolocigal and formal features still in place such that the feature constitution will be as follows (checked features are indicated by bracketing and nondistinctness marking by underlining).

20 [$_{TP}$ John$_{-D}$ T° [$_{\upsilon P}$ John-D υ° [$_{VP}$ likes Mary]]]

As far as linearization is concerned everything is fine except with the two nondistinct copies of John. These cannot be linearized since John sould precede itself in (19) and should precede and follow T° at once. Furthermore, the lower copy of John should violate Full Interpretation at PF since it contains an unchecked formal feature [D] which is assumed to be uninterpretable at PF. In order to get (20) conform PF BOCs we have two options to proceed.

21 CHAIN REDUCTION (John, John) → [$_{TP}$ T° [$_{\upsilon P}$ John-D υ° [$_{VP}$ likes Mary]]]
LINEARIZE ([$_{TP}$ T° [$_{\upsilon P}$ John-D υ° [$_{VP}$ likes Mary]]]) → <T°,John-D,υ°,likes,Mary>
FORMAL FEATURE ELIMINATION (<T°,John-D, υ°,likes,Mary>) → <T°,John,υ°,likes,Mary>

22 CHAIN REDUCTION (John, John) → [$_{TP}$ John$_{-D}$ T° [$_{\upsilon P}$ υ° [$_{VP}$ likes Mary]]]
LINEARIZE ([$_{TP}$ John$_{-D}$ T° [$_{\upsilon P}$ υ° [$_{VP}$ likes Mary]]]) → <John,T°,υ°,likes,Mary>

Both the continuations in (22) make the derivation converge at PF but by Economy (22) will block (23) since it involves fewer operations. Thus despite the fact that there are no inherent differences between heads and tails of chains (both can induce MLC effects, both can only have features checked if entering a checking relation locally, etc.), the correct PF string results.

Nunes' (1995) system differs from that of Chomsky (1995) in an number of additional interesting respects but they are not relevant to the present discussion and hence will stay unmentioned.

2.3 Minimalist Accounts of the Relevant Data

Uriagereka (1997) presents a strongly derivational model that can account for the data presented in section 1 by crucially relying on the left branch vs. main branch distinction. He starts by deriving the Linear Correspondence Axiom (Kayne 1994) according to which asymmetric c-command relations in a tree must be unambiguous with respect to each pair of nonterminal nodes immediately dominating a terminal. These relations may then be mapped onto a linear ordering. Uriagereka (1997) derives the base step of the LCA by assuming that a mechanism directly converting the asymmetric c-command relations into a linear ordering without any additional machinery would be conceptually simpler than one involving additional manipulations. Crucially this is possible since the nonambiguity conditions on asymmetric c-command are similar if not identical to those required for a linear ordering.

The induction step of the LCA - requiring that if α that dominates β precedes γ then β should also precede γ - is indirectly derived by assuming that it simply does not exist. Then left branches for instance could never have their terms linearized. Uriagereka concludes that it consequently must be the case that they are linearized, i.e. spelled out, independently of the main branch. On the assumption that linearization converts a set $\{\alpha,\beta\}$ into another set $\{\{\alpha\},\{\alpha,\beta\}\}$ and by strictly adhering to Chomsky's (1995) definition of syntactic object, Uriagereka concludes that the output of linearization is not a syntactic object and hence should behave as a word-like unit with respect to the Computational System. The strongly derivational design results from splitting the derivation in seperate derivational cascades each with terminal application SPELL-OUT with the objects assembled in such a way combined to a single phrase marker.

Consider for instance a Nominative Island violation as discussed by Uriagereka.

23 who$_i$ were [pictures of t$_i$] sold by Bill

Adopting the modifications of Nunes (1995), pictures of who is assembled in a separate derivational flow and a copy of who is made. pictures of who is then spelled out and inserted to the main branch. After assembling the final structure in (23), no chain could be read off and the unchecked feature of the lower copy of who cannot be eliminated since who inside the Nominative Island is no longer a syntactic object within the whole Nominative Island.

Although such examples as (23) - and all of the movement islands discussed in section 1 since they all involve left branches - are accounted for, it is unclear how such islands could be circumvented by parasitic gaps.

24 who$_i$ did [friends of t$_i$] admire t$_i$

Indeed, the unchecked WH-feature inside the island in (24) should make the derivation crash as it does in (23) since the addition of another copy of who does not alternate the word-like status of the island and thus should still block chain formation.

Furthermore, one wonders why the unchecked feature in (23) could not have been eliminated by FORMAL FEATURE ELIMINATION since if a feature is part of a subpart of a word it should still be accessible as a feature of the whole word, in this case the spelled out pictures of who just as the 3SG suffix and the stem in likes are parts of a word inaccessible to syntax but nevertheless visible to processes able to eliminate uninterpretable formal features (ordinary checking or FORMAL FEATURE ELIMINATION).

Further problems might be posed by such examples as the following

25 the man_i [$whose_i$ mother]$_j$ Mary talked to t_j
the man_i [pictures of who_i]$_j$ Mary saw t_j

which require a part of a "word" to enter a checking relation with the relative complementizer if pied-piping is interpreted in such a way. Since checking is obviously a syntactic operation these examples raise doubt about Uriagereka's point.

3. Preliminary Assumptions

Recall that the correct empirical generalization of section 1 seems to be that there is some ban on extracting out of a constituent that has been moved BEFORE extraction took place. That is we might propose to derive a system that predicts that constituents will act as barriers for extraction AFTER movement, but not so BEFORE being moved (relative ordering is crucial!). We will then be presented with some structural configuration as seen in (26).

26 $*\alpha...[_{\beta}...\alpha...]...[_{\gamma}...[_{\beta}...\alpha...]]$

This should be excluded on principled grounds, and we will now show that it is indeed possible to exclude (26) given Nunes' (1995) modifications of Minimalism are adopted.

The following assumptions are crucial to our analysis and thus each of them will be discussed before we present the core of the deduction.

27a Nondistinctness marking applies only with respect to the full constituent submitted to COPY, but not to its terms.
27b COPY immediately feeds MERGE, there is no copying for storage to be used later in the derivation.

There are two more crucial assumptions but these are already contained in Nunes' (1995) modifications and thus are not in need of any additional justification in the present paper.

28a MOVE as such does not exist. Rather, it is a purely epiphenomenal consequence of the displacement properties of natural language being captured by the copy relation followed by selective deletion for PF purposes (linearization).[1]
28b All instances of the same lexical item or complex constituent in a derivation are interpreted as distinct by default. Nondistinctness marking is introduced by the COPY operation, thus tracking its applications in a particular derivation.

As noted by Nunes (1995) and further developed by Uriagereka (1997), (28a) permits instances of what is generally referred to as "sideways movement" which possiblity was used to account for parasitic gaps and Across-the-Borad extraction by Nunes (1995).[2]

4. Justification of Preliminary Assumptions

Chomsky (1995) assumes that all instances of a single LI or identical complex constituents in a given derivation are interpreted as nondistinct by default. Distinctness is in turn introduced by different applications of his operation SELECT which takes a LI from the initial numeration, reduces its index by one, and introduces it into the derivation. That is, movement will automatically produce chains the members of which will be distinct but will nevertheless form a class since they were all introduced by a single application of SELECT, which took the foot or the terms of the foot of the respective chain into the derivation. Different classes of chain members will be introduced by different applications of SELECT since they are in chains footed by differently selected instances of this same LI.

In Nunes' (1995) system on the other hand, all instances of a LI or complex constituent are interpreted as distinct by default. Nondistinctness is thus introduced by application of the COPY operation only. If COPY applies to some given constituent, it will not only make a copy of its input but will also add nondistinctness markings to both the input and the output. Input and output are not otherwise taken as different, they are only marked as nondistinct but neither is marked as the copy of the other. If the input to COPY is already marked nondistinct from some other constituent, COPY will simply take over this marking to the copy of the input which will then result in a sequence of nondistinct instances of the respective constituent.

The crucial point here is to decide whether COPY marks all the terms of its input as nondistinct or nondistinctness marking applies only to the input as a whole without further reference to what it contains. This is not a trivial question as we shall see. In fact, given the modifications of the equivalent of projection in Bare Phrase Structure Theory (Chomsky 1994, 1995) laid out in Zellmayer (1997a), there may be an inherent difference between certain kinds of terms (to be concrete, the system might be able to differentiate between X° and all other terms of XP).

The COPY rule creates sets of nondistinct categories such that there are basically two ways of interpreting "nondistinctness".

29a Interpretative identity (technical)
29b Literal identity (intuitive)

Possibility (29a) is a certain technical way of treating constituents within a phrase marker whereas possibility (29b) is identity par excellance, i.e. identity with respect to whatever might be contained within the copied constituent. The latter interpretation may appear to be more in line with the traditional impression one might have of a COPY operation in requiring nondistinctness marking on all the terms of a copied item but this does not necessarily mean that it is the only way of treating COPY in grammatical theory. In what follows we will discuss conceptual and empirical arguments in favour of the purely technical interpretation (29a).[3]

A guiding idea of the Minimalist Program is that displacement is driven by immediate needs of the moved constituent (MOVE) or of some category within the local domain the moved item finally sits in (ATTRACT). Note that we formulated the attraction view rather liberally in order to allow for such cases as attraction of a head X° to a head Y° by the specifier of Y° as in Manzini & Roussou (1996) where the immediate needs of the moved head rather than those of the attractor will be satisfied or at least some joint satisfaction of the needs of both attractor and attractee is involved.[4] In any case, a checking or licensing relation (Chomsky loc.cit., Sportiche 1993) must be established between the moved item and some other category in the local domain it sits in. This domain which is traditionally labelled the "checking domain" is defined in Chomsky (1993) such as to include the head, the adjuncts to the head, the specifier, the adjuncts to the specifier, and the adjuncts to the maximal projection. The possibility of having adjuncts to specifiers is excluded by Chomsky's (1995) ban on joining to nonprojecting categories or by Kayne's (1994) ban on multiple adjunction that derives from his version of the LCA. Nevertheless, crucially there are no categories in the checking domain that are terms of the constituents in the structural positions just enumerated.[5]

Practically, this means that only full XP (or X°) can enter into checking relations if sitting in a checking configuration (see Chomsky (1995) for this distinction) but no term of XP can ever do so. The situation with X° is different due to the adjunction structure within complex heads. Compare in light of this the attempts to eliminate adjunction completely - which I sympathize with - and to shift X°-movement to PF (Chomsky 1995b).

Assume we have XP with a feature F. COPY then applies to XP creating a nondistinct copy which is then merged within the domain of some other category with a corresponding feature F', the two features entering a checking relation. But assume now, we have XP again but this time F is a feature of some YP contained in XP (dominated by XP, i.e. geometrically dominated by all the segments of XP). F of YP will then never enter into a checking relation with F' given what we have said so far about how the checking domain of some head Z° is defined (Chomsky 1993): The checking domain of Z° is the minimal residual domain of Z°. But if no checking relation is established, no chain will ever be read off and no nondistinctness marking will be necessary since this is only required by FORM CHAIN. We may thus draw the conclusion that it is not a priori necessary to take COPY to mark the terms of X as nonidentical in addition to full X itself.[6]

Given the upper considerations, we might apply some kind of Economy of inherent design to COPY. Recall that COPY is an operation that is not constitutive of a derivation as are SELECT, MERGE, or SPELL-OUT. Therefore it is assumed (Chomsky 1995, Nunes 1995) to be inherently costly whereas MERGE and similar operations would presumably be inherently costless. But if COPY is itself costly, we might expect it to be sentitive to "how much work is to be done" in each particular application - this is a rather intuitive notion but we will assume it suffices for the present purposes. If nondistinctness is necessary for a chain to be formed and feature checking is too, we would expect Economy to prevent COPY from marking anything nondistinct that does not enter into a checking relation and thus will never enter a chain formed from or within the two copies. Marking only the full input of COPY nondistinct would thus be the preferred option on Minimalist assumptions.

Another condition on FORM CHAIN seems to lend equal support to the present conclusion. Chain formation requires the members of the chain to be read off to be in a c-command relation (this is the residue of the former "movement is raising" constraint on Move α). Since no term within a copy of a moved constituent will ever c-command out of the complex object, no chain would ever be formed and thus no nondistinctness marking would be required for convergence. Thus, we would again expect COPY to keep to the more economical case of marking only the full constituent copied as nondistinct. This is again the desired conclusion.

If we adopt Kayne's (1994) view on phrase structure that specifiers are adjuncts to the maximal projections of their heads, which then uniformly project only once. But if his definition of c-command according to which crucially only categories but not segments enter into this relation is assumed, specifiers will be able to c-command out of their maximal projection, specifiers of specifiers will be able to do so too, and so on. - Admittedly, from a Minimalist point of view the definition is conceptually odd and is just needed to keep Kayne's (1994) system work. - Although he discusses empirical arguments in favour of his definition, we will just give the simplest examples in order to show that this cannot be the whole story.

30 *[[Mary]$_i$'s friend] admires herself$_i$
*[[[Mary]$_i$'s brother]'s dog] bites herself$_i$

We assume our (admittedly rather) simple counterargument to be valid, abstracting away from Kayne's (1994) evidence to the contrary.[7]

There might be also some empirical evidence in favour of the upper conclusion. Consider the following sentences.

31a who$_i$ did John speak [to t$_i$]
31b [to who]$_i$ did John speak t$_i$

As noted by Lasnik & Saito (1992), both of these examples can felicitously be answered by either (32a) or (32b).

32a Mary
32b to Mary

We might thus draw the conclusion that (31a) and (31b) are interpretatively identical, where interpretative identity basically means "LF output identity". Assume now that COPY automatically marks every term in a constituent copied as nondistinct. This entails that copying a larger input should be more costly that copying a smaller object given that such mechanisms as feature spreading and the like do not exist in the Minimalist Program. This would again entail that Economy conditions should block the more extensive option if both lead to convergence, i.e. (31a) should block (31b). This clearly is a wrong result given the judgements in (32).[8] We might thus take the assumption that COPY only marks its input but not the terms of the latter as nondistinct (if not already marked as such) to be well motivated on both conceptual and empirical grounds.

As far as the requirement for COPY to immediately feed MERGE is concerned, I do not have much to say about it here, apart from that it is crucial to the following discussion and I will just make some instructive suggestions how it could be derived on Minimalist assumptions.

There might be some requirement on the computational system to keep the array, i.e. the set of objects accessible to COPY and MERGE at any given point in the derivation, as small as possible (again relativized to any given stage) in order to reduce computational complexity. That is, there will be the less derivational options for proceeding from a given point the less objects there are in the array. This argument will only be valid if there is a true need to reduce computational complexity to a necessary minimum. Cf. Brody (1994) for arguments based on some different implementations of how to evaluate computational complexity and how to keep it small. Johnson & Lappin's (1997) critique of the Minimalist Program crucially relies on considerations of computational complexity too, at least in part.

Nevertheless, I personally do not very much sympathize with such lines of argument since syntax itself is just designed to satisfy conditions externally forced to it by interpretative systems, with which it communicates at the interfaces, in an optimal way. There is no true need to design the system in such a way that the derivational waste it generates will be kept minimal. The optimal solution will be to carry out all combinatorically possible convergent derivations and then sort out the most economical one without any stipulative restrictions on the generation process itself. Only if not restricted this way, Economy conditions will make any sense.

Similarily, application of a costly operation such as COPY might have a need of being "justified" by the following step making use of the output of the costly operation. This would then leave only a single possiblity, namely that MERGE must apply immediately after COPY since SPELL-OUT is barred from that derivational position because the rules of the phonological component would presumably be able to handle only with a single phrase marker submitted to them. A similar line of reasoning may be seen in the Last Resort condition, since this requires "movement" to be justified by a checking relation established within the output of MOVE/ATTRACT. In the latter instance, the idea (if it was truely intended) is formalized much more elegantly, since Last Resort is directly incorporated into the definition of MOVE/ATTRACT and the computational system thus has no way of circumventing it by definition.

If we did so with respect to COPY, i.e. tentatively formulate it as

33 Target α, produce a copy of α, add nondictinctness marking to α, if a is not already marked in such a way, and merge α to a category previously formed.

this amounts to reintroducing the situation with MERGE being a suboperation of MOVE/ATTRACT which was criticized by Nunes (1995). This problem - even if stated as initially formulated by Nunes - is only apparent however: Recall that there is an operation SELECT mediating the numeration and any applications of MERGE. SELECT involves an instance of COPY as its subopartation too (and I do not see any way of doing away with this). This is because if, for instance, the index of some LI is greater than one in the numeration, the LI in question must still be accessible to another application of SELECT which thus must involve copying rather than transferring (= change of location). Generalizing this to any instance of SELECT, this operation may informally be stated as (deviating from Chomsky (1995)):

34 Target α, produce a copy, put the copy into the present array, and reduce α's index by one.

No other formulation of SELECT seems to be appropriate in the system as it stands. Since SELECT cannot be reduced to anything other then (34) and even previous construction of the numeration which precedes any computational operation must involve copying items from the lexicon paired with another suboperation (that which add an index), nothing odd seems to be about having COPY being subpart of another operation. I owe the view that the relation between the lexicon and the assembly/syntax might be one of copying to a suggestion by Brody (1996).

There might be some conceptual evidence that COPY is even part of the definition of MERGE itself since some label must be designed for an object resulting from merger of two of its most immediate terms. Brody (1995) also instantiates categorial projection as an instance of the COPY relation; see furthermore Zellmayer (1997a). If we assume that an unlabelled object crashes at LF by FI, MERGE must contain some copying operation as an inherent suboperation. We might thus at least tentatively conclude that our initial assumption (27b) is conceptually supported, weakly in the worst case - i.e. if one considers the multiple instance of the copy relation in COPY, MERGE, SELECT, etc. and the occurence of the MERGE rule within COPY as redundant and therefore making the system wrong, see the arguments of Brody (1994) against chains and movement rules coexisting in a single theory (MERGE being a suboperation of COPY might be the only true problem since MERGE in turn contains a copy suboperation, i.e. we would be forced to strengthen our claim that projection involves a different kind of copying). See Zellmayer (1997b) for further discussion.[9]

5. Deriving the Ban on Extraction Out of Constituents Entering a Chain Relation

We have now justified the initial assumptions (27a) and (27b) at least to some degree and may proceed to the core of this paper. Recall the initial configuration, repeated below.

35 $*\alpha...[_{\beta}...\alpha...]...[_{\gamma}...[_{\beta}...\alpha...]]$

We will now show, that it is possible to derive from the system in Nunes (1995) and the assumptions (27a) and (27b) the fact that (35) is excluded whereas the similar configurations in (36) and (37) are not.

36 $[_{\beta}...\alpha...]...[_{\gamma}...\alpha...[_{\delta}...[_{\beta}...\alpha...]]]$
37 $\alpha...[_{\beta}...\alpha..[_{\delta}...\alpha...]]...[_{\gamma}...\alpha...[_{\beta}...\alpha...[_{\delta}...\alpha...]]]$

In (36) extraction of α out of β precedes movement of β and in (37) α has been moved inside β before extraction of α took place. We will show that both configurations are licit. That is, a moved item will act as a barrier for some particular term if this term is not moved prior to being extracted.

To derive (35) we assume the following derivation where we limit our discussion to the most relevant parts of this abstract example just for the purpose of making our proposal. Assume, α in (35) has some feature F that has to be checked by some head H° outside β for convergence reasons and assume furthermore that F of H° is strong, thus attracting α prior to SPELL-OUT.

We start by assembling the lower instance of β in order to reach

13 $[_{\beta}...\alpha...]$

We then proceed by merging β with other phrasal material which would include somd head H' which overtly attracts β to its checking domain. Depending on whether β was initially a right branch (part of main extended projection branch) or left branch rather we will have the structures (39a) or (39b), respectively.[10]

39a $[[_{\beta}...\alpha...]\ H^{\circ\prime}...[_{\beta}...\alpha...]]$
39b $[[_{\beta}...\alpha...]\ H^{\circ\prime}...[_{\beta}...\alpha...]...]$

In accordance with (35) we will continue to assume (39a) but. We wish to insist once again that the distinction between (39a) and (39b) does not affect our account.

The derivation now proceeds up to the point where H° is inserted and thus attracts α to become its specifier; we assume that α is a maximal projection or at least behaves as if it was one. Nondistinctness marking will be indicated by underlining in what follows (but only if relevant). Just a short look at the structures in (39) shows that it does not in any way matter whether α is copied from the higher or from the lower instance of β. Recall that according to the definition of COPY in (33) Minimality - either with respect to intervening potential attractees (Chomsky 1995) or potential attractors (Manzini 1996b) - is not relevant. Thus, either instance of α may be subjected to COPY, yielding either (40a) or (40b).

40a $\underline{\alpha}\ H^{\circ}...[[_{\beta}...\underline{\alpha}...]\ H^{\circ\prime}...[_{\beta}...\alpha...]]$
40b $\underline{\alpha}\ H^{\circ}...[[_{\beta}...\alpha...]\ H^{\circ\prime}...[_{\beta}...\underline{\alpha}...]]$

After application of SPELL-OUT, FORM CHAIN must apply to both the structures in (40) in order for CHAIN UNIFORMIZATION to be applicable. This latter operation is in turn required for convergence at LF since it deletes unchecked features of the "traces" of α that would otherwise induce LF crash.

Two chains could be read off from (40), CH1=(β,β) and CH2=(α,α), where we will focus the discussion to CH2 since only this is relevant. Crucially nondistinctness marking is a prerequisite for FORM CHAIN to apply and a chain is in turn the only possible input to CHAIN UNIFORMIZATION, such that nondistinctness marking will in effect be necessary for an object to converge in the relevant case. On the innocent assumption that both the lower copies of α in (40) have unchecked features that would make the structures crash at LF, two chains would be required to be formed, both with the highest instance of α as their heads. Nevertheless, since COPY could only have applied once, one copy of α will always lack nondistinctness marking. Thus one of the chains necessary for convergence could not be formed and one unchecked feature will survive till LF (since CHAIN UNIFORMIZATION can only apply to chains), inducing crash at this level of representation. We have thus shown that both the objects in (40) will crash at LF, the correct result.

We now see, that our particular interpretation of how nondistinctness marking is realized is a necessary presequisite for out account to work. Furthermore, it is clear why COPY is required to immediately feed MERGE - a situation that was captured by slightly modifying the definition of copy, since otherwise there would be no plausible way of formulating it. The scenario we wanted to exclude is the following.

COPY could have applied to α before it got fed with β such that it would have been possible to get nondistinctness marking onto both the embedded copies of α. The relevant steps of this alternative derivation are the following:

41 MERGE β (including α)
COPY α (thereby marking α nondistinct)
COPY β (thereby copying everything β contains, including nondistinctness marking on α)
MERGE β (in order to check H°')
MERGE α (in order to check H°)

Then, the relevant chains can be formed and the derivation is wrongly predicted to converge. Note that the derivation sketched in (41) has the same effect as a noncyclic derivation which involves covert movement of β would have had although it was assembled in a cyclic manner. We will show below, that covert movement involving noncyclicity indeed has the effect of circumventing the LF crash deduced in the preceding paragraph. The definition of COPY in (33) thus readily excludes the derivation in (41).[11]

6. Consequences of the Derivational Design of the Model

We are now in a position to show that interaderivational temporal ordering (in its most literal sense) of applications of the COPY operation crucially affects the output. The standard case that comes to mind is covert movement, i.e. movement that took place after application of SPELL-OUT. At first glance, one might suspect that the possibility of extraction out of covertly moved constituents depends on how exactly "covert movement" is implemented in the theory. Needless to say, from the upper discussion it should be clear that this is not the case. Our account is not affected by the particular way we implement covert movement, where we basically have three possibilities.

(i) We might assume that covertly moved constituents do not really move, i.e. they occupy their surface positions at LF. Although Brody (1994) for instance made pretty much effort to show that this cannot be the case since overt and covert movement behave alike with respect to island phenomena, one could still adopt some analysis involving empty operator movement as sketched out by Watanabe

(1991), at least in cases of A-bar-movement.[12] Such a way of implementing covert movement is explored in Zellmayer (1997a). Compare for instance, Chomsky's (1993) proposal that wh-in-situ does not move covertly to SpecCP. Clearly, nothing needs to be said with respect to the present proposal in connection to the in-situ view of covert movement since the problems discussed in section 3.3 will never arise in that case.

(ii) Alternatively, we might take the opposite view and claim that covert movement is completely alike to overt movement, i.e. both the head and the tail of the chain formed will be occupied by a copy of the moved constituent. This view is taken in Chomsky (1993) in the case of A-movement. Covert movement, if understood in this way, is necessarily countercyclic. On the assumption, that extraction is overt, the tail of the chain formed from the extracted item will be within the covertly moved constituent. Nevertheless, since overtness of extraction entails that COPY applied to the extracted item BEFORE covert movement takes place, nondistinctness marking will necessarily be present on the copy of the extracted item within the covertly moved constituent. Thus, if FORM CHAIN and CHAIN UNIFORMIZATION apply for some reason after covert movement, both the relevant chains can be read off and the structure will converge.

42a	$\underline{\alpha}...[_{\gamma}...[_{\beta}...\underline{\alpha}...]]$	overt structure
42b	$\underline{\alpha}...[_{\gamma}\,[_{\beta}...\underline{\alpha}...]...[_{\beta}...\underline{\alpha}...]]$	covert structure
42c	$[_{\beta}...\underline{\alpha}...]...\underline{\alpha}...[_{\gamma}...[_{\beta}...\underline{\alpha}...]]$	covert structure

This is obviously the desired result. Nothing will change, if β in (42) moves across α in (42a).

Problems could arise, if both extraction and raising of the extraction domain are covert as in (40) discussed above. Nevertheless, these problems are not substantial, since nothing forces the ordering of COPY operations where movement of the extraction domain precedes extraction. Since the targets of movement must be "there" before covert movement in any case, ordering could well be reversed since FORM CHAIN would not be affected.

There is another case to consider. Raising of the extraction domain may be overt whereas extraction itself may be covert, i.e. a case where the impossible ordering is forced. This situation can be treated in exactly the same fashion as the examples where both movements are overt and we thus predict it to be excluded. There is empirical evidence that shows that this prediction is true (note that contrary to what was argued for in Boskovic (1995) we assume that Accusative checking in English ECMed DPs is covert).

43 *Mary's prophecy was that John would expect Bill to win the race and [Bill to win the race] John expected t

We thus may conclude, that treating covert movement in the way Chomsky (1993) does, has no effect on our proposal. This far, the present paper gives indirect evidence that the temporal ordering of applications of particular operations may be relevant to determining the admissibility of a certain derivation and thus argues in favour of a derivational view of grammatical design.

Finally, (iii) nothing essential changes if the feature movement story (Chomsky 1993), a derivational adaption of the empty expletive (scope marker) account by Brody (1994), or some treatment involving empty operator movement (Watanabe 1991) of covert movement is adopted, since then the problem of extracting out of constituents entering a chain relation would never arise in this case.[13]

7. Circumventing Chain Islands

It is well known that configurations as Nominative Islands or Strong Islands in general can be circumvented by parasitic gap. It is the purpose of this section to show how this fact can be captured in the present account. Of course, the initial intuition is that given what has been said up to now there must be some way of getting the nondistinctness marking on both copies of the extracted item. As we did throughout this paper this we will crucially rely on the ordering of applications of the COPY operation without violating the ban on copying for storage incorporated to the COPY rule.

Before proceeding note that the situation of a configuration acting as a barrier for extraction allowing this effect to be neutralized by multiple chain formation strongly argues against a representational view of grammar which has no immediately plausible account for this fact. That is because representational models are inherently stative and thus the configurations that induce extraction domains to act as barriers will look alike both if extraction is only out of the barrier and if multiple extraction is involved (as in parasitic gaps). We may thus take the bare existence of such constructions as an argument in favour of a derivational model of grammar which crucially allows us to make recurse to the ordering of operations as we did in this paper.[14]

Given what has been said above it is fairly clear how the ameliorating effect of parasitic gaps can be handled in the present theory. In doing so we will adopt the treatment of Nunes (1995) which involves sideways movement - i.e. copying a particular constituent and merging it to an item that is not a term of the tree it was contained in before copying - followed by chain formation. The basic structure, which is abstracted from a Nominative Island configuration, is thus the following.

44 $\alpha ... [[_{\beta} ... \alpha ...] ... [[_{\beta} ... \alpha ...] ... [_{\gamma} ... \alpha ...](...)]$

In (44) it is irrelevant whether the constituent γ is a right branch or sits in some specifier position of the extended projection (Grimshaw 1991) of the main branch. Similarily, γ may well be α itself, as would be the case in a standard Nominative Island case. No c-command condition is necessary, what is relevant is indeed just the fact that there is some instance of α that could be copied into the moved extraction domain β just as its lowest instance - to be more accurate: the first instance of β assembled - is built. By copying β the nondistinctness marker on α

resulting from prior application of COPY to α in the course of assembling β would the automatically be taken over to any copy of β made in the derivation.

The derivation giving (44) looks as follows. Assume, as we did in section 3.3, we have two functional heads, each with a strong feature where H° attracts α and H°' attracts β. We then start by assembling γ.

45 $[_{\gamma}...\alpha...]$

The derivation proceeds up to the point where β is inserted. We could equally well have started assembling β immediately after having reached γ but nothing essential hinges on this distinction. Assume, we adopt the first possibility which then gives us β assembled by COPY applying to α within γ and merging the resulting copy with other material leading to

46 $[[_{\beta}...\underline{\alpha}...]...[_{\gamma}...\underline{\alpha}...](...)]$

from which point the derivation can proceed up to where the head H°' which attracts β is inserted. Note, that crucially we have nondistinctness marking on both the copies of α, i.e. also in the copy contained in the extraction domain β.

47 $H^{\circ\prime}...[[_{\beta}...\underline{\alpha}...]...[_{\gamma}...\underline{\alpha}...]...]$

COPY now applies to β copying everything β contains including the nondistinctness marker on α. Recalling the discussion from section 3.3 we immediately note that this is all that is needed to circumvent the "chain island" effect since crucially FORM CHAIN can now apply to all the copies of α as is required for the unchecked features of the lower instances of α to be eliminated in order for FI to be satisfied at LF.

48 $[_{\beta}...\underline{\alpha}...]\ H^{\circ\prime}...[[_{\beta}...\underline{\alpha}...]...[_{\gamma}...\underline{\alpha}...](...)]$

The derivation now proceeds up to the point of inserting H°, makes another copy of α, and inserts this to SpecHP thereby checking the strong features of H°.

49 $\underline{\alpha}\ H^{\circ}...[[_{\beta}...\underline{\alpha}...]\ H^{\circ\prime}...[[_{\beta}...\underline{\alpha}...]...[_{\gamma}...\underline{\alpha}...]...]$

FORM CHAIN can apply to (49) and read off three chains each of which is subjected to CHAIN UNIFORMIZATION leading to deletion of the unchecked features of each chain allowing the structure to converge at LF (if everything else runs fine).

We thus have reached a principled account of movement islands, including A-bar-Islands and that class of Strong Islands which can be circumvented by what Engdahl (1983) termed "obligatory parasitic gaps", i.e. those gaps that do not allow replacement by a pronoun.[15] Ordering of applications is again crucial to our proposal and given the problems this constructions rise for a representational approach (as pointed out above) we might add another argument for a derivational approach to the theory of grammar. Having completed our account we will now turn to discussion of a number of true and apparent counterexamples.

8. Some Apparent Counterexamples

First of all given what has been said so far about how chain islands are established - by the lack of nondistinctness marking on one of the copies of the extracted item and thus failure to form a chain and eliminate yet unchecked features - one might think of a possiblity to circumvent these islands without any need of multiple extraction. This could be done by independently assembling both instances of the extraction domain. Of course this is much less economical than simply copying the extraction domain but if leading to convergence it will be tolerated by the system. It is easy to show that this apparent escape hatch leads to crash in any case both at LF (unchecked features) and PF (failure of linearization).

Suppose, we have a configuration such as the one in (26) and (30), repeated here as (50).

50 $^*\alpha\ldots[_\beta\ldots\alpha\ldots]\ldots[_\gamma\ldots[_\beta\ldots\alpha\ldots]]$

Assume for ease of illustration, that β has hierarchically structured terms X, Y, and α such that (50) can therefore be explicated as (51).

51 $^*\alpha\ldots[_\beta\ X\ [Y\ \alpha]]\ldots[_\gamma\ldots[_\beta\ X\ [Y\ \alpha]]]$

As in sections 3.3 and 3.5 above we have two heads H° and H°' each of which is equipped with strong a feature where the former attracts α and the latter does so with respect to β. The derivation starts by assembling α and proceeds up to where H°' is inserted.

52 $H^{\circ\prime}\ldots[_\beta\ X\ [Y\ \alpha]]$

Instead of copying β as a whole we copy each of its terms and independently assemble the instance of β required by H°'.[16]

Having fully assembled the upper instance of β we merge it to the structure in (52).

53 $[[_\beta\ \underline{X}\ [\underline{Y}\ \underline{\alpha}]]\ H^{\circ\prime}\ldots[_\beta\ \underline{X}\ [\underline{Y}\ \underline{\alpha}]]]$

Note that we now have nondistinctness marking on both copies of α, as desired. The derivation proceeds up to the point where the equivalent of (49) is reached.

54 $\underline{\alpha}\ H^{\circ}\ldots[[_\beta\ \underline{X}\ [\underline{Y}\ \underline{\alpha}]]\ H^{\circ\prime}\ldots[_\beta\ \underline{X}\ [\underline{Y}\ \underline{\alpha}]]]$

Two chains involving α can now be read off and (29) should be expected to converge. Nevertheless it doesn't. Assume, H°' is T° or C° and β is either a nonoperator DP or a WH, rleativized, or topicalized DP (the standard case). Then, β will have to check some feature against H° as will generally be the case with movement induced by strong reatures with the exception of subject raising at least to SpecTP of ECMed infinitival subjects. All of β's terms are marked nondistinct from their counterparts in the other instance of β respectively, but the two instances of β themselves are nevertheless interpreted as distinct since they do not result from application of COPY. Therefore, no chain can be constructed and the Case feature of the lower instance of β cannot be deleted by CHAIN UNIFORMIZATION causing (54) to crash at LF.

Consider now the linearization of the two instances of β in (54). They are nondistinct, so they can be linearized since the upper instance of β asymmetrically c-commands the lower one. Nevertheless, each instance of β contains terms that are marked nondistinct independently of β itself. Thus, by the induction step of the LCA (cf. Nunes (1995) and Uriagereka (1997)) that reads as

55 If α asymmetrically c-commands β and α has a term γ then γ precedes β (my wording).

α in the upper copy of β should preced H°' by (55) since the upper copy of β asymmetrically c-commands H°'. By the same principle, α in the lower copy of β should be preceded by H°' since H°' asymmetrically c-commands this copy of α. Thus, α should both precede and be preceded by H°' since it is marked nondistinct and thus interpreted as a single category by PF. But this violates the irreflexivity (a_β → NOT (β_α)) and totality (EITHER α_β OR β_α) requirements on a well-formed linear ordering and the structure (54) therefore crashes at PF. The same is true of terms X and Y of β.

Since neither of the nondistinct terms in the upper instance of β c-commands its copy in the lower one, no chains can be read off and CHAIN REDUCTION cannot apply in order for LINEARIZE to give a well-formed output. Thus, this objection is readily excluded.

Second, given the account of circumventing "chain islands" that was sketched in section 6 we would expect all types of "movement islands" to allow being circumvented by multiple extraction constructions which prediction is obviously wrong.

56 *who$_i$ did Sue wonder [[which pictures of t$_i$]$_j$ Sam liked t$_j$]
57 *who$_i$ did Sue wonder [[which pictures of t$_i$]$_j$ Sam gave t$_j$ t$_i$]

(56) is correctly excluded, whereas (57) is predicted to be okay. The string in (57) can be parsed in two different ways and we will show in what follows that both are independently excluded.

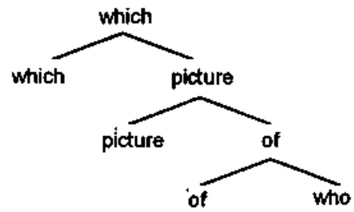
58a who did Sue wonder [which pictures of who] Sam gave who [which pictures of who]
58b who did Sue wonder [which pictures of who] Sam gave [which pictures of who] who

Since (58b) is obviously an unlikely result of parsing for semantic reasons, we will put it aside for the moment. Let us therefore focus on (58a) first.

By the Minimal Link Condition part of Nunes' (1995) definition of FORM CHAIN no chain can be read off from <u>who</u> in SpecCP and any other copy of <u>who</u> lower down in the tree. This is because in any case there is a c-commanding intervener - the head of <u>which pictures of who</u> or the whole DP - that is closer to <u>who</u> in SpecCP than any other copy of <u>who</u>. As far as the higher copy of <u>which pictures of</u>

who is concerned, the head of this latter phrase induces MLC effects since it c-commands who in complement position.

59



The Bare Phrase Structure (Chomsky 1994) representation of the complex WH illustrates this although only informally. Since the label if this whole phrase is identical to its head, the head must have the same feature constitution as the label and thus must be able to induce MLC effects whether or not its WH feature has been rendered invisible at LF by the checking operation. Any other chain involving who-in SpecCP will be blocked by the whole object in (59). One might object that this interpretation of the MLC should exclude sentences such as

60 what$_i$ did John wonder [why$_j$ Mary bought t$_i$ t$_j$]

which unfortunately are often enough wellformed, a fact that is seriously problematic for both Chomsky (1995) and Nunes (1995). There is nevertheless an alternative way of excluding (58).

No chain can be read off from the lower instance of which pictures of who in the direct object position and the copy in the embedded SpecCP since a copy of who intervenes. Thus, the WH features of the lower copy of which pictures of who cannot be deleted by CHAIN UNIFORMIZATION and the derivation crashes at LF. Similarily, even if such a chain was formed, the WH-feature of who in the lower copy of the complex WH could never be deleted. This is because FORM CHAIN cannot form a chain from this copy and who in the indirect object position because no checking with respect to the WH-feature takes place there and because the (undoubtedly) unchecked and undeleted WH-feature of the complex WH-phrase intervenes. There are more combinatorial possiblities but all are excluded in such a way that (58a) will always induce a crash at LF.[17]

But if (58a) is impossible the parser should generate structure (58b), repeated below, if presented with the string in (57).

61 who did Sue wonder [which pictures of who] Sam gave [which pictures of who] who

This is readily barred since chain formation involving the most deeply embedded copy of who is always excluded by the intervening lower copy of which pictures of who. No other chains involving the lowest copy of who can be read off for the lack of c-command and WH-feature checking. The data in (56) and (57) are thus in line with what one would expect.

9. Some True Counterexamples

The first and perhaps most devasting counterexample of all that might come to mind is that my account fails to capture the sensitivity of parasitic gaps to sitting within islands, i.e. a parasitic gap will not ameliorate an island violation if it is itself embedded to an island. This is captured by the Connectedness account of Kayne (1983) and is used as a diagnostics for empty operator movement in these structures by Chomsky (1986).

62 a book$_i$ that [anyone who finds [the first chapter of t$_i$]] usually ends up disliking t$_i$
*a book$_i$ that [anyone who expects [copies of t$_i$] to be on sale] usually ends up disliking t$_i$
*a book$_i$ that [anyone who left the room [after reading t$_i$]] usually ends up disliking t$_i$

Nevertheless these examples seem almost always to involve Adjunct Islands and Complex NP Islands which we did not treat in this paper. Therefore it may well be that these involve and additional mechanism specifically liked to these constructions.

Second, my account predicts that if an extracte item is moved within a moved extraction domain, this domain should not act as a barrier despite its being moved. To be concrete, I incorrectly predict ECMed infinitival subjects and idirect objects to be allowed to escape from Sentential Subject Islands; the case of ordinary subjects is different and probably involved Complementizer-Trace Effects.

63 *who$_i$ did [for John to expect t$_i$ to have kissed Mary] bother Bill
*who$_i$ did [for John to give t$_i$ pictures of Mary] bother Sue

I do not have anything to say with respect to these examples, they provide true counterexamples to my claim.

I have mentioned these firm counterexamples and leave it to the reader to judge whether my proposal is worth considering further. Despite the problems it may well be a step towards extending the Minimalist Program and approaching the explanatory power of its precursor, the Principles and Parameters Theory.

10. Conclusion

In this paper I argued that certain Island Constraints receive a natural explanation of certain modifications of the Minimalist Program (Chomsky (1995) as laid out in Nunes (1995) are adopted. The strict interpretation of how the COPY rule adds nondistinctness marking to its input easily derives from the overall architecture of the system. Furthermore I showed that the deduction becomes possible by crucially relying on the relative ordering of applications of COPY. This is a prototypical mechanism of manipulation and explanation available in a derivational system. Despite many problems remain and the data are handled in a much more elegant way in Manzini (1996a) for instance my account may be a first step towards a minimal solution.

11. Notes

1 In Nunes (1995), the displacement effect results from the interaction of six conceptually distinct operations (i.e. COPY, MERGE, FORM CHAIN, CHAIN UNIFORMIZATION, CHAIN REDUCTION, and FF-DELETION).

2 Note that although we will adopt this anlysis here, we do not sympathize with the idea of assmimilating parasitic gaps to ATB extraction as was proposed by Williams (1990) for instance since as pointed out by Postal (1992), there are good reasons not to treat these two phenomena as the same. Nevertheless, we believe that some kind of symmetry vs. asymmetry may be relevant in both constructions. As far as (28b) is concerned, the system deviates from what was assumed in Chomsky (1995).

3 Notice that Nunes (1995) makes initial use of "nondistinctness" in the conditioning part of his operation FORM CHAIN which crucially applies to nondistinct constituents only. Nevertheless, FORM CHAIN also incorporates a reformulation of Chomsky's (1995) Last Resort - a revision of the former local economy principle Greed (Chomsky 1993) that was incorporated into the definition of MOVE and later in the text reformulated in terms of ATTRACT in Chomsky (1995). It seems, that the Last Resort part of FORM CHAIN directly contradicts the literal interpretation in (29b) above: Last Resort requires that if a chain is to be read off from α and β, α has to have a feature F checked by some γ with β having a corresponding unchecked feature F' that could potentially have entered a checking relation with γ. That is, α and β will be nonidentical with respect to feature constitution since otherwise no chain could be formed. Since, however, nondistinctness is another necessary precondition for chain formation, it must be case that nondistinctness is understood in its technical interpretation. This fact is noted by Nunes (1995) since he remarks that the notion of "sameness" used by FORM CHAIN must be his nondistinctness (= our interpretation (29a)) rather than "identity" (= our interpretation (29b)). This argument from feature constitution presents us with initial motivation for preferently adopting (29a).

4 The trigger for discplacement may also be a mixture of needs of both moved item and target (Chomsky 1993) or may result from duplication of features and a FI driven need to eliminate such redundancies (Brody 1995). As pointed out by Prinzhorn (class, 1997), the former view suffers from the problem of redundancy and should thus be avoided. Reducing all movement to ATTRACT faces problems concerning the location of the trigger for movement, see Zellmayer (1997b).

5 Since containment may also be defined in terms of domination rather than "being a term of" (domination was actually used in Chomsky 1993), specifiers of SpecXP would indeed be included in the checking domain of X° if specifiers are treated as adjuncts in Kayne (1994); let us put this aside.

6 Note that this implies a particular view on where features are located in a complex category - on the maximal projection of the category in question. If it was the head, no checking relation could ever take place and we would only expect head movement but no phrasal movement. It is not of relevance to the present paper how this view on the launching site of the features is implemented but see Zellmayer (1997a) for some discussion.

7 On the assumption that Reinhart & Reuland's (1993) theory of reflexivity is correct, the examples in (30) will be readily excluded since the predicates in question are reflexive-marked although no two arguments of them are coreferent, since specifiers to arguments of a predicate P are not coarguments of other arguments of P. Nevertheless, there are empirical problems with this theory and we are not in need of immediately incorporating it into the Minimalist Program just because it has been proposed.

8 Note that preposition stranding seems to be a rather marked option not possible at all in many languages and frequently replaced by pied-piping. This runs against what one would expect.

9 This section provides an argument against rejecting having two operations A and B where A is a suboperation of B and A and B differ in derivational cost. It is easy to see that in Chomsky (1995), SELECT, MERGE, MOVE/ATTRACT, and SPELL-OUT all contain the derivationally costly suboperation COPY although they differ in derivational cost. It thus seems reasonable to decompose these macrooperations into just two microoperations <u>COPY</u> and <u>MERGE</u> and show that individual

subcomponents may be responsible for (alledged) differences in derivational cost. See Zellmayer (1997b); compare also Brody (1994, 1995) on his COPY suboperation which is involved in his CHAIN and PROJECT procedures.

10 As will become clear as we proceed this does not affect our account in any relevant way, i.e. the often used left vs. right branch distinction will not be relevant to our account which in this respect differs from that of Kayne (1983) and its Minimalist translation by Manzini (1996a). Similarily it differs from that of Uriagereka (1997) although this author does not make use of Kayne's (1983) initial ideas.

11 Note, that if we had used an Economy condition instead of (33), we could not have excluded (41), since only convergent derivations are allowed to block other derivations and thus, the derivation leading to the illicit structures in (40) could never have blocked (41) since it does not converge.

12 Brody (1994) himself admits that there is no a priori reason to prefer his primary/secondary-chain approach to Watanabe's (1991) empty operator movement approach. This is because the latter proposal assimilates covert movement to overt movement in such a way as to capture the similarity with respect to island sensitivity of these constructions that Brody (1994) intended to show. He furthermore argues that a distinction primary vs. secondary is superior to an overt vs. covert distinction where he claims that secondary chains are parasitic altough he does not explain the parasitic nature of secondary chains in a principled way comparable to Kayne's (1983) approach. The no-movement view easily captures the nonsensitivity to single islands but may have difficulties in treating multiple islands (cf. Longobardi (1991)).

13 Neither a bundle of formal features of the covertly moved item, nor a phonologically empty expletive, or an empty operator will contain any term of the item it is extracted out (Chomsky 1995), corefers with (Brody 1995), or is base adjoined to (Watanabe 1993).

14 Needless to say such an account is conceptually superior only if ordering is intrinsically determined by output conditions rather than extrinsically stipulated as was the usual practice in the early days of generative grammar.

15 As Engdahl (1983) herself notes, the obligatoriness (ban on lexical resumptives) may be due to the WCO effect. This argues against a null resumptive pronoun analysis of parasitic gaps.

16 Note that this requires starting with Y and α which in turn is only possible if both Y and α are copied before being merged. But this entails that there must be at least a single application of COPY that does not immediately feed MERGE. Thus, the derivation should be readily excluded. Assume this exceptional situation is allowed by somehow fooling the system and see how the derivation can nevertheless be excluded on independent grounds.

17 In a much more trivial way, (58a) could be excluded since it involved questioning of an indirect object which is otherwise excluded.

(i) *who$_i$ did John give t_i books

I nevertheless do not agree with this, since in multiple WH instances such extractions seem quite acceptable.

(ii) who$_i$ did John give t_i what

However things might be (58a) is always excluded on rather principled grounds.

12. References

Branigan (1992): Subjects and Complementizers. Ph.D. Thesis, MIT.
Boskovic (1995): Principles of Economy and the Syntax of Nonfinite Complementation. Ph.D. Thesis, UConnecticut.
Brody (1994): Lexicological Form. A Radically Minimalist Theory. MIT Press.
Brody (1995): Towards Perfect Syntax. Working Papers in the Theory of Grammar 2.
Brody (1996): Recent Developments in the Theory of Grammar. Notes of class lecture, UVienna.
Cinque (1990): Types of A'-Dependencies. MIT Press.
Chomsky (1986): Barriers. MIT Press.
Chomsky (1993): A Minimalist Program for Linguistic Theory. In Hale & Keyser (1993).
Chomsky (1994): Bare Phrase Structure. In Webelhuth (1995).
Chomsky (1995a): Categories and Transformations. In Chomsky (1995c).
Chomsky (1995b): The MIT Autumn Lectures. Notes of Class Lecture, MIT.
Chomsky (1995c): The Minimalist Program. MIT Press.
Engdahl (1983): Parasitic Gaps. Linguistics and Philosophy 6.
Grimshaw (1991): Extended Projection. Manuscript, Brandeis University.
Hale & Keyser (1993): The View from Building 20: Essays in Linguistics in Honour of Sylvain Bromberger. MIT Press.
Hornstein (1995): Logical Form. Blackwell.
Johnson & Lappin (1997): A Critique of the Minimalist Program. Natural Language and Linguistic Theory 20.
Kayne (1983): Connectedness. Linguistic Inquiry 15.
Kayne (1994): The Antisymmetry of Syntax. MIT Press.
Koizumi (1995): Phrase Structure in Minimalist Syntax. Ph.D. Thesis, MIT.
Lasnik & Saito (1992): Move α: Conditions On Its Application and Output. MIT Press.
Manzini (1996a): Dependencies, Phrase Structure, and Extractions. Ms.
Manzini (1996b): A Minimalist Theory of Weak Islands. Ms.
Manzini & Roussou (1996): A Minimalist Theory of Control. Manuscript, UCLA.
Nunes (1995): The Copy Theory of Movement and the Linearization of Chains in the Minimalist Program. Ph.D. Thesis, UMaryland.
Postal (1992): Parasitic Gaps and the Across-the-Board Phenomenon. Linguistic Inquiry 24.
Reinhart & Reuland (1993): Reflexivity. Linguistic Inquiry 25.
Rochemont (1997): Phonological Focus and Structural Focus. Ms., UBC.
Sportiche (1988): A Theory of Floating Quantifiers and Ist Corrolaries on Constituent Structure. Linguistic Inquiry 20.
Sportiche (1993): A Reductionist Approach to Syntactic Variation and Dependencies. Ms., UCLA.
Uriagereka (1997): Multiple Spell-Out. Groninger Arbeiten zur Germanistischen Linguistik 40.
Watanabe (1991): WH-In-Situ, Subjacency, and Chain Formation. Ms., MIT.
Webelhuth (1995): Government and Binding Theory and the Minimalist Program. Blackwell.
Williams (1990): The ATB Theory of Parasitic Gaps. The Linguistic Review 6.
Zellmayer (1997a): Maximizing Minimalism. Ms., UVienna.
Zellmayer (1997b): Review of Nunes (1995). Ms. in Prep., UVienna.

Author´s adress:

Florian Zellmayer
Department of Linguistics
General Linguistics
University of Vienna
Berggasse 11/3/3
1090 Vienna
Austria