On the Role of Resumptive Pronouns in Amnestying
Island Constraint Violations
Anthony S. Kroch
University of Pennsylvania

1. In a recent article, Kayne (1981) remarks that the relative clauses in (1) and (2) are grammatical in colloquial English\(^1\) while those in (3) and (4) are just as impossible in the colloquial language as they are in the written standard:

(1) [(Kayne's (83a))] the guy who they don't know whether he wants to come or not
(2) [(83b)] the guy who I hate almost everything he does
(3) [(84a)] *the guy who they don't know whether wants to come or not
(4) [(84b)] *the guy who I hate almost everything does

Examples (3) and (4) can, of course, be ruled out by any one of a number of formulations of the COMP-trace effect (e.g., Chomsky and Lasnik 1977, Bresnan 1977, Chomsky 1980). Kayne's version is the Empty Category Principle (ECP)\(^2\), a constraint on the proper binding of traces recently proposed by Chomsky (forthcoming); and he naturally wants the grammaticality of (1) and (2) to follow from the fact that the substitution of a pronoun for a trace in these sentences makes the ECP inapplicable to them. To achieve this result, he suggests that there is a pronoun insertion transformation applying after wh- movement to produce (1) and (2) from the structure underlying (3) and (4), thereby eliminating the improperly bound empty category. The difference between the colloquial language and the standard is then simply that the standard language does not contain the pronoun insertion rule.

Kayne's analysis is immediately attractive and not the least of its attractions is that it accounts for the contrast between (1)/(2) and (3)/(4) without modification of grammatical theory. Unfortunately, detailed examination of Kayne's proposal, which he sketches only briefly in his paper, shows it to be untenable. What we hope to demonstrate in these remarks is that when one tries to work out the details, Kayne's insertion rule cannot be stated within the theory that he is defending. Furthermore, consideration of a slightly wider range of cases shows the rule to be inadequate empirically; and we shall find that accounting for (1) and (2) in the grammar requires abandoning the movement analysis of wh- in favor of an analysis in which wh- words and phrases are base-generated in COMP. In addition, however, one serious problem of the movement analysis — the ad hoc character of the constraints needed to block the overgeneration of resumptive pronouns in environments where they cannot occur — will turn out to be shared by any base-generated account. Since this problem appears to have no elegant formal solution, we shall be led in the end to reconsider the grammatical status of (1)/(2), whose acceptability is, of course, more problematic than Kayne's
straight-forward grammaticality judgment suggests. Indeed, although forms like (1) and (2) occur in speech with some frequency (Kroch forthcoming), their users are generally reluctant to judge them acceptable. We shall propose, therefore, in place of a grammatical account, an explanation for the relative acceptability of (1)/(2) and for the occurrence of such forms in speech which derives from a recently proposed model (MacDonald 1980) of how the grammar of wh- dependencies is incorporated into the real-time sentence generator. Our claim will be that the generator produces forms like (1) and (2) even though the grammar does not allow them and that it does so when the speaker embarks on the utterance of a relative clause without realizing that the required gap will be in an impermissible position.

2. The first and most obvious problem for Kayne’s Resumptive Pronoun Insertion rule (RPI) is the one that it will turn out to share with all grammar based approaches; namely, that the rule must be constrained to apply only where the ECP would otherwise be violated later. If not, pronoun insertion will replace traces with pronouns in all relative clauses and thereby produce examples like (5) and (6), which are no more acceptable in colloquial English than they would be in the standard.

(5) *She’s the only woman here that I know her very well.
(6) *The student that he got the highest score solved the hardest problem.

The most obvious way to avoid generating (5) and (6) is to constrain the pronoun insertion directly by having it apply only if the trace it replaces is not properly bound. If we do this, however, then we are stating the ECP in a different place in the grammar of the colloquial language than in the grammar of the standard language. In the colloquial language we have a pronoun insertion transformation which is constrained by the ECP while in the standard language the same principle is a constraint on logical form (LF). This is clearly an impossible result if we take seriously the autonomy of the various components of the grammar and the explanatory power of the ECP. What we would be left with is an unexplained conspiracy among rules of different parts of the grammar in different dialects to avoid the generation of certain forms. This result could be avoided only if we generalized RPI to the standard dialect and then added to that grammar a filter excluding all wh- constructions in which the wh-word failed to bind a trace in the deep structure position where it originated. Of course, such a filter would not be local, given the unbounded distance between the wh-word and its source position, and so writing it would increase the power of the grammar in an unwanted way. Also problematic is the fact that making the ECP a constraint on the RPI rule means that it is no longer a constraint on binding, at the least an unnatural result.
A second problem with Kayne’s RPI rule is that it would have to be stipulated to be obligatory, and this stipulation cannot be made in Kayne’s framework (the Revised Extended Standard Theory as formulated in Chomsky and Lasnik (1977) and later work) since all transformations are there taken to be optional and the effects of obligatory rules are obtained through the use of filters on surface structure or LF. The RPI rule, of course, cannot be optional because if it were, we would have to state the ECP twice in the grammar of the colloquial language, once negatively as a constraint against RPI applying too widely and again as a constraint on LF to rule out the cases where the transformation was needed but had failed to apply.

A further problem for RPI is posed by the following data:

(7) There was one guy who I didn’t think would come.
(8) There was one guy who I didn’t think that he would come.
(9) *There was one guy who I didn’t think he would come.
(10) *There was one guy who I didn’t think that would come.

The same colloquial language that allows (1) and (2) also allows (8) as a variant of (7) while excluding (10) and also (9). These facts cannot be captured if (8) is generated by a pronoun insertion rule. If we assume deletion in COMP is responsible for sentences in which the object complement clause is not introduced by that, then pronoun insertion must apply after that deletion. Otherwise we will generate (9) just in case pronoun insertion applies and then the deletion takes place. Worse, we will have no way of generating (7), the preferred variant. This follows because pronoun insertion must be obligatory, as demonstrated above. This ordering is, however, impossible since RPI is a rule of the syntax and deletion in COMP is a rule of the phonology, the input to which is the output of the syntax (see Chomsky and Lasnik (1977) and later work).

3. Given the problems posed by RPI, we must look elsewhere for an explanation of the acceptability of (1)/(2). One possibility, which would involve the minimal change from Kayne’s proposal, would be to replace RPI with a morphological rule that spelled out trace as a pronoun in the appropriate environment. If we consider pronouns to be bundles of syntactic features, then they differ from traces minimally, that is in not being anaphors and in having phonetic content. They share all other syntactic features — case, number and gender. That traces have number and gender is shown by examples like (11):

(11)a. The men, who, we thought t would shave themselves, didn’t bother.
b. The woman, who we thought she would cut
herself, *himself, with the paring
knife didn't.

If the opacity constraints (PIC, SSC) on anaphors, treated as
constraints on LF, are to account for (11), as in Chomsky (1980)
and elsewhere, we must assume that trace serves as the antecedent
to the reflexives in (11). Then if the gender and number features
of reflexives are selected (either directly or through a filtering
mechanism) to match those of their antecedents, as seems
necessary, trace must be marked for those features, perhaps by
some rule that passes them down from the head of the relative to
the relative pronoun and another rule that guarantees that traces
will receive the features of their antecedents. The same point
can be made for the number feature by subject–verb agreement, as
in (12):

(12) The men, who you thought they were was
responsible

Number-agreement is clearly a local phenomenon but it can only be
stated as such if the trace in (12) is marked for number. The
only alternative to marking trace for syntactic features would be
to have number-agreement apply before wh-movement and to have wh-
words marked in the base for their syntactic features. Then,
after wh-movement, if the features on the head differed from
those on the wh-word, the relative clause would be stranded by a
filter. This approach would force us to treat both S and S’ as
cyclic nodes to avoid the extrinsic ordering of wh-movement after
subject–verb agreement. More seriously, it would force us to
abandon the surface filter approach to reflexivization in order to
guarantee that the reflexive pronoun will agree in its features
with the wh-word. Since the wh-word is in the proper binding
position with respect to the reflexive only before wh-movement,
the reflexive would have to receive its syntactic features either
through a deep structure filter or a transformational copying
process ordered before wh-movement.

If we assume that traces contain all of the relevant
syntactic information needed for selecting the proper resumptive
pronoun in cases like (1)/(2), then we could write a simple
feature changing rule to transform trace into a pronoun in the
appropriate environment. Such a Morphological Resumptive Pronoun
Realization (MRPR) rule would belong not to the syntax but to the
phonology of a REST grammar. Therefore, the impossible ordering
forced by NPI in handling (7)–(10) would no longer exist, since
both deletion in COMP and MRPR would now be in the same component.
Of course, it would still be necessary to order deletion before
MRPR if we wish to avoid generating (9). If we were to decide
that (9) is acceptable in the colloquial language, however, (since
such forms sometimes do occur), we could eliminate this ordering
stipulation. The difficulty with the morphological rule, of
course, is that it is equivalent to the transformational one in the crucial feature that the ECP would have to be a feature on it rather than on LF. Again we are faced with the unacceptable choices outlined above.

4. At this point let us introduce a new bit of data which simply eliminates both RPI and MRPR as candidates for the explanation of (1)/(2). The colloquial language contains ECP violating relative clauses not only with resumptive personal pronouns, but also with the demonstrative that and full (non-pronominal) noun phrases in the trace position. Thus, we find in speech examples like those of (13), which are as acceptable as (1)/(2):

(13)a. There was one prisoner that we didn't understand why the guy was even in jail.
   b. It came up in the charge to the jury, which we have been discussing what that meant.

These cases cannot be generated either by an insertion rule or by a morphological spell-out rule since the noun phrases that stand in place of the trace differ from trace by more than a simple syntactic feature. Therefore, the sentences of (13), if they are grammatical, must be base-generated. But if these cases are base-generated then there may be no reason why any relative clause should involve wh-movement, for the interpretive principle that accounts for (13) may be sufficient to account for the more usual kinds of relative as well. Thus, it would follow from a successful incorporation into the grammar of relative clauses with resumptive elements that base-generation rather than movement was the correct approach to describing unbounded dependencies.

Stating a single principle of interpretation that will handle relatives containing either gaps or resumptive elements is not difficult. In all languages, whether they form relative clauses with gaps or with resumptive pronouns, the occurrence of relative clauses is governed by an absolute semantic constraint that the clause be a statement about its head (Kuno 1976). In English this constraint is normally met because the relative clause contains a trace that is coreferential with the head (or with the wh-in COMP) so that the proposition expressed by the clause can always be taken as a function with the head as its argument. But when a relative contains no gap, it still must be interpretable as such a function. This means that the clause must contain some pronoun or full noun phrase coreferential with the head or wh-element. Thus, we can interpret a relative clause like (14) but not one like (15):

(14) the day that we were gonna leave that night
(15) *the boy that we saw the school

If we state this universal semantic constraint as requiring coindexing between the head and some NP in the clause below and if
we generate traces freely, as Chomsky (1977) already proposes for
more limited purposes, we can generate all the cases, including
those in (13), without wh-movement. Traces will be freely
generated but they will be allowable in LF only if they are
cointexed to an antecedent by a rule of interpretation (see Koster
(1978) for one such treatment). A filter will reject any
structure that contains an unbound trace. The relative clause
cointexing rule can be written so as to allow any NP in a relative
clause to be cointexed with the wh- in COMP (or with the head of
the relative), so long as its syntactic features are consistent
with those of the c-commanding antecedent phrase. To rule out
(3)/(4), all we need say is that the application of the cointexing
rule is constrained by the ECP. Then such cases will be ruled out
by the filter that rejects unbound traces. The relatives in
(1)/(2) and (13) will, of course, be fine because the ECP will not
constrain the cointexing of non-empty NP’s, so that the embedded
subject pronouns and noun phrases will be cointexed with the heads
of their clauses and the basic semantic constraint will be
honored. The only difference between the standard and colloquial
dialects on this account would be that the former allows only
traces to be cointexed by the relative clause interpretation rule,
thus blocking (1)/(2) and (13).

The solution just sketched is again attractive but again it
cannot be correct, for like Kayne’s analysis, it predicts, unless
constrained in an ad hoc way, that simple relative clauses in
English should be grammatical when they occur with resumptive
pronouns, which they are not. To prevent such overgeneration
under this account, we would have to write the cointexing rule so
that it applied only to traces in non-ECP violating environments
and only to lexical NP’s in environments where the ECP was
violated, clearly an ad hoc and unexplanatory stipulation. Here
it might seem that we could find general constraints on the
coreference of non-empty NP’s that would eliminate the
overgeneration, but this approach will not work either. If it
were the case that coreference between pronouns and preceding noun
phrases were blocked in just those environments where resumptive
pronouns are not allowed, we would have an independent reason for
ruling them out and, therefore, could allow the base component to
generate them freely. Unfortunately, the desired situation simply
fails to obtain. Even in cases where resumptive pronouns are
ungrammatical, they do occur in speech from time to time, and when
they do, they refer back to the head NP without referential
anomaly, as in (16):

(16)a. He’s very good at those gold leaf letters that
you put them on from the inside.
b. It was John H., who I’ve heard of him.
c. I went with a group that it takes people
out of a nursing home on excursions.

The problem here is not semantic, it is syntactic and a solution
that appeals to constraints on coreference will not work. Indeed, there is good reason to think that the interpretation of resumptive pronouns in relative clauses follows the same rule that governs coreference in left dislocations (Kroch, forthcoming). If this turns out to be true, then a solution to the overgeneration problem in terms of constraints on coreference is definitely ruled out, since left dislocations are not limited to "island" environments.

5. As we mentioned in starting our discussion, the unattractiveness of formal solutions to the overgeneration problem has led us to revise our opinion of the grammatical status of relative clauses with resumptive pronouns. We justify this change by the often cited rule of methodology that unclear cases should be decided by the theory, for the status of resumptive pronouns in English relatives is certainly unclear. If we go by the acceptability judgments of naive speakers, we shall conclude that all sentences containing resumptive pronouns are ungrammatical. Speakers generally reject even forms like (1)/(2) and do so even when the speakers themselves have just produced them. On the other hand, linguists, whose judgments are less influenced by normative considerations but are more theory biased, are divided in their opinions.7 What everyone admits is that there is an enormous difference between (1)/(2) and (3)/(4). The former are conceivable utterances while the latter are clearly impossible, a judgment that corresponds exactly to the facts of usage. Forms like (1)/(2) occur in speech but forms like (3)/(4) never do. This fact can be accounted for by either a grammatical or a performance theory, so that the choice of how to classify the cases can reasonably be made on the grounds of which theoretical approach is the more plausible.

The occurrence of resumptive pronouns in environments where the ECP prohibits the occurrence of gaps can be made to follow easily from the structure of a real-time sentence generator designed to model certain obvious characteristics of human speech; namely, that it is rapid and fluent but still exhibits a fair number of errors and false starts. MacDonald (1980) has designed just such a generator, which models these characteristics by having limited "lookahead" or advanced planning capacity and by having the capacity to provide output before the linguistic unit under construction is completely formed. With the slightest of modifications, MacDonald's device can be made to generate forms like (1)/(2) and to avoid (3)/(4). Only by increasing its planning capacity, however, or by decreasing its fluency can it avoid generating (1)/(2), even if the grammar that underlies it does not allow resumptive pronouns.

In handling relative clauses and other wh- dependencies, MacDonald's generator produces the constructions with the wh- element in fronted position from the outset; that is, at no point in the construction of a wh- clause is the wh- element moved from a deep structure position to the front of the clause. Instead, at
the message level (the pre-linguistic level of organization at which propositions are formulated in a non-language notation) the wh- element is already fronted. It is inserted into a tree as the sister of a clause-like structure in which the wh- element appears again in the position appropriate to its functional role (subject, object etc.). At the same time, a wh- routine is activated which keeps track of the identity of the fronted wh- message element. When the first repeated occurrence of that message element appears, this routine signals to the controller of production to realize it as a zero. By starting out with the wh- element in fronted position, MacDonald’s generator is able to provide output from a wh- clause before the entire clause up to the gap position has been formed. Since the distance between the beginning of the clause and the gap position can be indefinitely long, this ability is necessary to produce fluent speech.

From our perspective the important thing about MacDonald’s wh- routine is that it works by zeroing a repeated element rather than by movement. Because of this, the generator can begin its relative clauses without knowing what their internal syntax will be. If the repeated occurrence of the wh- message element turns out to be realized syntactically in an ECP violating environment, the generator will produce an ungrammatical sentence. As MacDonald points out, a number of things can be done to avoid creating message structures that eventually lead to syntactic violations, but it is also possible to allow such structures to occur where human beings seem to produce them. If the ECP is made a condition on the zeroing signal, then clauses containing resumptive pronouns and noun phrases in ECP violating environments will be produced. Zeroing by the wh- routine will be blocked in these cases, and other independently needed routines will then realize the undeleted message element as an appropriate discourse pronoun or anaphoric noun phrase.

In the above account, as in the earlier grammatical analyses we discussed, the ECP appears twice, once in grammar and once in the generator. One might ask, therefore, whether anything is gained by invoking processing to account for the ability of resumptive elements to amnestify ECP violations. In our view the gain is considerable. We now have a possible explanation, tentative though it may be, instead of just a description of the phenomenon. If the human sentence generator behaves like MacDonald’s model, it cannot in fluent speech take the ECP into account in choosing the message structure from which it derives its relative clauses. That would require too much advance planning. Nor can it apply the ECP as that condition appears in the grammar, that is, as an output filter, for doing so would destroy fluency. It can, however, apply the ECP as a condition on the zeroing of the repeated wh- message element, for at the point where this zeroing occurs all the syntactic information necessary to check for compliance with the ECP is available. Thus, the architecture of the generator dictates where the ECP must apply if it is to apply at all, and since ECP violating traces seem to be
as uninterpretable as they are ungrammatical, it must apply somewhere. In consequence, the ECP will apply too late to prevent initiation of the production of a clause in which a violation turns out to occur. Since, moreover, the failure to zero the repeated wh-element does not involve the direct violation of an explicit rule of the grammar, the generator will not halt but rather will proceed to finish the clause. Thus, the resulting relative clause with resumptive element, while not derivable from the grammar, is not avoidable for the generator. Therefore, it will occur in speech and will seem reasonably acceptable compared to its ECP violating alternant, which is not only ungrammatical but also uninterpretable. In the standard language, which is essentially a written mode, the planning and fluency constraints on speech are radically weakened and the avoidance of resumptive elements becomes possible, hence mandatory, given their ungrammaticality.

Footnotes
* This work was supported by NIE Grant number G78-0169, a study of the syntactic differences between written and spoken language. Thanks are due to Mark Baltin, Howard Lasnik, Ellen Prince and Don Hindle for many fruitful discussions.

1. It is unclear what Kayne means by "colloquial". We shall take him to mean the everyday speech of people who control the standard language in their writing, since our empirical study (Kroch forthcoming) suggests that forms like (1)/(2) are most common in the speech of just such people.

2. We shall have nothing to say about whether the ECP is the correct way to rule out (3)/(4). We shall simply assume it to be correct since our argument is unaffected by the question of whether it or a competing analysis gives a superior grammatical account.

3. Of course, if we assume, as in Chomsky (forthcoming) that there is no deletion in COMP but rather optional expansion of COMP in the base, then this problem disappears. As Lasnik (personal communication) has pointed out to me, however, it is not at all clear that deletion in COMP should be dispensed with, given the need to delete relative pronouns in COMP and to account for such facts as the absence of for in the complement of want just when for is immediately after want in the string, as in:

   a. I want very much *(for) you to come along
   b. I want (*for) you to come along.

4. The investigation of this possibility was suggested to me by Howard Lasnik.

5. As Kuno states it, this constraint, which he calls "the
thematic constraint on relative clauses," is a pragmatic one. It requires that the gap or resumptive element in a relative clause be in a possible topic position in the clause. What we are proposing here is a much weaker and purely semantic constraint.

6. Even in a phrase structure grammar like that given in Gazdar (1981), which is radically different in many ways from the framework we have been assuming, the constraint needed to prevent the overgeneration of resumptive pronouns is similarly ad hoc. In the case of Gazdar's grammar, the required constraint amounts to stating Gazdar's basic island constraint, the Generalized Left Branch Condition, in two different places.

7. Both Langendoen (1970) and Bever, Carroll and Hurtig (1976) assume that sentences like (1)/(2) are ungrammatical and that they occur because the speaker's sentence production routines do not always faithfully implement his/her grammar. Our analysis proceeds along the same lines.

References


—. (forthcoming) Lectures on Binding and Government.


