John Frampton

Parasitic Gaps and the Theory of Wh-Chains

1. Introduction

The aim of this article is to propose and justify some revisions of Chomsky’s (1986) barriers theory, particularly as it pertains to Wh Movement. These revisions focus on modifying the constraints on adjunction to maximal projections in Wh Movement. The constraints are relaxed in one direction by allowing adjunction to IP. But they are made more stringent in another direction by imposing a canonical head government condition as a licensing condition on adjunction. The main justification for these modifications will be an improved account of parasitic gap phenomena and a simplified account of the that-trace effect. An improved account of some of the phenomena associated with extraction from wh-islands provides further justification for the proposals.

The effect of the changes is to incorporate into the barriers theory some aspects of Kayne’s (1983) connectedness theory, particularly as it was developed by Longobardi (1984), Bennis and Hoekstra (1984), and Koster (1984). A simple example will illustrate some of the proposals that will be developed in the later sections of this article. Consider a typical subject island violation in English and its corresponding parasitic gap construction:

(1) a. *Alex, who [friends of t] admire Bill, . . .
   b. Alex, who [friends of t] admire t, . . .

The structure of the relative clause in (1b) that I will try to justify in the following pages is essentially as follows:

(2) \[ cp \ [wh_0 \ Comp \ [IP: t_1] [NP \ friends \ of \ t_2] \ Infl \ [VP: t_3, \ admire \ t_4]] \]

Primary chain: (wh_0, t_1, t_3, t_4)
Parasitic chain: (wh_0, t_1, t_2)

I am indebted especially to Richard Kayne for a careful reading of an earlier version of this article as well as for his advice and encouragement. Thanks also to Luigi Rizzi for much needed encouragement on reading the first version and to two LI reviewers for their careful and helpful criticism. Parts of this article were presented at the 1988 GLOW Colloquium and I thank the audience for their comments and questions. I have also had the benefit of comments by Andrew Barss, M. A. Browning, Noam Chomsky, Sam Gutmann, Marco Haverkort, Kyle Johnson, Pierre Pica, Tanya Reinhart, Tim Stowell, Tarald Taraldsen, and Juan Uriagereka.
For the sake of conciseness, I use the symbol IP:t to indicate the maximal projection IP with the trace t adjoined to it.\textsuperscript{1} I also omit bracketing of nonmaximal projections when the structure is sufficiently clear without them.

Several aspects of this representation should be particularly noted. First, there are two wh-chains that are headed by the same operator. Each chain, however, satisfies local binding (that is, each of its members, except for the tail, locally binds the succeeding member). The element t1, locally binds both t2 and t3 since neither of the later traces c-commands the other. Second, I depart from Chomsky’s assumption that adjunction to IP does not occur. In fact, adjunction to IP plays a crucial role in this example, the adjoined position being the locus of the connection between the primary chain originating in the VP and the parasitic chain originating in the subject NP. Third, the parasitic chain satisfies 1-Subjacency since the only barrier separating t1 and t2 is the subject NP.\textsuperscript{2}

Now consider the ungrammatical \textsuperscript{(1a)}. The structure of the relative clause in \textsuperscript{(1a)} that corresponds to \textsuperscript{(2)} is shown in \textsuperscript{(3)}:

\begin{equation}
\text{(3)} \quad \text{[CP who, Comp \textsubscript{IP:t} \textsubscript{1} [NP friends of t2]} \text{Infl [VP admire Bill]]}
\end{equation}

How is \textsuperscript{(3)} to be ruled out without also ruling out \textsuperscript{(2)}? It must be the case that adjunction is constrained in some way so that, in \textsuperscript{(2)}, t3 is involved in licensing t1. I will propose that adjunction of a wh-element to a maximal projection can only take place from a position that is canonically governed by the head of that maximal projection.\textsuperscript{3} In \textsuperscript{(2)} t3 is canonically governed by Infl, the head of IP, so that adjunction to IP is licensed. But in \textsuperscript{(3)} Infl does not govern t2 because of the intervening NP barrier.

Adjunction to IP and the licensing condition on adjunction play crucial roles in what follows. Within the original barriers framework, Chomsky makes essential use of the impossibility of adjunction to IP in establishing various constraints on wh-extraction. We need to demonstrate that these constraints can be derived even if adjunction to IP is permitted. The licensing condition on adjunction provides the needed mechanism.

I will proceed by reviewing and revising the barriers theory, showing that the constraints on extraction hold up under this theory, and taking up an explanation of the phenomenon of parasitic gaps. A simpler account of the core phenomena explained by Lasnik and Saito’s (1984) theory will also be possible on the basis of the suggested revisions of the barriers theory. Next I will propose a revised account of extraction from wh-islands. Then I will give an account of some of the subtle features of the distribution of parasitic gaps that have been studied by Cinque. This account extends to the distribution of gaps resulting from wh-extraction from wh-islands as well as to the island

\textsuperscript{1} \textsubscript{IP:t} is therefore equivalent to the standard [IP t [IP ]].

\textsuperscript{2} I assume here that the PP internal to the NP is not a Subjacency barrier, being L-marked by the head noun. Other assumptions are possible. All I require is either that who is originally in a position 1-subjacent to the trace adjoined to the IP or that it can move internally within the NP to such a position. The possibility of extracting who from friends of who, when this phrase is in object position, shows that this is the case.

\textsuperscript{3} I assume here that each language chooses a canonical direction of government, either to the right or to the left, which determines the relation of heads to their complements. Government (or m-command) in the canonical direction will be called canonical government (or m-command).
effects studied by Obenauer that involve certain adverbial operators. Finally I will propose an explanation for the tense effects observed in parasitic chains and the chains involved in extraction from wh-islands.

The scope of inquiry is limited. It deals only with Wh Movement. It does not take up LF movement, although there are indications of how the theory might be developed in this direction. It also does not discuss empty operator movement and the licensing conditions on empty operators.

2. Preliminary Modifications of the Barriers System

In this section I will assume a familiarity with the barriers system developed in Chomsky (1986) and restrict myself to presenting some modifications of that system. I will assume Chomsky's definitions of "segment," "include," "exclude," "L-mark," and "inherent barrier" without discussion. Chomsky uses the notion "minimality barrier" as well as the notion of a (Subjacency) "barrier." I will use "barrier" only in the latter sense. The notion of a "minimality barrier" does not appear in this article. Other notions should be taken as defined here.

2.1. The Special Role of IP, Subjacency

Infl and its maximal projection play a very special role in the barriers system. Three distinct special properties of IP can be singled out:

\textit{IP1}

With respect to its own status, IP always behaves as if it were L-marked, even when it is the sister of a nonovert Comp. IP is therefore never a barrier except when barrierhood is inherited from a lower maximal projection. But with respect to its own ability to induce barrierhood of higher maximal projections, IP behaves as if it were not L-marked. That is, higher maximal projections always inherit barrierhood from IP. This property is characteristic of maximal projections that are themselves barriers. (p. 14)\footnote{Chomsky's suggestion (p. 15) that the "exception for IP would be removed . . . if we were to assume that C L-marks its IP complement" is only partially correct. Without further stipulation, this would remove IP's ability to induce barrierhood in the higher CP.}

\textit{IP2}

Adjunction to IP is excluded, even though adjunction to most other nonarguments is permitted. (p. 32)

\textit{IP3}

Infl does not induce minimality effects. In general, a head $X_0$ can prevent an element $\alpha$ from governing an element $\beta$ if $X_0$ is a more local governor of $\beta$ than $\alpha$ is. It is assumed that Infl does not have this property. (p. 48)

In order to account for wh-islands (in English), Chomsky needs both to assume IP2
and to formulate the Subjacency Condition so that crossing a single barrier yields a violation. The relevant configuration for extraction from wh-islands (assuming IP2) is (4):

\[(4) \ [\text{VP}_t \ V [CP \ wh-\text{Op} [C : \text{Comp} [IP \ NP [I : \text{Infl} [\text{VP}_t \ldots ]]
\]

Only a single barrier separates the two traces. The IP is not a barrier. The CP is L-marked but inherits barrierhood from IP. If adjunction to IP is allowed, there would be no barrier separating the higher VP-adjoined trace from the lower IP-adjoined trace and there would be no Subjacency violation.

My approach will be to suppose that adjunction to IP is permitted so that extraction from wh-islands does not necessarily involve crossing any barriers at all. Since a 0-subjacency requirement will then be of no use in accounting for wh-islands and is questionable in other contexts, I will also assume that crossing a single barrier does not result in a Subjacency violation. This will mean that an account of island violations, particularly Wh-Island Constraint violations, must be developed that is quite different from the traditional subjacency account. The licensing condition on adjunction will play the key role. For precision, let us make explicit some definitions that will prove useful in what follows. I will say that X separates Y₁ from Y₂ if X includes Y₂ and excludes Y₁. I will say that X is subjacent to Y if there is at most a single barrier that separates Y and X. Eliminating IP2 also allows us to eliminate IP3. Adjunction to IP will provide a way to bypass the minimality effects of Infl. Elimination of IP2 and IP3 goes partway toward eliminating the special role of IP. But IP1 remains and continues to play a key role in establishing clausal structure in syntax.

2.2. The Head Government Condition on Adjunction

In this section I will discuss the conditions under which A-movement can exploit adjunction to maximal projections in forming chains. Chomsky assumes that wh-chain formation can exploit adjunction to nonargument maximal projections. This is qualified by stipulating that adjunction to IP is excluded as well, but we have removed this exception. Unfortunately, it must also be stipulated that adjunction to PP is excluded, even in certain cases where it does not occupy an argument position. I will assume, for the purposes of this article, that adjunction to NP, PP, and CP is excluded. These are the categories that typically represent arguments, but I assume that the exclusion does not depend on whether or not the phrase actually occurs in an argument position.

In addition to restricting adjunction by narrowing the class of maximal projections

---

5 Actually, Chomsky motivates the ban on adjunction to IP on the basis of the Subject Island Condition. Chomsky's (61a) is (i):

(i) *the man who [IP[NP pictures of t] are on the table]

But if 0-subjacency is required for grammaticality, accounting for the violation in (i) does not require a ban on adjunction to IP. The NP is not L-marked and a Subjacency violation would therefore result even if adjunction to IP is possible.
that can serve as its target, we need a further restriction. I will assume that adjunction is further restricted by the following condition:

**Head Government Condition on Adjunction (HGCA)**

A wh-element can only be adjoined to a maximal projection XP from a position that is canonically governed by the head of XP.

For the sake of concreteness, the HGCA is framed as a constraint on movement. It must be satisfied at each step in the derivation of S-Structure from D-Structure. For convenience in the following discussion, I will say that $x$ is *g-related* to $y$ if $x$ is adjoined to a maximal projection whose head canonically governs $y$.

To illustrate the HGCA, we can draw two immediate consequences. First, Wh Movement cannot adjoin the subject of IP to IP in a language (like English) in which Infl does not canonically govern the subject position. This will have important consequences later when we discuss the *that*-trace effect. Adjunction of the subject would yield the following configuration:

(5) \ldots [\text{IP} \_ t\_1, t\_2 \_ [\text{Infl} \_ [\text{VP} \ldots

But in this configuration $t\_1$ is not g-related to $t\_2$ since Infl does not canonically govern $t\_2$. Second, various island facts follow immediately. Consider subject islands. Suppose NP is in Spec(IP). Just as above, Wh Movement from inside the NP to the IP-adjoined position is not possible. Movement is forced to go at least to the Spec(CP) position. This yields a Subjacency violation since the NP (which is not L-marked) and the IP (which is a barrier by inheritance from the NP) must be crossed. This is true regardless of the direction of canonical government.

Similar considerations apply to extraction out of an adjunct clause. Consider (6), for example:

(6) *Jack, who we talked to Bill [\text{PP} because we like t], \ldots

We assume the PP is not L-marked and is included in IP or VP. Adjunction to PP is excluded, and neither the matrix Infl nor V can govern into the PP (since it is not L-marked). The extraction chain cannot therefore exploit adjunction to IP or VP, and the movement must be from a position included in PP directly to the Spec(CP). But in this case at least two barriers are crossed: the adjunct clause itself, and the IP (a barrier by inheritance). If the adjunct is VP-internal, an additional barrier must be crossed. A Subjacency violation therefore results in either case. Extraction of an adjunct itself, as opposed to extraction out of an adjunct clause, can proceed through adjunction to IP if we suppose that Infl (or V) canonically governs adjuncts.

---

6. It could equally well be framed as a condition on S-Structure representation. This would entail various modifications to later formulations.

7. The term is meant to suggest the notion of "g-projection" introduced by Kayne (1983, 167). The key insight of Kayne's connectedness theory, which I have taken over in the HGCA, is the crucial role that canonical head government plays in permitting the upward "projection" of extraction chains.

8. This assumption plays a parallel role in Longobardi (1984).
Note that in order to satisfy the HGCA, extraction from a complement CP (in English) depends upon the ability of the matrix verb to govern the Spec(CP) position of the complement. The most plausible analysis of nonbridge verbs within the current framework is that their complements are head-governed by V but are not L-marked. They therefore act like adjunct islands. The HGCA also makes a strong prediction about extraction from NP and PP complements. In the configuration [V NP] it is generally assumed that the head of the NP shields its complement from government by V. Extraction from the complement of the head of the NP via adjunction to VP will therefore require either movement through a Spec(NP) position or some form of reanalysis that nullifies the minimality effect of the intervening head. For proposals that extraction from NP proceeds via a Specifier position, see Giorgi and Longobardi (1988) and Sportiche (1989). See Browning (1987) for a discussion of extraction from PPs via a Specifier position. See Pollock (1989b) for arguments against this proposal and an alternative proposal that extraction from NP and PP depends upon a form of reanalysis. I will postpone a discussion of extraction from wh-islands until after the question of minimality is taken up in a later section.

3. Chain Formation and Parasitic Gaps

The main motivation for the HGCA is the account of parasitic chain formation that it allows. It is first necessary to discuss the general question of wh-chain formation. What I will be aiming for is a theory of chain formation that allows the primary gap and the parasitic gap to be bound by the same operator.

9 I will also propose later that, in some circumstances, the matrix verb can directly govern the embedded IP, permitting extraction that does not go through Spec(CP).

10 This was proposed by Fukui (1986), adapting an earlier proposal of Stowell (1981) to the barriers framework.

11 An anonymous reviewer suggests that I must explicitly prohibit Wh Movement to the Spec(VP) position, or lose an explanation for the island characteristics of the complements of nonbridge verbs (quip, murmur, whisper, and so on). I have no objection to supposing that Spec(VP) is an A-position and therefore unavailable for A-movement. But, with respect to extraction from the complements of nonbridge verbs, such considerations are not necessary.

In languages where VP is a barrier, movement to Spec(VP) is clearly a dead end for Wh Movement in the present system. The HGCA will not permit further movement via adjunction to VP (because V does not canonically govern its specifier) or to IP (because VP is a barrier to government by Infl). Movement is forced to go all the way to the higher Spec(CP) position, incurring a Subjacency violation.

If VP is not a barrier, as I assume later for certain Romance languages, then movement to Spec(VP) from the Spec(CP) position of a non-L-marked complement is ruled out on other grounds. The trace in Spec(VP) does not antecedent-govern the trace in Spec(CP). I will propose later that such intermediate traces in Spec(CP) must (at least) be antecedent-governed.

12 Pollock proposes that, for example, it is possible for a preposition in English to be “categorically underspecified” so that, in a configuration like the following, PP can be analyzed as a segment of the VP:

(i) [VP V [PP P NP]] → [VP V [VP P NP]]

This yields a configuration in which V and NP are not separated by a full maximal projection. It follows from an assumption that I will adopt later, that this allows V to govern NP directly.

13 The idea that the operator that binds the parasitic gap is the same operator that binds the primary gap is standard in those theories that analyze parasitic gaps as pronominal elements. Most theories that analyze parasitic gaps as traces have assumed that the parasitic chain is headed by a null operator that is licensed in some way by the primary chain. Kiss’s (1985) treatment is an exception.
3.1. Chain Formation

I will adopt the view of wh-chain formation that Rizzi (1986) originally proposed for A-chain formation. The key idea is that a chain is not a straightforward history of movement and that chain formation is a distinct process. Chains are formed at S-Structure (or some other level) under the condition that each chain element, except the tail, must locally bind its successor in the chain. I will also assume that each member of the chain must be subjacent to its predecessor. In order to make the notion of “locally bound” precise, several definitions are needed. I will say that the phrases X and Y are disjoint if they are not equal and if no segment of one of the phrases dominates a segment of the other phrase. I will say that the phrase X c-commands the phrase Y if X and Y are disjoint and if every phrase that includes X also includes Y. I will say that the phrase X binds the phrase Y if X c-commands and is coindexed with Y. I will say that the phrase X locally binds the phrase Y if X binds Y and if there is no phrase Z such that X binds Z and Z binds Y.

Rizzi’s idea leaves open the question of how traces arise. The idea is equally compatible with free insertion of traces at S-Structure or the assumption that traces result from a derivational history. I will assume that indexed traces are created by the application of Move α in the derivation of S-Structure from D-Structure. Chain formation then takes place at S-Structure and gathers these coindexed traces into chains. Recoverability of Deletion requires the content of traces to be recovered. Recovery is accomplished through chain formation at S-Structure. I will also assume that Move α is a two-step process. Typically, an element is first copied to its target location and the two elements are coindexed. Then the source is deleted, leaving an indexed trace in its place. But I will also assume that either copying or trace formation can operate independently. In general, if a trace is formed, recoverability of deletion will be impossible unless copying has applied first so that the trace can enter into chain formation. A parasitic chain will be an atypical case where recoverability is possible even though a trace is formed without prior copying. The content of the trace can be recovered by exploiting the presence of another chain.

It could be argued that the possibility of deleting an element and leaving a radically empty category, rather than a coindexed empty category (a trace), should be allowed. This option has been suggested in the literature as a means to avoid ECP violations of intermediate traces. Presumably there is no a priori Recoverability of Deletion argument against this option because it is the content of the initial trace that is at stake. I will, however, take the point of view that traces are subject to Recoverability of Deletion.

---

14 Rizzi explicitly leaves open the possibility that chain formation may be subject to further restrictions.
15 It perhaps should not be assumed a priori that the content of traces can be recovered only by chain formation. However, that discussion is outside the scope of this article.
16 Certain wh-extraction strategies that appear to involve what have been analyzed as “scope markers” could possibly involve copying and reduction (a mild form of deletion in which a reduced feature set is retained). See Maracz (1987) and Koster (1986, 84) (which contains further references).
This implies not only that “free deletion” is not an option but also that traces cannot be stranded. That is, traces must enter into chain formation.

3.2. Parasitic Gaps

We can now return to describe the process of chain formation in example (2), repeated here:

(2) \[ \text{CP who, Comp [IP:tl}, [\text{NP friends of t2}_j] \text{ Infl [VP:t3}_j, \text{admire t4}_j]] \]

At D-Structure who appears both in the position of t2 and in the position of t4. I assume that indexing is free at D-Structure and that the two instances of who bear the same index. The chain (who, t1, t3, t4) is a straightforward history of movement. Note that the requirements of the HGCA are satisfied since t3 is g-related to t4 and t1 is g-related to t3. The second chain is not a history of movement. The who that appears in the t2 position simply deletes, leaving a coindexed trace. Chain formation then takes place at S-Structure. One chain that is formed is the history of movement, (who, t1, t3, t4). The second chain is (who, t1, t2). Crucially, the second chain cannot arise as a simple history of movement because of the requirements of the HGCA. Infl does not govern the t2 position, so that adjunction to IP from the t2 position would violate the HGCA.\footnote{Note that the link between the parasitic gap and the primary chain was made on the basis of coindexing at D-Structure, not on the basis of coindexing as a reflex of movement. It is possible to exploit this fact to construct an explanation for the fact that parasitic gaps are restricted to category NP by supposing that only NPs can be indexed at D-Structure. An example of this effect follows:}

(7) Alex, who we liked t before we met t, \ldots

A sketch of the structure of the relative clause is shown in (8):

(8) \[ \text{CP who, Comp [ we Infl [ like t3}_j]] [ before [ Comp [\ldots CP IP:t1}_j] VP:t2}_j] PP CP IP:t4}_j] \]

Chains: (who, t1, t2, t3) and (who, t1, t4, \ldots)

\footnote{Note that the link between the parasitic gap and the primary chain was made on the basis of coindexing at D-Structure, not on the basis of coindexing as a reflex of movement. It is possible to exploit this fact to construct an explanation for the fact that parasitic gaps are restricted to category NP by supposing that only NPs can be indexed at D-Structure. An example of this effect follows:}

(i) a. Who did you write to t?
   b. [To whom] did you write t?
   c. Who did you write to t after talking to t?
   d. *(To whom) did you write t after talking t?

Although this is an appealing explanation if only parasitic gap constructions are considered, it is unlikely to be the correct explanation. The restriction to NP gaps that is characteristic of parasitic gap constructions is shared by a wide class of \textit{wh}-extraction constructions—including extractions from adjuncts, empty-operator constructions, and many cases of extraction from NP. Belletti (reported in Chomsky (1986, 32)), for example, has noted the following contrast:

(ii) a. Who did you leave England without speaking to t?
   b. *(To whom did you leave England without speaking t?}

An explanation based on restrictions on D-Structure indexing will extend to empty-operator constructions, but it seems dubious that the full range of cases involve empty operator constructions.
The crucial link for parasitic chain formation in this case is the link from \( t1 \) to \( t4 \). The HGCA allows adjunction to the matrix IP from the \( t2 \) position, which is governed by Infl, but it does not permit adjunction to the matrix IP from the \( t4 \) position, since that position is not governed by the matrix Infl. The formation of the parasitic chain therefore depends upon the existence of the primary chain. Note that \( t4 \) is subjacent to \( t1 \) since the only intervening barrier is the adjunct PP, assuming that the preposition L-marks the CP in the adjunct phrase. Note also that neither \( t2 \) nor \( t3 \) blocks the local binding of \( t4 \) by \( t1 \) since neither c-commands the adjunct phrase.

I have made several assumptions about the structure in (8) that are controversial. Note, however, that the above analysis is compatible with many alternative assumptions. If the adjunct phrase is in fact VP-internal, for instance, \( t2 \) would play exactly the same role that \( t1 \) plays above. The crucial point would then be the assumption that the verb governs the adjunct phrase, but its object does not c-command the adjunct phrase. I assumed that the adjunct phrase is a PP dominating a CP and that the parasitic chain reaches up to the IP-adjoined position before deletion (that is, trace formation without prior copying, as above).\(^{18}\) Other possible structures can be analyzed in roughly the same way. If the chain can reach up to the Spec(CP) position, chain formation can proceed as above. If the adjunct phrase is analyzed as a PP dominating a bare IP, chain formation can proceed via a link between the embedded IP-adjoined position and the matrix IP-adjoined position.

In the example above the actual history of movement within the adjunct clause must be to a point that is subjacent to the matrix IP in order to establish the parasitic chain. The analysis therefore immediately yields the fact that restrictions on \( wh \)-chain formation within the adjunct clause will be reflected in restrictions on the distributions of parasitic gaps. This is indeed the case.\(^{19}\) There are other restrictions on the distribution of parasitic gaps. Later I will give an explanation for the fact that parasitic gaps cannot occur in nominative subject positions in English and (relevantly) related languages. I will also return to another major case of parasitic gaps, those that occur inside relative clauses.

Note that, although I propose that the two \( wh \)-chains involved in parasitic chain formation are headed by the same operator, the distinction between the primary chain

---

See Cinque (1986) for an extensive discussion and a proposal to explain these facts by supposing that the relevant gaps are in fact pro, hence restricted to category NP. See Browning (1987) for an approach in which the relevant gaps are assumed to be traces.

\(^{18}\) Larson (1987) gives evidence that, at least for some types of adverbial adjuncts, the PP over CP structure is correct and that the Spec(CP) position is filled with a null adverbial operator. It is for this reason that I assume the parasitic chain can only reach the IP-adjoined position via movement. Note that if Larson is right, there is a strong argument for adjunction to IP. If the Spec(CP) position is filled and the IP-adjoined site is unavailable, then the parasitic chain can only reach the VP-adjoined position within the adjunct clause. We would then be faced with explaining how the connection to the primary chain could be made over two barriers: the CP (a barrier by inheritance from IP even if it is L-marked) and the adjunct PP.

\(^{19}\) Chomsky (1986, 54–56) is a good review of the evidence that the distribution of parasitic gaps obeys all of the island constraints that are expected if parasitic gaps are the tails of \( wh \)-chains. I will not repeat it here.
and the parasitic chain has not been lost. The primary chain is a history of movement. The parasitic chain is only formed on the basis of chain formation at S-Structure. It is important to be able to draw this syntactic distinction between the parasitic chain and the primary chain. Kearny (1983) gives the following examples:

(9) a. Which books about himself did John file before Mary read?  
    b. *Which books about herself did John file before Mary read?

These examples show that the binding theory properties of the fronted \(wh\)-phrase must be computed with respect to the tail of the primary chain, not the parasitic chain. Under the present analysis, this is a natural result. The operator that appears in S-Structure originated in the position of the primary gap.\(^{20}\) Parasitic gaps in relative clauses have been particularly difficult to analyze within the framework of movement theory. Adjunction to IP offers an improvement, although there are still problems. Consider (10a–c):

    c. *Jack, who I met [everyone who likes t], . . .\(^{21}\)

The parasitic character of the gap in the NP in (10a) is shown by (10b). But the remarkable fact is that extraction is not even possible in (10c), where the NP is an L-marked object.

I will make the common assumption that the relative clause is adjoined to the NP.\(^{22}\) This yields the structure shown in (11):

\(^{20}\) According to Taraldsen (MIT class lecture, 1986), in Finnish the fronted \(wh\)-phrase takes the Case marking of the tail of the primary chain, not the parasitic chain. In Hungarian there appears to be a Case-matching requirement. See Kiss (1985). I do not take up these questions here, but it is worth pointing out that the distinctions between the primary chain and the secondary chain that are available under the present theory provide a theoretical framework within which they can be addressed.

\(^{21}\) Chomsky (1986, 48) relates the existence of parasitic gaps in relative clauses to the "Vacuous Movement Hypothesis." See also Cinque (1986). The idea is that a subject \(wh\)-phrase can remain in situ, leaving the Spec(CP) position open as an escape hatch. Evidence for this claim, however, is inferred from faulty examples that violate crossing constraints (in the sense of Pesetsky's (1982) Path Containment Condition). The following are typical:

(i) a. *a book which [anyone, to whom we'll give t \(t_j\)] will like t  
    b. *a man who [anyone, to whom you introduce \(t_j\)] dislikes t

Both of these examples have crossed \(wh\)-chains. If this is avoided, acceptable examples are possible for which assumptions about vacuous movement are irrelevant. This was pointed out to me by Richard Kayne.

(ii) a. a guy that [\{every joke\}, we told \(t_j\) to] delighted t  
    b. a wall that [anything, you put \(t_j\) on] would ruin t  
    c. a paper that [everyone, we asked \(t_j\) about] praised t

The fact that the distribution of parasitic gaps is subject to crossing conditions is further evidence that parasitic gaps are traces left by \(Wh\) Movement. See Pesetsky (1987, 104) for an argument that such crossing conditions are a diagnostic for movement.

\(^{22}\) For some discussion, see Holmgren (1987).
The crucial question for parasitic chain formation is the relation between $t_1$ and $t_4$. Note that the relative clause is not *included* in the NP since it is only dominated by the outer segment of the NP. It follows that the NP does not separate $t_1$ from $t_4$ and therefore is not a barrier for $t_4$.$^{23}$ The relation between $t_1$ and $t_4$ therefore satisfies Subjacency since the only barrier that separates $t_1$ from $t_4$ is the CP, which is not L-marked. Parasitic chain formation is therefore possible as in the previous examples.

Why doesn't the possibility of adjunction to IP lead to island violations when the relativized NP is in object position? The configuration would be as follows:

\[(12) \quad [\text{VP:}t_1, V \ [\text{NP[NP ...]} \ [\text{CP who}\ [1_{\text{IP:t_2}} ...]]]]\]

$^{23}$ This contradicts Belletti and Rizzi (1988, 327), who assume that segments of a maximal projection can act as barriers to extraction.
This configuration, however, violates the HGCA because \( t_1 \) is not \( g \)-related to \( t_2 \). The head of VP, V, does not govern \( t_2 \) because of the intervening CP barrier.

This analysis of parasitic gaps in relative clauses has some weaknesses. First, it ignores the effect of the determiner of the subject NP on the possibility of a parasitic gap. The following contrast is typical:

\[(13)\]
\[
\begin{align*}
&\quad a. \quad \text{Jack, who [everyone who likes t] visited t, \ldots} \\
&\quad b. \quad \text{Jack, who [the man who likes t] visited t, \ldots}
\end{align*}
\]

Since the whole question of the effect of the determiner on the possibility of extraction from NPs is not very well understood, I will leave this problem for further work. A second weakness is that it provides no explanation for the fact that corresponding examples in Dutch are completely impossible.\(^{24}\)

A few general remarks are in order before we leave the section on parasitic gaps.

Why do parasitic gaps have the somewhat marked character that they do? Some speakers simply do not consider parasitic gap examples as fully acceptable. This deviance does not derive from the marginality of the parasitic chain. According to our analysis, parasitic chains should be fully grammatical. It is revealing to consider examples like the following:

\[(14)\]
\[
\begin{align*}
&\quad a. \quad \text{Jack, who I have persuaded friends of t that I hate t, \ldots} \\
&\quad b. \quad \text{Jack, who I have persuaded friends of t that I hate Joe, \ldots} \\
&\quad c. \quad \text{Jack, who I have persuaded friends of Joe that I hate t, \ldots}
\end{align*}
\]

\((14b,c)\) show that neither of the chains involved in \((14a)\) is deviant in itself. Nevertheless, \((14a)\) groups with other parasitic gap examples with respect to its marked character. The natural conclusion to draw is that the presence of branching chains, regardless of the character of the individual \(wh\)-chains involved, is considered (more or less, depending on the speaker) as a departure from full grammaticality.\(^{25}\) See Koopman and Sportiche (1982) and Safir (1984) for proposals along these lines.

The assumption that a parasitic gap and the primary gap are both in chains headed by the same operator, along with the assumption that variables must be A-free in the domain of the head of their chain (Principle C of the binding theory for variables), leads to the conclusion that neither the primary gap nor the parasitic gap can c-command the other. Since the "head of the chain" is identical in each case and each variable is clearly in the domain of the head of its own chain, it obviously follows that each of the variables is in the domain of the head of the other's chain.\(^{26}\) This is just the well-known anti-

\(^{24}\) It is plausible to relate this to the fact that the head of the relative clause is on the left in Dutch so that it does not canonically govern the relative clause. This is the approach taken by Bennis and Hoekstra (1984). The problem with this approach is that one loses the ability to correctly account for the fact that nonparasitic extraction is impossible from object position.

\(^{25}\) This could be viewed as saying that a chain that is not a history of movement has a marked character.

\(^{26}\) Conceptually, this argument is an oversimplification. If chain formation at S-Structure is accepted, then variables should be removed from the scope of Principle C of the binding theory. The consequences of Principle C for variables follow directly from the requirement of local binding in chain formation and the assumption that all coreferential items (not just traces) enter into the determination of the local binding relations. In my opinion, this result is a strong argument in favor of detaching chain formation from derivational history.
c-command condition, noted by Taraldsen (1981), which has played an important role in many discussions of parasitic gaps. Note that it arises as a consequence of our assumptions, not as an independent stipulation. Such a stipulation appears to be needed in movement theories of parasitic gaps that assume that the parasitic chain is headed by an empty operator.

Finally, I want to emphasize what I think is a significant achievement of the analysis of parasitic gaps that has been developed. No licensing conditions on parasitic gaps have been stipulated. The two main ingredients, the HGCA and Rizzi's chain formation algorithm, are quite generally applicable. Chomsky has emphasized the point that parasitic gaps are such a marginal phenomenon that it is highly unlikely that any knowledge specific to the construction is learned in the course of language acquisition. It is not a question of learnability, but it also seems unlikely that Universal Grammar incorporates any principles that are relevant only to this particular phenomenon.

4. Minimality, Antecedent Government, and the ECP

In this section I will consider the question of minimality—the ability of intervening governors to block a government relation that would otherwise hold. With respect to government by heads, this leads directly to an account of wh-islands. Then I will take up the question of antecedent government and propose that whereas head government can be blocked by intervening heads, antecedent government can be blocked not only by intervening heads but also by intervening operators. This will lead to an account of various facts about the distribution of parasitic gaps, the distribution of the gaps in extraction from wh-islands, the that-trace effect, and the phenomenon of pseudo-opacity.

4.1. Head Government, Minimality, and Wh-Islands

Let us first focus on the question of head government, which is relevant to the HGCA. The intuitive idea of "minimality" in this case is simple: the complement of a head cannot simultaneously be governed by a more distant head. The following purely structural definitions (abstracting away from the character of the elements involved) will prove useful:

**Definition 1**

We say that X s-governs Y if X m-commands Y and there are no barriers for Y that separate X from Y. We say that X1 intervenes between X and Y if X and Y are separated by a maximal projection, X1 canonically m-commands Y, and there is a phrase that contains X1 and Y but does not include X.

To illustrate, consider the following configuration:

(15) \[ CP \ X \ [C' \ Comp \ [IP;Z \ Y \ [I' \ Infl \ . . .]]] \]

Here Comp intervenes between X and Y. Comp does not intervene between X and Z because the latter two are not separated by a maximal projection. Infl does not intervene between X and Y because although Infl m-commands Y, Infl does not canonically
m-command Y. To anticipate, although Infl is a structurally closer governor for Y than X is, it does not block the government of Y by X.\textsuperscript{27} In terms of these structural definitions, we can define head government as follows:

\textit{Definition 2}

X \textit{head-governs} Y if X is a head that s-governs Y and there are no nonempty heads that intervene between X and Y.

With adjunction to IP, the relevant configuration for extraction from a \textit{wh}-island is as follows:

\begin{align*}
\text{(16)} & \quad \ldots \text{[VP:i]} \text{ V [CP } \text{wh-Op [C'} \text{ Comp [IP:i2]} \ldots
\end{align*}

What is at issue is the status of \textit{t1}. The HGCA demands that \textit{t1} be g-related to \textit{t2}. In effect, the requirement is that V must canonically govern \textit{t2}. Does it? The IP does not include \textit{t2}, so the CP cannot inherit barrierhood from the IP. The CP is L-marked by V. There are therefore no intervening barriers. In fact, \textit{t2} is \textit{0-subjacent} to \textit{t1}. The issue is therefore the head, Comp, which intervenes between V and \textit{t2}. Without the "nonempty" clause of Definition 2, the Comp would block government of \textit{t2} by V. This would correspond to an absolute exclusion of \textit{wh}-extraction from \textit{wh}-islands via adjunction to the matrix VP.\textsuperscript{28} The empirical facts are quite complicated. English is not usually considered to allow Wh-Island Constraint violations. But there are many examples of \textit{wh}-extraction from \textit{wh}-islands in English that are fully grammatical and many others that are best characterized as marginal rather than ungrammatical. Grammaticality judgments appear to be influenced by several well-known factors. Extraction is more difficult in an interrogative construction than a relative construction, or if the \textit{wh}-island is tensed rather than infinitival, or if the \textit{wh}-operator in Spec(CP) is associated with an argument rather than an adjunct, or if the phrase being extracted is an adjunct rather than a \textit{0}-marked complement. There are also further systematic effects if the extraction site is in a clause embedded in the \textit{wh}-island.

Consider the tensed/nontensed \textit{wh}-complement distinction in the most favorable context: a relative clause construction, an adjunct-related operator in the specifier of the \textit{wh}-complement, and a \textit{0}-marked complement being extracted. My judgments are that the infinitival case is essentially perfect.

\begin{align*}
\text{(17) a.} & \quad \text{My old Dodge, which my mechanic knew how to fix, \ldots} \\
\text{b.} & \quad \text{??My old Dodge, which I know how my mechanic fixed, \ldots} \\
\text{c.} & \quad \text{Tomatoes, which I don't know when to plant, \ldots} \\
\text{d.} & \quad \text{??Tomatoes, which I don't know when you planted, \ldots} \\
\text{e.} & \quad \text{Jack, who you know where to find, \ldots} \\
\text{f.} & \quad \text{??Jack, who you know where I found, \ldots}
\end{align*}

\textsuperscript{27} But note that Rizzi (1986; class lectures) has proposed that in Italian, as opposed to English, Infl does block government of the subject by an element in Spec(CP). This is related to the "strength" of Infl and the apparent absence of "short extraction" of the subject in Italian.

\textsuperscript{28} I will propose later that Italian can exploit another extraction strategy that is unavailable in English.
If we analyze the Comp that appears in wh-complements in English as radically empty in the case of infinitival wh-complements and as "marginally empty" in the case of tensed complements, these results follow from Definition 2 and the HGCA. Referring to (15) and (16), Comp does not block government of t2 by V in the infinitival case, so that the HGCA licenses the adjunction of t1 to the matrix VP. The grammaticality of (17a,c,e) follows. In the tensed case there is a weak violation of the HGCA since in this case Comp does (marginally) permit government of t2 by V. The marginal status of (17b,d,f) follows.

If this analysis is correct, the more severe ungrammaticality that is associated with interrogative constructions and argument-associated operators must be due to other factors. I will not pursue these problems here.

It is perhaps controversial that extraction from an infinitival wh-island should not be regarded as a violation, at least with respect to the core principles under discussion here. There may, of course, be violations due to unexplained factors associated with interrogative constructions, factors related to particular wh-phrases, and so on. It is worth noting that alongside examples like (17a,c,e), it is even possible to find acceptable examples of extraction from infinitival wh-islands that are embedded in infinitival wh-islands:

(18) a. ?These are the only vegetables which I don’t know where to find out how to plant.
   b. ?These are the only vegetables which I don’t know how to find out where to plant.

Given the complexity of these examples, their relative acceptability is striking. It is difficult to explain the acceptability of examples like these if it is maintained that extraction from infinitival wh-islands in English is essentially ungrammatical.

The case of infinitival wh-islands is important because it demonstrates that it is possible for a wh-phrase to move out of an infinitival complement without moving through Spec(CP). This will have important consequences later on, playing a key role in explaining various tense effects in wh-chain formation.

The possibility of adjunction to IP offers a very different explanation of the differences between Italian and English with respect to extraction from wh-islands. It has been suggested that "strong Infl" in the pro-drop Romance languages effectively L-marks VP so that VP is not a barrier (Kayne (class lectures, fall 1987)). Let us adopt this proposal. Consider the following structure:

(19) [CP t1j . . . [IP . . . Infl [VP V [CP wh-Op [Comp [IP:t2j . . .

Since neither the embedded CP (which is L-marked by V) nor the matrix VP is a barrier, t2 is actually 0-subject to t1. Note that the HGCA only restricts adjunction so that it imposes no condition on the movement from the t2 position to the specifier of the matrix CP.

Note also that if the specifier of the higher CP were not available for the "long
jump," then extraction would not be possible. Because of the HGCA, movement would have to go all the way to the next higher Spec(CP) position. This would entail a serious Subjacency violation because of the barriers induced by the higher IP. The prediction is that extraction from a (tensed) wh-island that is itself embedded in a wh-island should not be possible in Italian because the movement out of the first wh-island does not have a Spec(CP) available as a target for a long jump. This is indeed the case.29

If this view of wh-island extraction in Italian is correct, those Romance languages in which VP is not a barrier should pattern with Italian with respect to this phenomenon. On the basis of limited investigation, this does appear to be the case.30 If this idea proves correct, it would mean that "extractability from wh-islands" (in whatever theory-internal terms this is expressed) is not a question of independent parametric variation. Rather, it is a reflection of the status of VP as a barrier and related to the complex of syntactic phenomena that include pro-drop, subject postponing, and clitic climbing. This would certainly be a desirable result. It is hard to draw firm conclusions, however, since the question of the status of VP as a barrier in various languages has only begun to be addressed.31

It should be noted that explaining the facts of extraction from wh-islands and the Italian/English contrast is a stiff test of the attempt by the barriers theory to unify bounding theory and government theory. Rizzi's (1982) classic explanation could use parametric variation in bounding theory to distinguish the two language types. Barriers theory has, in a sense, tried to do away with bounding theory. The best result would be that only the Subjacency Condition, not subject to independent parametric variation, would remain from bounding theory. I will return in a later section to reanalyze the complicated cases of extraction from clauses embedded within wh-islands that Rizzi analyzed so successfully from the standpoint of bounding theory.

4.2. Antecedent Government, Minimality, and the ECP

Let us first make the notion of antecedent government precise.

\[ \text{Definition 3} \]

\[ X \text{ antecedent-governs Y if X is coindexed with Y, if X s-governs Y, and if there is } \]

\[ \text{no element Z that intervenes between X and Y that blocks this government.} \]

29 See examples (13)–(15) in Rizzi (1982, 54). My treatment of wh-islands obviously owes a big debt to Rizzi's essay even though it employs rather different and more modern theoretical tools. Consideration of other issues raised in that essay will be taken up in the last section of this article.

30 On the basis of the data in Torrego (1984), extraction from wh-islands in Spanish appears to pattern more or less with Italian except for complications that arise from verb preposing. This is disputed by an anonymous LI reviewer, who claims that Spanish "is at least as strict as English (more so, for some speakers) with respect to wh-island extraction." More work is clearly needed to settle the issue.

The ease of extraction from wh-islands in French appears to be intermediate between that of English and that of Italian. See Sportiche (1981). I have no explanation for this, although it is plausibly related to verb raising (to Infl) in French and its possible effect on the barrierhood of VP.

31 If Pollock's (1989a) ideas are on the right track, the barrierhood of maximal projections resulting from the ramification of Infl into its various aspects (agreement, tense, aspect, and so on) will also enter into any calculation of subjacency in a "long jump strategy" for extraction.
This definition, of course, begs the question until we specify which intervening elements can block government. Chomsky (1986) assumes that only heads can block antecedent government. Rizzi (1987) assumes that the elements that can block antecedent government depend upon the type of chain that is involved. For \textit{wh}-chains, A-specifiers (essentially operators) can block antecedent government. I will assume that antecedent government is fundamentally more fragile than head government and that \textit{either heads or operators} can block antecedent government in \textit{wh}-chains.

Jaeggli (1981, 148) views the ECP as a condition on the identification of empty elements, recognizing two forms of identification, lexical and antecedent. I will follow this approach. I will say that a \textit{wh}-trace is \textit{lexically identified} if it is an A-position that is either $\theta$-marked or Case-marked by a lexical item.\footnote{Chomsky's original formulation of the ECP employed government by a lexical category ($V^{0}$, $P^{0}$, $A^{0}$, or $N^{0}$). Stowell (1981) argued that this was not strong enough and that $\theta$-government was required. See also Rizzi (1987). Here I follow a suggestion of Lasnik and Saito (1984, 277), so that exceptional Case marking (ECM) subjects are treated like $\theta$-governed objects with respect to the ECP. Note that since inherent Case is assigned in conjunction with $\theta$-marking, only structural Case marking is relevant to the definition of lexical identification. Stowell takes a different point of view on the basis of the behavior of ECM subject gaps in empty operator constructions. I will defer a discussion of the issue since, although it is very important, it is tangential to the main line of the discussion. Stowell's assumption would restrict lexical identification to $\theta$-government. In parasitic gap constructions and extractions from \textit{wh}-islands, ECM subjects pattern with $\theta$-governed objects. I therefore will regard ECM subjects as lexically identified and assume that additional factors are at work in empty operator constructions. Note that I have excluded Case marking in Spec(CP) from consideration by requiring lexically identified positions to be A-positions.}

I will say that a \textit{wh}-trace is \textit{chain-identified} if it is connected to its operator by a chain, each of whose links satisfies antecedent government.\footnote{Implicit in this is the assumption that the operator itself is "identified." For the purposes of this article, I bypass here the question of chains headed by null operators. In this case there are likely to be conditions on the identification of the head of the chain that must be satisfied. Stowell (1985) proposes that chains headed by empty operators cannot chain-identify a trace. The distribution of gaps in these constructions will therefore share many features with the distribution of parasitic gaps and the gaps in extraction from \textit{wh}-islands.}

I will say that a \textit{wh}-trace is \textit{identified} if it is either lexically identified or chain-identified. In these terms, we can state the ECP for \textit{wh}-traces:

\textit{ECP (for \textit{wh}-traces)}

\textit{Wh}-traces in base-generated positions must be identified.

Lexical identification is only possible for the initial trace and is not an option for intermediate traces. I therefore am making the strong prediction that \textit{wh}-traces in Spec(CP) must be chain-identified even if the initial trace is lexically identified.\footnote{The original proposal that intermediate traces must be antecedent-governed is due to Lasnik and Saito (1984). They, however, allow deletion of intermediate traces associated with lexically identified elements. See Stowell (1985) and Adams (1984) for proposals more along the lines of the one made here.} Implicitly, I am anticipating that intermediate traces in adjoined positions need not be chain-identified.

The simplest way to explain the nuances of this formulation of the ECP is to proceed directly to some applications. In the remainder of this section I will concentrate on the terminal trace; in the next I will consider intermediate traces in Spec(CP).
4.2.1. The That-Trace Effect. Consider the that-trace effect in English. The configuration is as follows:

\[(20) \quad [\text{CP } x_j \text{ [Comp [IP } t_j \ldots \text{]}}, \text{ with } x_j \text{ either an operator or a trace}\]

The subject position is not lexically identified. The subject \(\theta\)-role is assigned, not by a lexical item, but by VP (or perhaps Infl, depending upon the assumptions made). Further, the subject is not Case-marked lexically. The standard Government-Binding explanation for the that-trace effect is that a filled Comp, as opposed to an empty Comp, blocks antecedent government of \(t_j\). I think this is basically correct. But a refinement is needed under the assumptions we have made. At least in the case of a tensed CP, Comp will be filled with a nonovert (but syntactically active) element. Why doesn't it block antecedent government? We will assume that spec-head coindexing is permitted with a null Comp and that coindexed intervening heads do not block antecedent government.\(^{35}\)\(^{36}\) Consider now the case of extraction of a nonsubject. The configuration is as follows:

\[(21) \quad [\text{CP } x_j \text{ [Comp [IP:} t_j \ldots\text{]}]\]

In this case Comp cannot block antecedent government of \(t_j\) by \(x_j\) since these elements are not separated by a maximal projection (since IP does not include \(t_j\)). Recall that, according to Definition 1, if two elements are not separated by a maximal projection, then no element can intervene between them. There is therefore no possibility that an intervening element can block government. This gives the well-known result that an overt Comp only blocks the extraction of the subject. Adjuncts, for example, freely extract over that in English.

Note that this account of the that-trace effect in English makes crucial use of the canonical government requirement of the HGCA. If simple government by the head of a maximal projection were sufficient to permit adjunction, then subject extraction could evade the effect of the intervening Comp by first adjoining to IP. The subject-object asymmetry would then be lost. This is a major difference between the present system and Chomsky's original barriers theory. Here the subject-object asymmetry (in English) derives from canonical government requirements, whereas in the original system it derives from the assumption that adjunction to IP is not possible.

By way of contrast, consider Dutch. I will follow much recent work and assume that Dutch is an SOV language with canonical government to the left. In this case the subject of IP is canonically governed by Infl, so that extraction of the subject can proceed via adjunction to IP. The prediction is that Dutch should not exhibit a that-trace effect.

\(^{35}\) See Chomsky (1986) for a general discussion of spec-head agreement. See Rizzi (1987) for a discussion of spec-head agreement in relation to the that-trace effect. Although it is used in a very different way, I borrow the idea of using spec-head agreement from Rizzi. It is an unsatisfactory feature of both analyses that they must stipulate that that does not enter into spec-head agreement. Otherwise, the accounts fail.

\(^{36}\) Another possibility is that the coindexed head actually becomes part of the extraction chain along the lines of the "extended chains" discussed by Chomsky (1986). Although the actual mechanism is important in other contexts, it is not important here.
This is apparently the case. Note that in Dutch, just as in English, extraction out of a subject cannot proceed via adjunction to IP since although Infl does canonically govern the subject NP, the subject NP is not L-marked. The subject NP is therefore a barrier for elements inside it, preventing government by Infl and thereby preventing adjunction to IP.

4.2.2. Wh-Islands. Consider (16) again, the structure relevant for extraction from wh-islands in English:

\[(16) \ldots [\text{VP:t}] \ V \ [\text{CP wh-Op} \ [C \ Comp \ [\text{IP:t2}] \ldots] \]

Although V head-governs t2 (perhaps marginally, depending on the character of Comp), t1 does not antecedent-govern t2 because of the intervening operator. It follows that wh-extraction from wh-islands always involves a wh-chain that does not satisfy antecedent government in at least one link. We can draw two conclusions from the ECP. First, the terminal trace must be lexically identified (since it must be identified and is not chain-identified). Extraction of adjuncts and nominative subjects from wh-islands should therefore be excluded. Second, there cannot be an intermediate trace in a Spec(CP) position within the wh-island since such a trace would have to be identified, and can be neither lexically identified nor chain-identified. The latter conclusion will be explored in the final section of this article. The former conclusion is easily verified. Consider the extraction of adjuncts from wh-islands. Adjunct extraction is sharply ungrammatical:

\[(22) \quad \text{a. } \text{??What}_j \text{ did he wonder whether to fix } t_j? \]
\[\text{b. } \text{**How}_j \text{ did he wonder whether to fix the car } t_j? \]
\[\text{c. } \text{?Which}_j \text{ tools do you know how to use } t_j? \]
\[\text{d. } \text{**When}_j \text{ do you know how to use a pipe wrench } t_j? \]

Now consider the extraction of a nominative subject.

\[(23) \quad \text{a. } \text{??He} \text{ is the guy that I found out when you said you saw } t. \]
\[\text{b. } \text{*He} \text{ is the guy that I found out when you said } t \text{ saw you.} \]
\[\text{c. } \text{**He} \text{ is the guy that I found out when you said that you saw you.} \]

(23a) is not good, but (23b) is terrible. (23a) has a weak HGCA violation, but (23b) has an ECP violation since the initial trace is neither lexically identified nor chain-identified. Note that (23c) is significantly worse than (23b). In the latter example, antecedent government in the extraction chain is broken in two places. In (19), the Italian case of “long” extraction, antecedent government is blocked not only by the intervening op-

37 There has been some controversy on this point. See Koster (1986, 206) for some discussion.
38 Note that the interrogative is marginal, whereas a corresponding relative—my old car, which I’ve wondered whether to bother fixing t, . . . —is perfect. I give the interrogative for contrastive purposes.
39 In fact, we will see later that it has two.
40 It may also be the case that violations that occur closer to the terminal trace produce sharper violations. See Rothstein (1988) for related discussion.
erator but also by several heads. Thus, in spite of the fact that Italian permits extraction of lexically identified items from \(wh\)-islands, it patterns with English with respect to the extraction of items that are not lexically identified.

4.2.3. Parasitic Gaps. Recall that in the analysis of parasitic chains given earlier, the link established by chain formation rather than actual movement crossed a (single) barrier. Parasitic chains will therefore have a link that does not satisfy antecedent government and, like extraction chains from \(wh\)-islands, must therefore originate in a lexically identified position. Parasitic gaps associated with the extraction of adjuncts will therefore be ruled out. Parasitic gaps in nominative subject positions are also ruled out. This effect was first noted by Taraldsen (1981).

(24) a. Jack, who I heard about \(t\) before we hired \(t\), . . .  
b. Jack, who you said we would hire \(t\), . . .  
c. Jack, who you said \(t\) would hire us, . . .  
d. ??Jack, who I heard about \(t\) before you said we would hire \(t\), . . .  
e. *Jack, who I heard about \(t\) before you said \(t\) would hire us, . . .

(24a) is a standard parasitic gap construction. (24b) and (24c) are both perfectly grammatical examples of \(wh\)-extraction, one from object position and one from subject position. (24d) is marginal,\(^41\) but (24e) is completely out of the question. The tail of the chain is not identified because it is neither lexically identified (IP subjects are only lexically identified in exceptional Case marking (ECM) constructions in English) nor chain-identified (because its chain does not satisfy antecedent government). If the subject position is lexically identified, as in an ECM construction, a parasitic gap is acceptable:

(25) Jack, who we hired \(t\) although we believed \(t\) to be incompetent, . . .

ECM constructions in French differ from their English analogues in that the matrix verb does not govern the embedded subject. Exceptional Case marking appears to take place in the Spec(CP) position, so that ECM constructions are only possible when there is \(wh\)-extraction that passes through Spec(CP).\(^42\) Analogues of (25) should therefore be ungrammatical in French since there will be a trace in Spec(CP) that is not identified. That trace is not lexically identified because it is not in an A-position and it is not chain-identified because the parasitic chain will have a link that violates the antecedent government requirement. According to Cinque (1986), this is indeed the case. His (40a) and (43a) are shown in (26):

(26) a. *L'homme que nous apprécions \(t\) sans croire \(t\) être intelligent . . .  
   'The man that we appreciated \(t\) without believing \(t\) to be intelligent . . .'

b. L'homme que nous apprécions \(t\) sans croire \(t\) intelligent . . .  
   'The man that we appreciated \(t\) without believing \(t\) intelligent . . .'

\(^{41}\) The deviance is due to another characteristic property of parasitic chains—they degenerate rapidly when passing through tensed clauses. This will be taken up later.

\(^{42}\) Here I follow an aspect of Kayne's (1983, 110) well-known analysis but continue to assume that ECM constructions in English involve a bare IP.
Cinque notes that (26b), a case of extraction from a small clause, is grammatical since the embedded subject position is lexically identified.

4.2.4. Pseudo-Opacity. Obenauer (1986) discusses an interesting effect in French, which he calls pseudo-opacity, which is a good illustration of the blocking effect (on antecedent government) of intervening operators. His paradigm is shown in (27):

(27) a. [Combien de films], a-t-il beaucoup vu t₁?
   how-many of films has he much seen
   ‘How many of the films has he seen often?’
   b. *Combien, a-t-il beaucoup vu [t₁ de films]?
   c. Combien, a-t-il vu [t₁ de films]?

(27c) shows that combien can be extracted from the NP combien de films. Why can’t it be extracted in (27b)? Obenauer proposes that beaucoup is an indexed operator that must be coindexed with t₁ in (27b) because it has the appropriate categorial features to bind t₁ and it is in an appropriate structural relationship with t₁ as well. This coindexing then prevents combien from locally binding its trace. (27a) is grammatical because beaucoup does not have the appropriate categorial features to bind the trace of the moved NP.

There is an alternative explanation based on the ECP.43 The trace in (27a) is lexically identified, whereas the trace in (27b) is not. If we assume, with Obenauer, that beaucoup is an operator in this construction, the facts follow from the assumptions we were already led to in explaining wh-islands. That is, beaucoup blocks antecedent government in the extraction chain, leading to ungrammaticality when the tail of the chain is not lexically identified, as in (27b).44

5. Intermediate Traces in Spec(CP) and the Tense Effect

Consider the contrast in acceptability in the following cases of extraction from clauses embedded within wh-islands:

(28) a. Sam, who I know when to try to see t, ...
b. ??Sam, who I know when to say I saw t, ...

---

43 See Rizzi (1987) for another alternative, which differs only in technical details from the one proposed here. Rizzi assumes that beaucoup is actually in the Spec(VP) position rather than simply adjoined to VP. The main difference between Obenauer’s proposal and the alternatives is that Obenauer views the problem essentially as a problem of chain formation, rather than a problem of antecedent government.

44 The details are technical. Assuming that beaucoup is adjoined to VP, the structure is (i):

(i)  . . . [VPₜ₁j, beaucoup [VP V [NP . . . t₂j . . .]]] . . .

The tricky part is to verify that beaucoup intervenes between t₁ and t₂. In order to verify this, we must show that beaucoup canonically m-commands t₂ (it does) and that there is a phrase that contains beaucoup and t₂ but does not include t₁. The VP is such a phrase. The relative order of t₁ and beaucoup (which are both adjoined to VP) is immaterial to the argument.

I assume here that beaucoup is in a standard adverb position, VP-adjointed. If I had assumed that beaucoup was not a simple adverb but instead headed a maximal projection (as in many theories of negation operators), the definition of “intervene” given earlier could be simplified.
c. *Sam, who I know why you tried to see t, . . .

d. *Sam, who I know why you said you saw t, . . .

(29) a. The Matterhorn, which I’ve decided when to attempt to climb t, . . .

b. *The Matterhorn, which I’ve decided when to announce we climbed t, . . .

c. ??The Matterhorn, which I found out why he attempted to climb t, . . .

d. *The Matterhorn, which I found out why he announced that he climbed t, . . .

(30) a. [+wh, −tense] [−wh, −tense]

b. [+wh, −tense] [−wh, + tense]

c. [+wh, + tense] [−wh, −tense]

d. [+wh, + tense] [−wh, + tense]

The complementizer patterns in (28) and (29) are as shown in (30). Judgments are not crystal clear in examples such as these, but the general pattern is apparent. If the extraction is from a tensed clause embedded within the wh-island (30b,d), there is a marked degradation in acceptability. The effect must be related to the extraction from a wh-island because, in general, tensed embedded clauses do not create obstacles to extraction in English. A sentence like He is the guy who I told you that Jack said that Mary hates t is perfectly grammatical.

Rizzi (1982) noted this tense effect, which holds for Italian as well as English. In Italian, unlike English, the analogues to both (28c) and (29c) are fully grammatical, but the analogues to (28b,d) and (29b,d) are degraded just as in English. Rizzi explained the effect on the basis of bounding theory, lending support to his theory of extraction from wh-islands and the differences between Italian and English in this regard.45

What makes the effect particularly interesting is that a sequence of two steps appears to be deviant, even though neither step is deviant in itself. Recall that I have claimed that extraction from an infinitival wh-island in English is not inherently deviant, so that we have no explanation for the deviance of the (b) examples above. In Italian there is no apparent explanation for the deviance of either the (b) or (d) type examples.

Consider the following possible structure for the (b) examples:

\[
\begin{align*}
\text{(31)} & \quad \ldots \ [V \ [\text{wh-Op}} \ Comp_1 \ [NP \ Infl] \ [V \ [t_{4j} \ Comp_2 \ [\ldots}}
\end{align*}
\]

This is ruled out by the ECP since *t4 is not identified. Recall that wh-Op blocks antecedent government, so the chain connecting *t4 to the operator does not chain-identify *t4. Another possibility is the structure shown in (32):

\[
\begin{align*}
\text{(32)} & \quad \ldots \ [V \ [\text{wh-Op}} \ Comp_1 \ [NP \ Infl] \ [V \ [\text{Comp}_2 \ [\ldots}}
\end{align*}
\]

This structure has a weak HGCA violation since the tensed Comp2 only marginally

45 Chomsky (1986, 37–39) discusses similar cases. Bordelois (1985), who gives many examples, proposes that the relevant variable in these cases has an anaphoric character and that the tense effect is due to the binding theory.
allows head government of $t4$ by $V$ exactly as in a Wh-Island Constraint violation. In the case of the (a) examples, we have the structure (33):

\[(33) \ldots [V \text{ [wh-Op} \text{ Comp}_1 [NP \text{ Infl} [V [\text{ Comp}_2 [\ldots}\]
\]

This is not an HGCA violation under our assumption that a – tense Comp does not block head government. In (33), as opposed to (32), the HGCA does license the adjunction of $t3$ because in this case the $V$ can head-govern across the infinitival Comp$_2$ just as in the case of extraction from an infinitival wh-island. We therefore have a fairly detailed explanation of the tense effect. The details are different in the case of the long jump strategy that Italian can employ, but it yields exactly the same results. In both cases the results follow from the requirement that intermediate traces in Spec(CP) must be chain-identified.

Parasitic chains display this same tense effect. Consider, for example, (34a–f):

\[(34)\begin{align*}
\text{a. } & \text{Jack, who I talked to t before Jack interviewed t, } \ldots \\
\text{b. } & \text{Jack, who I talked to t before you told me to interview t, } \ldots \\
\text{c. } & \text{??Jack, who I talked to t before you told me Jack interviewed t, } \ldots \\
\text{d. } & \text{Those papers, which I filed t without reading t, } \ldots \\
\text{e. } & \text{Those papers, which I filed t without telling you to read t, } \ldots \\
\text{f. } & \text{??Those papers, which I filed t without telling you I read t, } \ldots
\end{align*}\]

The last sentence in each group is distinctly worse than the first two. The contrast, however, does not seem to be as sharp as it is in the case of extractions from wh-clauses.\footnote{Stowell (1985), in fact, finds parasitic gap constructions to be "immune to tense effect." I do not share this judgment.} Perhaps this is a consequence of the fact that parasitic gaps are already marginal in some way. I leave this as a possible problem and turn to some other consequences of the assumption that intermediate traces in Spec(CP) must be chain-identified.

5.1. Some Consequences for Dutch

The idea that intermediate traces in Spec(CP) must be chain-identified may provide an explanation for certain puzzling aspects of the distribution of parasitic gaps in Dutch. The empirical fact is that Dutch has parasitic gaps in (tenseless) adjunct clauses, but the parasitic gap cannot be within a clause embedded in the adjunct clause, even if the embedded clause is infinitival. This was taken by Bennis and Hoekstra (1984, 67), who studied the phenomenon, to argue against a chain theory of parasitic gaps. It is possible to show, however, that if we adopt one of their assumptions, the requirement that traces in Spec(CP) must be chain-identified furnishes the basis for the correct prediction. Dutch has the peculiarity that clausal complements appear to the right of the verb whereas nonclausal complements appear to the left. We have assumed that Dutch is an SOV language so that clausal complements are not canonically governed by the higher verb.
Wh-extraction from embedded clauses is nevertheless possible. This poses a problem for accounts of wh-extraction (such as ours) in which canonical government plays a role in licensing Wh Movement. Bennis and Hoekstra, following a suggestion of Van Riemsdijk, assume that the licensing requirement for movement from Spec(CP) is weakened in Dutch so that the directionality of canonical government does not apply to this position.\footnote{The paper is written within the connectedness theory framework so the assumption is actually stated in terms of the licensing of g-projections.} This assumption certainly has an ad hoc character, but let us consider its consequences nonetheless.

The crucial consequence of this assumption is that extraction from embedded clauses (either tensed or not) in Dutch must necessarily proceed via movement through Spec(CP) since it is only this position that allows "noncanonical" extraction. As we have seen, in English it is possible for a wh-element to move out of an embedded nontensed clause without passing through Spec(CP). In Dutch this option is not available. One immediate prediction is that extraction from wh-islands should be strictly excluded in Dutch. The Spec(CP) position is filled and the IP position is not canonically governed. But only the Spec(CP) position can escape the canonical government requirement of the HGCA. The prediction is confirmed: extraction from wh-clauses is uniformly ungrammatical in Dutch.\footnote{Bennis and Hoekstra note this fact explicitly in terms that are very close to mine: "... extraction in Dutch necessarily involves an intermediate step through Comp, whereas in English a g-projection can be built directly, i.e. without an intermediate step through Comp, given the fact that the embedded clause is canonically governed by the matrix verb" (p. 38).} A second prediction is that parasitic gaps inside clauses that are embedded in adjunct clauses (tensed or not) should be impossible. The parasitic chain does not satisfy antecedent government. It follows from the ECP that the parasitic chain cannot pass through any Spec(CP) position within the adjunct clause. Since it cannot escape from the clause embedded in the adjunct clause without passing through the Spec(CP) position of the embedded clause, the conclusion follows. We thus have a chain-theoretic account of the fact noted by Bennis and Hoekstra.

5.2. Successive Cyclicity

Consider now extraction through a series of tensed clauses, as in (35):

(35) He’s the guy who Jack said that Bill thinks that Mary likes t.

In English, we have seen that the extraction must pass through each of the successive Spec(CP) positions. In a language like Italian, the "long jump strategy" allows skipping a Spec(CP) position, just as in extraction from a tensed wh-island in Italian. Recall that a "long jump" breaks antecedent government in the extraction chain. It follows from the ECP that it is only the lowest tensed Spec(CP) position that can be skipped. To see this, suppose the lowest Spec(CP) is skipped. The "long jump" must go to the next highest Spec(CP) position. It must go to a Spec(CP) position, rather than an adjoined position, because of the HGCA. It cannot go to a higher Spec(CP) because of the sub-
jumpancy condition on chain formation. Once movement goes to a Spec(CP) position, the chain must satisfy antecedent government in every higher link because the trace in Spec(CP) must be chain-identified. No further tensed Spec(CP) positions can be skipped since this entails breaking antecedent government, leaving the first trace in Spec(CP) unidentified.

We can see this at work in Spanish (which I assume allows the ‘‘long jump strategy’’ as in Italian). The phenomenon of verb fronting in Spanish is a particularly clear demonstration of the successive cyclic character of Wh Movement.49 In Spanish the extraction of wh-phrases associated with arguments causes the verb to front. Fronting probably consists of head-to-head movement of Infl to Comp, the verb having undergone head-to-head movement to Infl. The precise analysis is not important for our purposes. In the Spanish equivalent to a sentence like (35), verb fronting is obligatory in each of the embedded clauses except for the most deeply embedded clause. It is natural to regard verb fronting as triggered by the presence of a wh-operator in Spec(CP). If so, this is exactly what our analysis predicts. The first Spec(CP) can be skipped using the ‘‘long jump strategy.’’ Since the Spec(CP) is empty, unlike the wh-island case, skipping the position is possible, but not obligatory. But then the extraction is forced to pass through each higher Spec(CP) position.

There is a related phenomenon in Modern Irish that was pointed out by Chung and McCloskey (1987).50 In the Irish analogue to (35), the complementizer that is equivalent to that (go) is replaced by the complementizer aL if a wh-phrase has been extracted from the clause it heads. As in Spanish verb fronting, a ‘‘trail’’ is left so that the passage of the wh-phrase out of successive clauses can be read directly off the S-Structure. I will adopt Rizzi’s (1987) suggestion that the go → aL complementizer shift is an instance of obligatory spec-head agreement. The presence of aL reflects agreement between the head of CP and the +wh feature of the specifier of CP and therefore indicates the presence of an intermediate trace in Spec(CP). What is important about the phenomenon in the present context is that the shift is obligatory, except for the deepest complementizer. The deepest Comp can remain go at the cost of somewhat marginal acceptability. I interpret this to mean that the extraction chain can avoid the first Spec(CP) position (perhaps at the cost of degraded acceptability), but it cannot avoid the higher Spec(CP) positions. This is exactly what we would expect in English if English had a that → aL spec-head agreement phenomenon. Skipping the first Spec(CP) would entail a weak HGCA violation akin to a tensed Wh-Island Constraint violation. Skipping the first two Spec(CP) positions would entail a double weak HGCA violation, presumably significantly worse. Once a Spec(CP) position is used, the trace in that position must be identified. It can only be chain-identified. Skipping a higher Spec(CP) position would entail an ECP violation.

49 Its existence and importance were recognized by Torrego (1984). She analyzed the phenomenon along the lines of Rizzi’s analysis of extraction from wh-islands. I rely heavily on her analysis and completely on her data.

50 See footnote 33 in particular.
6. The Role of Canonical Head Government

The closest antecedent to the theory that has been developed here is that of Stowell (1985). Although the framework is very different, several of Stowell's ideas have been incorporated into the present theory, most notably the relationship between tense effects, successive cyclicity, and the ECP. Stowell's work approached questions related to the ones addressed above by adopting a form of the ECP that requires traces to satisfy both a head government requirement and a local binding condition. Specifically, Stowell requires that (1) traces must be canonically head-governed, and (2) traces must be locally bound (this includes a subjacency requirement, as does my requirement for chain formation). In addition, Stowell adopts a principle very much like what I have called the ECP as an independent principle.

I have subsumed the requirement that traces be locally bound under the general theory of chain formation. The comparison of the two theories then reduces (approximately) to a comparison between the HGCA and Stowell's requirement that traces be canonically governed. It is impossible to compare the two theories in a thorough way without untangling the extensive web of assumptions that underlie the two differing approaches, a task too large to be undertaken here. The main obstacle to a direct comparison is that Stowell's work is not done within the barriers framework so that (1) he is not forced to make assumptions about bounding nodes on the basis of assumptions about government and (2) the question of how (or whether) the ECP constrains intermediate adjoined traces does not arise.

In simplest terms (perhaps overly so), the HGCA requires that an element that moves to an adjoined position must be canonically governed (by a particular item related to its target position), whereas Stowell requires that an element that moves anywhere must be canonically governed (by something). The danger of oversimplification comes from the fact that in my theory the HGCA is viewed as a licensing condition only on intermediate adjoined traces, elements that do not even appear in Stowell's theory. But it does appear that there is little to choose between the two theories with respect to simplicity.

With respect to empirical coverage, the immediate place to look for differences is subject extraction since there is a question of canonical head government for this position. We do not require canonical head government for movement from Spec(IP) to Spec(CP) since no adjunction is involved. But it does not appear to yield a decisive test. Stowell obtains the that-trace effect under his assumptions by assuming that in the absence of that, the matrix verb can canonically govern the subject position, satisfying the canonical head government condition. The assumption that the matrix verb governs into the embedded IP is, of course, inconsistent with assumptions of barriers theory. But there is no inconsistency in Stowell's theory because it does not demand a relationship between subjacency barriers and barriers to government. It is perhaps unsettling that movement internal to CP is licensed by an element external to CP, but each theory holds up in its own terms.\footnote{See Rizzi (1987) for a closely related proposal that uses a nonovert element in Comp to provide a...}
The unified account of parasitic gaps on the one hand, and subject and adjunct islands on the other, is the most compelling argument in favor of the HGCA. All chain theories of parasitic gaps have shared the view, in one form or another, that the parasitic chain has to get close enough to the primary chain in order to be licensed. It is also apparent that the "point of contact" is an intermediate trace. The following question then arises, in a rather sharp way I think. If the parasitic chain gets close enough to a position in the primary chain to be licensed, why isn't it close enough to actually move to the position in question? In other words, why should a licensing condition on the position of the head of a parasitic chain involve a weaker relation of subjacency than that required for movement? Unless movement out of islands is constrained by some condition other than distance (measured by subjacency), there is no ready answer. The HGCA is the answer to this puzzle. It is not just subjacency that prevents the parasitic chain from leaving its island. Subjacency prevents a long movement, but the HGCA prevents a short movement. Extraction from an island is ruled out ultimately because external heads cannot govern into the island. In addition to providing an answer to this puzzle, it is no small point that the HGCA allows us to avoid having to burden Universal Grammar with a principle of parasitic chain licensing.

7. Conclusions

The most interesting result of the present investigation is the sharp split between the properties of intermediate adjoined traces and traces that appear in base-generated positions. The latter are subject to the ECP, the former to the HGCA. It is possible that intermediate adjoined traces should not be viewed as traces at all but should be viewed as some kind of scope marker. It may be possible to construct a theory along these lines. But it is not clear how this can be done without introducing a corresponding conceptual problem. Intermediate adjoined traces do share many of the properties of those traces that are subject to the ECP—they bear referential indices, enter into chain formation, and serve as antecedent governors.

Another possible direction to take in conceptual unification might be to claim that all traces are subject to the ECP but that intermediate adjoined traces can delete, along the lines of Lasnik and Saito's (1984) approach. Unlike their assumptions, however, deletion would have to take place after chain formation since intermediate adjoined traces seem to play a crucial role in parasitic chain formation. In order to give that approach force, a way would have to be found to justify deleting only adjoined traces. Further, one is still left with the fact that only adjoined traces are subject to the HGCA, as a condition on movement.

---

canonical head governor. This version does not face the objection that CP-internal movement is licensed by a CP-external element, but one may question why an overt Comp like that is an insufficient head governor while a nonovert Comp is sufficient. Rizzi assumes that heads play no role in blocking antecedent government, so that antecedent government is no problem, only the need for a canonical head governor. The fact that Rizzi must assume that the nonovert Comp fulfills the role of head governor only when it is coindexed with the subject position perhaps indicates that it is, in fact, antecedent government that is at stake, not head government.
I pointed out earlier that the HGCA incorporates some aspects of connectedness theory into the barriers theory. The split between the properties of traces in adjoined and base-generated positions is a reflection of this merger and shows that the contradictions between the theories have not yet been completely resolved.

Putting aside this major conceptual question, the present analysis offers evidence bearing on several current questions. The success of the analysis of parasitic gaps in predicting the detailed structure of the distribution of parasitic gaps, without resort to stipulation, supports the view that parasitic gaps are indeed traces and not null resumptive pronouns. It is technically possible to avoid Rizzi's Chain Formation Algorithm in generating the split chains that arise in parasitic gap phenomena, but only at the cost of a substantial loss of elegance and simplicity. This is support for the concept of chain formation as distinct from derivational history. The fact that fairly detailed results on tense effects and successive cyclicity were obtained that were previously obtained only on the basis of bounding theory parametrization lends support to a leading idea of Chomsky (1986), the idea that bounding theory and government theory can be unified.

References

PARASITIC GAPS AND THE THEORY OF WH-CHAiNS


Pollock, J.-Y. (1989b) "Opacity, Genitive Subjects and Extraction from NP in English and French," ms., Université de Haute Bretagne, Rennes II.


Department of Mathematics
Northeastern University
Boston, Massachusetts 02115

Bitnet: frampton@nuhub