3.8 Conclusion

In this chapter, we have considered a number of widely held assumptions about the nature of complementation in English, arguing in particular that small clause analyses fail to explain a variety of important facts. We have also presented an analysis of infinitival complements in English based on the premise, which we have justified at length, that such complements are not clausal as standardly assumed. In the analysis we have sketched, the interpretation of the unexpressed subjects of these complements is determined by lexical properties of the verbs and adjectives that co-occur with them. We have summarized the familiar distinction between raising and equi expressions and provided an account of their differing properties. Equi dependencies are treated in terms of coindexing, and nothing more. The dependency in raising constructions, on the other hand, is both syntactic and semantic in nature. This dependency is correctly treated by our Raising Principle, which, as we have shown, also predicts key properties of raising constructions discussed by Jacobson (1990). We have also sketched our treatment of expletive pronoun constructions, presenting a detailed treatment of their behavior in raising constructions.

In Chapter 7, we return to the analysis of unsaturated complements, showing that the dependencies isolated here need not be lexically stipulated, but rather follow from the interaction of principles of syntactic argument structure and semantically based principles of controller assignment whose application is more general than the complements of verbal expressions.

4

Unbounded Dependency Constructions

4.1 Introduction

In this chapter, we discuss in some detail how the theory of HPSG can be brought to bear on the analysis of unbounded dependency constructions (UDCs). This term was introduced by Gazdar (1981) to refer to a class of constructions standardly analyzed by transformational grammarians in terms of *wh*-movement, or—to use more recent transformational terminology—movement to the nonargument position SPECIFIER of CP (Chomsky 1986b). For English, the class of unbounded dependency constructions includes such phenomena as topicalization, *wh*-questions, relative clauses, *it*-clefts, pseudoclefts, purpose infinitives, and *tough ‘movement’,* which are illustrated in (1) and (2):

(1)

a. Kim, Sandy loves —_,

b. I wonder [who, Sandy loves —_].

c. This is the politician [who, Sandy loves —_].

d. It's Kim [who, Sandy loves —_].

e. [What, Kim loves —_] is Sandy.

(2)

a. I bought it, for Sandy to eat —_.

b. Sandy, is hard to love —_.

c. This is the politician, [Sandy loves —_].

d. It's Kim, [Sandy loves —_].

Such constructions fall naturally into two classes. In those in (1), there is an overt constituent in a nonargument position—either a topic or an expression containing a *wh*-phrase—that can be thought of as strongly associated with (or *filling*) the gap or trace, indicated by —_. We will refer to the constructions in this class as *filler-gap* constructions, or strong UDCs.

In the second class of unbounded dependency constructions, such as those in (2), there is no overt filler in a nonargument position; instead there is a con-

1. We thank Georgia Green and Elisabet Engdahl for detailed comments on earlier drafts of this chapter and for numerous improvements.
stituent in an argument position that is—loosely speaking—interpreted as coreferential with the trace. Constructions of this subclass are the ones that are analyzed by current transformational grammar in terms of an empty operator. The analysis we will propose, however, involves no such empty operators. Rather, it will treat the relation between the argument and the trace as one of coin dexing. We will refer to these constructions as weak UDCs.

Initially, we will be concerned with topicalization and tough movement as representative of the two main subtypes of unbounded dependency, but the analyses we will propose can be extended or adapted in natural ways to deal with UDCs in general. Before turning to specifics of the analyses, however, we must mention two basic points about unbounded dependencies that should be borne in mind. One is that they are indeed unbounded, which means that the dependency in question may extend across arbitrarily many clause boundaries; and the other is that, at least in the case of strong UDCs, there is a syntactic category matching condition between the filler and the gap. Both of these points are illustrated by the examples in (3)–(5):

(3)  
   a. Kim, Sandy trusts ———.  
   b. [On Kim], Sandy depends ———.  
   c. *[On Kim], Sandy trusts ———.  
   d. *Kim, Sandy depends ———.  

(4)  
   a. Kim, Chris knows Sandy trusts ———.  
   b. [On Kim], Chris knows Sandy depends ———.  
   c. *[On Kim], Chris knows Sandy trusts ———.  
   d. *Kim, Chris knows Sandy depends ———.  

(5)  
   a. Kim, Dana believes Chris knows Sandy trusts ———.  
   b. [On Kim], Dana believes Chris knows Sandy depends ———.  
   c. *[On Kim], Dana believes Chris knows Sandy trusts ———.  
   d. *Kim, Dana believes Chris knows Sandy depends ———.  

In (3) the trace is an argument of the main clause, in (4) it is an argument of an embedded complement clause, and in (5) it is an argument of a doubly embedded complement clause within a complement clause; in principle there is no bound on the depth of embedding.

The other point is illustrated by the contrast between the (a) examples, where the trace is in the object position of trust, which requires a noun phrase, and the (b) examples, where the trace is in the object position of depend, which requires a prepositional phrase headed by the preposition on. In either case, the filler must respect the syntactic category requirements imposed by the local environment of the trace; thus the (c) and (d) examples where it does not are ungrammatical. The point here is simply that in filler-gap constructions (strong UDCs) there is a genuine syntactic dependency between the filler and the gap.

4.2 Filler-Gap Constructions

As mentioned in Chapter 1, the nonlocal features are analogous to GPSG’s foot features. However, there are two important differences that we want to mention here. First, the inventory of features is different. Both GPSG and HPSG employ a slash feature for dealing with traces, but unlike GPSG, HPSG does not use a refl feature for analyzing reflexive pronouns. In fact, in HPSG reflexives are not treated as unbounded dependencies at all, but instead are handled by the binding theory developed in Chapter 6. In addition, instead of a single wh-feature, HPSG employs distinct features Que and rel for interrogatives and relatives. This distinction is motivated, among other things, by familiar distributional differences between interrogative and relative pronouns, for example, with respect to pied piping, as shown in (6):

(6)  
   a. This is the farmer pictures of whom appeared in Newsweek.  
   b. *Pictures of whom appeared in Newsweek?

The second difference between HPSG nonlocal features and GPSG foot features is that the data types are different. In GPSG all foot features take the same kind of value, namely a syntactic category. But in HPSG (following a suggestion made originally by Maling and Zaenen (1982)), nonlocal features take sets as values. This will allow for the treatment of multiple unbounded dependencies, such as those illustrated in (7):

(7)  
   a. [A violin this well crafted], even [the most difficult sonata], will be easy to play ——— on ———.  
   b. This is a problem which, John, is difficult to talk to ——— about ———.  

What kind of members those set values have depends on the feature, as shown in (8). This is because in an interrogative dependency, the only information that has to be kept track of is the nominal–object corresponding to the wh-phrase, and in a relative dependency, all that is needed is the referential index associated with the relative pronoun.

(8)  

---

2. We provide no HPSG analysis of interrogatives in the present volume. However, this task has been undertaken by Ginzburg (1992).

3. Implicit in our decision to treat the value of nonlocal features as sets, rather than lists, is a decision to treat Fodor’s (1978) nested dependency constraint as a matter of processing, rather than grammar. See n. 9 below.

4. See the HPSG analysis of interrogatives of Ginzburg (1992), whose restricted parameters are for present purposes equivalent to our npros. In an alternative approach, alluded to in P&S-87, the value of Que was assumed to be an interrogative generalized quantifier.
But in a strong UDC, all the local information—including head features, subcategorization, and content, has to be propagated from the trace position up to the filler. To illustrate this, let us first consider a somewhat simplified example of a filler-gap dependency. We will return to the same example below, filling in a number of details. The example in question is the sentence (9a), and the analysis itself is sketched in (9b):

(9) a. Kim, we know Sandy claims Dana hates 

b. 

\[
\begin{array}{c}
S \\
\text{NP} \quad [\text{LOCAL} \{1\}] \\
\text{VP} \quad [\text{SLASH} \{1\}] \\
\text{Kim} \\
\text{NP} \quad [\text{SLASH} \{1\}] \\
\text{we} \\
\text{V} \quad [\text{SLASH} \{1\}] \\
\text{know} \\
\text{NP} \quad [\text{SLASH} \{1\}] \\
\text{VP} \quad [\text{SLASH} \{1\}] \\
\text{2} \\
\text{Sandy} \\
\text{V} \quad [\text{SLASH} \{1\}] \\
\text{claims} \\
\text{NP} \quad [\text{SLASH} \{1\}] \\
\text{Dana} \\
\text{V} \quad [\text{SLASH} \{1\}] \\
\text{hates} \\
\end{array}
\]

This analysis is similar in spirit to the one proposed by Gazdar (1981), especially as developed by Gazdar et al. (1985), though the technical details are slightly different. The essence of the analysis is that an unbounded dependency has three parts: a bottom, a middle, and a top. These three parts are labelled in (9) as 1, 2, and 3, respectively. The bottom of the unbounded dependency construction is where the dependency is introduced; the middle of the dependency is where it is successively passed from daughter to mother up the tree; and the top is where the dependency is discharged or bound off. Let us consider these one by one.

First, the bottom, where the dependency is introduced. We assume that every unbounded dependency is introduced at a terminal node by a special sign that has a nonempty value for the appropriate nonlocal feature. In a relative construction, this sign is a relative word (e.g. who, which), containing a nonempty value for the REL feature; and for a wh-question, it is an interrogative pronoun with a nonempty value for the QUE feature. In the present case, what we have is a SLASH-type dependency, and the special sign that introduces it is a trace.

But what is a trace? In our theory a trace is just a special lexical item, which is shown in (10a):

(10) a. Trace as it appears in the lexicon (simplified version):

\[
\begin{array}{c}
\text{PHONOLOGY} \{ \} \\
\text{SYNSEM} \\
\text{LOCAL} \{ 1 \} \\
\text{NONLOCAL} \\
\text{QUE} \{ \} \\
\text{REL} \{ \} \\
\end{array}
\]

As is clear, a trace is a quite impoverished structure. It has no phonology, of course; it specifies no local features, and as far as nonlocal features go, the only nonempty value is the SLASH value, which is a singleton set. And what is the member of that set? It is just whatever the local features of the trace are. That structure sharing is indicated by the two occurrences of the tag \{1\}.

Now a trace, as it occurs in the lexicon, appears to be rather useless. But once a trace occurs as a complement of some head, then it will structure-share whatever local features are specified for that complement by the head. And of course whatever those local features are, they will also show up in the SLASH value set of the trace. Thus, the trace in the structure (9b), which is shown in close-up as (10b), has the local structure imposed by the verb hates on its object position, and that same local structure is found in the SLASH value set of the trace:

(10) b. Trace in (9b) with LOCAL features imposed by hates on its object position (simplified version):

\[
\begin{array}{c}
\text{PHONOLOGY} \{ \} \\
\text{SYNSEM} \\
\text{LOCAL} \{ 1 \} \\
\text{CAT} \{ \text{HEAD noun [CASE acc]} \} \\
\text{SUBCAT} \{ \} \\
\end{array}
\]

\[
\begin{array}{c}
\text{LOCAL} \{ 1 \} \\
\text{SLASH} \{ 1 \} \\
\text{NONLOCAL} \\
\text{QUE} \{ \} \\
\text{REL} \{ \} \\
\end{array}
\]
The second part of our unbounded dependency in (9b), labelled 2, is the middle, where the information that there is an unbound trace is propagated up the tree. The mechanism responsible for this propagation is one of the principles of universal grammar posited in HPSG, called the Nonlocal Feature Principle. This is analogous to the Foot Feature Principle of GPSG, and a simplified version of it is stated in item (11):

(11) **Nonlocal Feature Principle** (simplified version):

The value of each nonlocal feature on a phrasal sign is the union of the values on the daughters.

In the case of a structure like (9b), where just one unbounded dependency is introduced, namely the single member of the SLASH value set on the trace, the effect of the Nonlocal Feature Principle is simply that every node in the tree that dominates the trace will also have just that one element in its SLASH value set. Thus the SLASH value is passed from the trace to its mother, then to that node’s mother, and so forth up the tree.

Finally, the third part of the unbounded dependency is the top, where the SLASH value introduced by the trace is bound off or discharged, in this case by identification with the local features of the filler, namely the noun phrase Kim. Again, as in Gazdar’s theory, this is guaranteed by a phrase structure rule, or more properly an immediate dominance schema, to the effect that one way to form a phrase is from a finite sentence containing a trace that is unbound within that sentence and a filler whose local features match those of the trace. This rule is expressed informally in rewrite form in (12a), and somewhat more formally as a feature description in (12b). This rule is responsible for licensing the top node in (9b), labelled 3. All other phrasal nodes in this tree are licensed by the head-complement rules discussed in previous chapters.

(12) **Head-Filler Rule** (Schema 6) (preliminary version):

a. $X \rightarrow \text{[Local [L] [SLASH (]) ... ]}$

b. $\text{[FILLER [DTR | SYNSEM [Local [L]]] \text{DAUGHTERS [COMPLEMENT-DTRS ( ) \text{HEAD-DTR \ SYNSEM [Local [CAT [HEAD void [VFORM fin]]] \ SUBCAT ( ) \ NONLOCAL [SLASH (]) ... ]]} \ ]]}$

The key point to note about this analysis of strong UDCs is that the local features of the trace position and those of the filler are both identical with the local structure passed up as a SLASH value, and therefore by the transitivity of equality they are identical with each other. In other words, the filler and the trace are the same thing, at least as far as their local structure is concerned. Thus an effect similar to that of wh-movement is achieved, without actually moving anything.

This is typical of a more general difference between derivational theories and constraint-based theories. In derivational theories, like GB, the fundamental explanatory mechanisms are transformations, such as move-a, that successively convert one level of structure into another. But, however well the facts may appear to motivate different levels of structure, we are not aware of any evidence that transformations themselves correspond to anything in the empirical domain. In constraint-based theories like HPSG, by contrast, the fundamental explanatory mechanism is conditions of identity or sharing of substructure by different attributes of a common structure.

Before proceeding, we must consider a problem that was glossed over in the simplified analysis of topicalization in (9b). The problem is that in the account we gave of (9), there is nothing in the analysis that prevents the SLASH feature from simply continuing to be passed up the tree, even after it has been identified with the filler, or even from being bound again higher in the tree. Thus, as matters stand, there is nothing in the theory to rule out ungrammatical sentences like the ones in (13):

(13) a. *Bagels, I know that bagels, they like —.-

b. *Who, did you wonder who, Kim saw —.-?

c. *Kim is hard to find the student who, Sandy met —.-

What is needed is a mechanism to guarantee that once a trace has been bound, the member of the SLASH value set that corresponds to it will not be passed any further up the tree. What we propose in this connection is to distinguish between unbounded dependencies that are required by the grammar to become bound and those that continue to be inherited upward. More specifically, we replace nonlocal structures like that in (8) with structures like the one shown in (14):

(14) Internal structure of nonlocal value:

```
[INHERITED [QUE (set of npros) [SLASH (set of local structures)]]]
[REL (set of ref’ indices) [TO-BIND [QUE (set of npros) [SLASH (set of local structures)]]]]
```

Given this change, the structure of a trace has to be changed correspondingly, from the form given in (10a) to the form shown in (15):
(15) Trace as it appears in the lexicon (final version):

\[
\begin{align*}
\text{PHONOLGY} & \quad (\quad) \\
\text{LOCAL} & \quad [1] \\
\text{SYNSEM} & \quad (\quad) \\
\text{NONLOCAL} & \quad \text{INHERITED} \\
\text{TO-BIND} & \quad \text{QUE} \\
\text{REL} & \quad \text{SLASH} \\
\text{SLASH} & \quad \text{SLASH} \\
\end{align*}
\]

Likewise, the Nonlocal Feature Principle has to be adjusted to guarantee that those nonlocal dependencies that become bound off are subtracted from the set of nonlocal feature values that are passed up to the mother. The way we account for this is by assuming that the binding of a nonlocal dependency is always declared on the head daughter of the structure in question. Thus the Nonlocal Feature Principle has to be changed from the simplified version shown in (11) to the final version given in (16):

(16) **Nonlocal Feature Principle** (final version):

For each nonlocal feature, the INHERITED value on the mother is the union of the INHERITED values on the daughters minus the TO-BIND value on the head daughter.

Now one way that a nonlocal dependency can be bound off is for some grammar rule to do it. Our topicalization example (9) is an instance of this. But to make this precise, we must change the rule that licenses head-filler structures from the preliminary version given in (12) to the final version given in (17):

(17) **Head-Filler Rule** (Schema 6) (final version):

a. \( X \rightarrow [\text{LOCAL} \ [1], \ [f/fn, \text{INHER} \ | \text{SLASH} \ [1, \ldots]], \text{TO-BIND} \ | \text{SLASH} \ [1]] \)

b. \[
\begin{align*}
\text{FILLER-DTR} & \quad \text{SYNSEM} \quad [\text{LOCAL} \ [1]] \\
\text{COMPLEMENT-DTRs} & \quad (\quad) \\
\text{DAUGHTERS} & \quad \text{HEAD-DTR} \\
\text{HEAD-DTR} & \quad \text{SYNSEM} \\
\text{LOCAL} & \quad [\text{CAT} \quad \text{HEAD-NP} \mid \text{VFORM} \mid \text{fin}] \\
\text{SUBCAT} & \quad (\quad) \\
\text{LOCAL} & \quad [\text{INHER} \mid \text{SLASH} \ [1, \ldots]] \\
\text{TO-BIND} & \quad [\text{SLASH} \ [1]] \\
\end{align*}
\]

5. In addition, we must modify the other schemata of our theory that introduce phrasal heads so that the specification \[\text{TO-BIND} \mid \text{SLASH} \ [1]\] is added to the head daughter.

With these adjustments in place, the final analysis of our topicalization example is as shown in (18):

(18) a. Kim, we know Sandy claims Dana hates ———.

b. 
\[
\begin{array}{c}
\text{S} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{TO-BIND} \mid \text{SLASH} \ [1] \\
\end{array}
\]

\[
\begin{array}{c}
\text{NP} \\
\text{we} \\
\text{know} \\
\text{V} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{S} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{V} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{NP} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{want} \\
\text{NP} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{Dana} \\
\text{V} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{NP} \\
[\text{INHER} \mid \text{SLASH} \ [1]] \\
\text{hates} \\
\end{array}
\]

6. In light of the well-known violations of strong connectivity illustrated in (i) and (ii), it may ultimately be necessary to modify the account of topicalized sentences presented in the text:

(i) *You can rely on that Kim will help you.

(ii) That Kim will help you, you can rely on.

Such a modification might involve classifying head-filler structures into a variety of subsorts, one of which involves a filler that is merely coindexed with the INHER|SLASH value of the sentential head. In this way, a kind of topicalization would be recognized (e.g. (ii)) that involved only a weak UDC, i.e. where no identity is required between the category of the filler and that of the gap. This analysis would of course have to be
4.3 Tough Constructions

Let us now turn to the phenomenon known as tough movement. This term refers to examples like the ones in (19), but the analysis we will provide can also be adapted to deal with examples involving verbs like cost and take, as shown in (20) and (21):

(19) a. Kim, would be easy to bribe.
b. Kim, would be easy to prove Sandy bribed.

(20) a. [This theorem], will take only five minutes to prove.
b. [This theorem], will take only five minutes to establish that Gödel proved, in 1930.

(21) a. [This widget], will cost Kim $500 to fix.
b. [This $500 bribe], will cost the government $500,000 to prove that Senator Jones accepted.

Examples like these all belong to the subclass of unbounded dependency constructions that we referred to as weak UDCs, and they differ from strong UDCs in the key respect that there is no filler corresponding to the trace. Instead there is a constituent in an argument position that is coindexed with—that is, has the same referential index as—the trace. In all the examples in (19)—(21), this constituent is the subject. The fact that this is not a filler-gap dependency is underscored by the fact, illustrated in all of (19)—(21), that the trace and the coindexed subject need not have the same case.

(22) "I, nom" am easy to please, "acc".

Such constructions have been problematic for transformational grammar because they do not fit comfortably with any of the usual subvarieties of move-α. On the one hand they cannot be wh-movement, since the constituent coindexed with the trace is in an argument position; on the other hand they cannot be NP-movement either, since the trace is in a case-assigned position.

GB theory's solution to this problem (Chomsky 1977) is to propose an empty operator, a phonetically null constituent in a nonargument topic position that binds the trace and is coindexed with the subject, as shown in (23):

(23) Empty operator analysis:
I, am easy [S O₁ [S PRO to please.]]

On our analysis, by contrast, it is just a lexical fact about predicates like easy, take, and cost that they subcategorize for infinitive complements containing an accusative NP gap coindexed with the subject. Thus the heart of our analysis is captured in item (24), which is a partial representation of the SYNSEM value of the lexical entry for the tough-class adjective easy:

(24) Partial representation of SYNSEM value for easy:

```
LOCAL|CAT [HEAD adjective
  SUBCAT (NP, [PP, for.])
  VP [inf, INHER, SLASH (2) NP[acc]; ppro [i]]]

NONLOCAL, TO-BIND, SLASH (2)
```

8. Here NP[acc]; ppro[i] abbreviates an accusative NP whose content is of sort ppro and whose index is tagged as [i]. We are assuming here that the subject of easy-class adjectives is assigned a semantic role and hence that lexical forms like (24) do not constitute a violation of the Raising Principle. To see the potential problem here, consider the fact that identifying the unexpressed subject of the VP complement with the subject of easy would produce the wrong interpretation (the unexpressed subject must be associated with the for phrase, if the latter is expressed). Yet the Raising Principle requires that any SUBCAT member not assigned a role (other than a lexically specified expletive) be a raising controller.

There is considerable evidence in favor of our claim that the subject of easy is role-assigned. First, this immediately explains why examples like (i) are ill-formed, since the expletive there bears an index that is nonreferential:

(i) *There is easy to believe to be a unicorn in the garden.

Second, we also obtain an explanation for the fact that easy and the like may undergo null complement anaphora (see section 5 of Chapter 3), as illustrated in (ii):

(ii) Kim is hard to talk to but Sandy is easy.

Finally, assigning the subject of easy a role allows an account of the well-known fact that examples like (iii) and (iv) differ in interpretation:

(iii) This sonata is easy to play on that violin.
(iv) That violin is easy to play this sonata on.
Thus sentence (22) is assigned the structure sketched in (25):

```
      s
     / \  
    /   \  
   /     \  
  VP  [SUBCAT (3)]
  /     \
 l      
  V     /  
 SUBCAT (3) [INHER | SLASH {1}]
 /     \  
 am    /  
 [TO-BIND | SLASH (2|NP)] [INHER | SLASH {2}]
 /     \  
 easy /  
 VP[inf] [INHER | SLASH {2}]
 /     \  
 to   /  
 v[base] [INHER | SLASH {2}]
 /     \  
 please /  
 NP[1] [INHER | SLASH {2}]
```

Note that this structure is generated by the usual head-complement rules. The coindexing of the trace with the subject obtains because it is directly specified in the lexical entry for easy. And the SLASH value corresponding to the trace is not inherited beyond the infinitive VP because the head daughter of the dominating AP, namely the adjective easy, specifies that the inherited SLASH value in the infinitive complement is to be bound.

We should point out here that the idea of having tough-class adjectives subcategorize for a complement containing an accusative gap is borrowed from the GPSG analysis of tough-movement presented in Gazdar, Klein, Pullum, and Sag 1985 (hereafter, GKPS). And, like that analysis, our account predicts facts like those shown in (26):

```
(26)  a. John is easy to please ——.
 b. John is easy to persuade —— to be reasonable.
 c. John is easy to believe —— capable of doing something that stupid.
 d. John is easy to persuade Mary to kiss ——.
 e. John is easy to believe Mary would kiss ——.
 f. *John is easy to persuade Mary —— is capable of doing something that stupid.
```

Here the (f) and (g) examples are ungrammatical because the infinitive complements would have to contain nominative gaps.

There are a couple of significant differences between our analysis of weak UDCs and the GKPS analysis, however. First, the GKPS account runs into difficulties explaining why a case conflict does not arise between the subject and the gap in examples like (26) (see Hukari and Levine 1991 for some discussion). Our account avoids this problem because it requires identity only of the referential index, not of any syntactic features.

Second, on our analysis the trace is assumed to be pronominal, in the sense to be explicitly defined in Chapter 6; as explained there, our theory differs from GB in not requiring that a trace be pronominal. If the trace were an anaphor, Principle A of our binding theory (also set forth in Chapter 6) would wrongly require it to be coindexed with the subject of please in (26)a; and if it were a nonpronoun, coindexation with John in (26)a would constitute a violation of our Principle C.

And third, since the GKPS account does not permit set values for foot features, it disallows the possibility of a constituent containing more than one unbound trace. Thus it is unable to account for well-known 'multiple extractions' examples like those in (7), repeated here as (27): 9

```
(27)  a. [A violin this well crafted], even [the most difficult sonata], will be easy to play —— on ——.
 b. This is a problem which, John, is difficult to talk to —— about ——.
```

But in HPSG, since nonlocal features have set values, this problem does not arise. For example, the relative clause in (27b) will be assigned a structure like the one sketched in (28):

```
9. We do not attempt here to deal with nested dependency constraint effects such as the contrast between (27a) and (i):

(i)  *A sonata this simple, even the most crudely crafted violin would be easy to play on.
```

It is likely that a plausible solution is available in terms of a modification to the theory wherein the value of SLASH is treated as a list rather than a set, and TO-BIND values always coincide with the FIRST of the SLASH list (when this is an NP, at least). In order to develop this analysis, however, a number of technical issues involving parasitic gaps remain to be worked out.
(28) \[
\begin{array}{c}
S \\
\text{[INHER | SLASH ( ) ]}
\end{array}
\]

\[
\begin{array}{c}
\text{NP} \\
\text{[LOCAL (}) \\
\text{INHER | SLASH ( ) ]} \\
\text{TO-BIND} | SLASH ( ) \\
\text{which}
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \\
\text{[SUBCAT ( ) ]} \\
\text{INHER | SLASH ( ) ]} \\
\text{John}
\end{array}
\]

\[
\begin{array}{c}
\text{V} \\
\text{[SUBCAT ( ) ]} \\
\text{INHER | SLASH ( ) ]} \\
\text{is}
\end{array}
\]

\[
\begin{array}{c}
\text{A} \\
\text{[TO-BIND | SLASH ( ) ]} \\
\text{[INHER | SLASH ( ) ]} \\
\text{easy}
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \\
\text{[INHER | SLASH ( ) ]} \\
\text{to}
\end{array}
\]

\[
\begin{array}{c}
\text{V} \\
\text{[INHER | SLASH ( ) ]} \\
\text{[INHER | SLASH ( ) ]}
\end{array}
\]

\[
\text{talk to about}
\]

In this example the verb phrase *talk to ___ about ___* has two unbound traces. One, the object of the preposition to, gives rise to the INHER | SLASH element labelled 2; the other, the object of about, gives rise to the INHER | SLASH value element 3. The first of these becomes bound as specified on the adjective *easy*, and at the same time is coindexed with the subject of *easy*. Notice that this element is not inherited onto the AP node. The remaining SLASH element, the one labelled 3, is bound at the top of the tree, where it is identified with the local features of the relative pronoun, in accordance with the filler-gap rule that licenses the top node. At the top of the relative clause, the INHER | SLASH value is empty.

As noted by Maling and Zaenen (1982), there are severe constraints on where doubleton INHER | SLASH values can occur in English. On the basis of the ungrammaticality of examples like (29), Maling and Zaenen assume that English obeys a constraint preventing finite clauses from being doubly ‘slashed’:

\[
\text{(29) a. *Violins this well crafted, even the most difficult sonata will be easy to guarantee that Winnie can play ___ on ___].}
\]

\[
\text{b. *Violins made this badly, even the easiest sonata would be hard to make it possible for Winnie to play ___ on ___].}
\]

Although some such constraint is doubtless at work, the acceptability for many speakers of examples like (30) suggests that Maling and Zaenen’s formulation may not be correct:

\[
\text{(30) Someone that stupid, how much time [do we really want to waste ___ arguing with ___]?}
\]

Finally, we should point out that the lexically based analysis of the *easy-*class of expressions we have presented is consistent with the observation made by Elisabet Engdahl (personal communication, 1992) that languages vary considerably with respect to which predicates allow gap-containing complements. For a more complete discussion of how HPSG lexicons express the relevant generalizations about these and related expressions, see Flickinger and Nerbonne 1992.

4.4 The Trace Principle

One of the most important issues in modern syntactic theory has been the question of where traces can occur. This question has a long history, dating back to Ross’s dissertation (Ross 1967), and since that time a great many so-called island constraints have been proposed in the literature. But in recent years, many have attempted to deduce extraction islands from a very small number of universal principles, perhaps involving certain parameters of cross-linguistic variation. For example, in GB theory, the distribution of wh-traces is constrained by two principles, the Empty Category Principle (or ECP) and Subjacency (Chomsky 1981). It would be too much of a digression to discuss these two principles in detail here, but in essence the ECP says that a wh-trace has to occur in the right kind of local environment, while Subjacency says that the movement from trace position to the final landing site has to proceed step-by-step, with each step being permissibly small in some appropriate sense. Note that the effect of Subjacency is closely paralleled by that of the Nonlocal Feature Principle, which enforces a particular step-by-step constraint on the inheritance of all nonlocal information, including SLASH. In the next section, we will consider an additional principle that places further constraints on the local inheritance of SLASH, and hence functions as an even closer HPSG analog of Subjacency, but here we will direct our attention to a constraint on traces whose role in the theory is largely analogous to that of the ECP in GB.

The Trace Principle is stated as in (31):
(31) **TRACE PRINCIPLE:**

Every trace must be subcategorized by a substantive head.

For all the analogies between GB and HPSG, however, we believe that once one unravels the chains of definitions that the GB trace theory depends on, the HPSG theory of traces can be seen to be much clearer and simpler. There are also some important differences of a more specific nature. For example, unlike the GB account, there is nothing in our theory of universal grammar that rules out so-called *that*-trace sentences like the one in (32a). Given the fact that similar sentences in certain Scandinavian languages are grammatical, we think our approach is preferable. Relevant examples are the Icelandic sentence (32b) cited by Maling and Zaeecn (1982), and the Norwegian sentence (32c) cited by Engdahl (1983):

(32) a. *Who, did Kim claim that — left?*
    
b. Hver sagðir þu að varí kominn til Rekjavíkur.
    who(NOM) said you that was come to Reykjavik
    ‘Who did you say came to Reykjavik?’
    
c. Desse konstruksjonar trur eg at er meir
    these constructions think I that are more
    naturlege uttrykksmåtar.
    natural expressions
    ‘These constructions, I think (*that) are more natural expressions.’

To account for the badness of (32a), we simply propose a further English-

10. The locution ‘subcategorized by’ is here to be understood as ‘having its SYNSEM
    value on the SUBCAT list of.’

11. Our formulation of trace synchrony carries with it an implicit denial of the central
    claim embodied in analyses that appeal to the ECP as a constraint on logical forms—
    namely the claim that quantifier scope, trace binding, and negative particles in various
    Romance languages all exhibit certain ‘subject-object asymmetries’ that must be given
    a unified account in terms of a prohibition (the ECP) against binding from ‘too far
    away.’ Indeed, Ladusaw (n.d.) casts doubt on the semantic predictions made by the
    widely accepted analyses of Romance negative particles put forth by, e.g., Kayne
    (1984b) and Rizzi (1982). Ladusaw also presents a plausible alternative account of these
    data that makes different semantic predictions and that, as we have assumed here, separ-
    rates the account of negative particles from considerations of extraction phenomena.
    Similarly, as May’s (1984) discussion makes clear, the ECP can provide a treatment of
    quantifier scope asymmetries in English only at the cost of abandoning the idea that
    logical forms (for us, CONTENT values) represent relative scope information, a conse-
    quence that we regard as both unwarranted and unnecessary. For these reasons, we have
    formulated our Trace Principle as a constraint on the distribution of traces that is unre-
    lated to constraints on the timing or the interpretation of negative particles.

specific constraint on the distribution of traces, a constraint we will formulate
as an English-particular parametrization of the Trace Principle:

(33) **TRACE PRINCIPLE (parametrized for English):**

Every trace must be strictly subcategorized by a substantive head.

This modification adds the condition that traces not only be subcategorized,
but also be strictly subcategorized, in the sense that its SYNSEM value must be
a noninitial member of a substantive head’s SUBCAT list. We assume that
languages simply differ with respect to whether or not this further constraint
exists. The revised Trace Principle predicts the ungrammaticality of (32a), as that
example contains a subject trace.

There is still more to be said about subjects, of course, since nothing we have
said so far accounts for the grammaticality of apparent subject extractions like
the one in (34a):

(34) a. Who, did Kim claim — left?

The solution we suggest here, adapting an idea originally proposed by Gazdar
(1981), is that such sentences do not actually contain finite sentential comple-
ments at all. Instead, we analyze such examples as having a structure like the
one in (34b):

(34) b.

12. Note that we have said nothing thus far that would differentiate objects of case-
marking prepositions like of from subjects, since both involve initial members of SUB-
CAT lists. Hence we have not yet provided an account of why traces are permitted in
examples like (i):

(i) What were you thinking of —?

This deficiency will be remedied in the analysis of Chapter 9, where subject and non-
subject complements are selected by different features.
In order to license such structures, we posit the lexical rule sketched informally in (35):

(35) SUBJEXTR.

x

[(SUBCAT \{ \ldots , [unmarked] , \ldots \})]

INHER SLASH ( )

SUBCAT \{ \ldots , [SUBCAT \{ [LOC[ ]) \ldots ]

The idea here is that any English verb that subcategorizes for a nonsubject\(^{13}\) S complement gives rise to a new lexical entry that subcategorizes for an (unslashed) VP complement. Readers familiar with GPSG will recognize the similarity between this lexical rule and GKPS's SLASH Termination Metarule 2 (sometimes referred to as the Finite VP Metarule). Note in particular that SELR, like the GKPS metarule, applies only if the sentential complement in the input is compatible with the specification [MARKING unmarked], and hence produces no violations of Bresnan's (1972) 'Fixed Subject Condition.' Unlike the GKPS account, however, the outputs of this lexical rule will interact with the binding theory presented in Chapter 6 to provide an account of the strong crossover examples in (36):

(36) a. *Who, did he, say Mary likes ---?
   b. *Who, did he, say ---, likes Mary?

Under the assumption that gerund phrases are projections of V[ger] (i.e. that a gerund phrase like him stepping in front of me is also an instance of S[unmarked]), SELR will also provide an account of the fact (mentioned in section 3 of Chapter 3) that subjects of gerunds may also be extracted, as in (37):

(37) a. The man that I resented --- stepping in front of me in line.
   b. Which man did you resent --- stepping in front of you in line?
   c. It was Sandy that I resented --- stepping in front of me in line.

In these examples, the SELR output is a slashed form of resent that subcategorizes for a subject and an unsaturated gerund complement.

13. \(Y\) is a variable ranging over synsem objects. Hence the presence of \(Y\) in (35) guarantees that the S[unmarked] is not a subject. This apparently ad hoc restriction will be eliminated in Chapter 9.
(see Chapter 1, section 8); and in (41b) the trace is the head of its phrase, not a complement at all. The Trace Principle thus provides a unified account of numerous island constraint phenomena.

Two points implicit in these remarks warrant further discussion. First, though prenominal possessor traces are correctly ruled out by the Trace Principle, it is a fact (as noted by Grosu (1974)) that extraction of possessor phrases is quite generally impossible in English. This fact is illustrated by contrasts like the following:

(42)  
- a. I met a friend of Leslie’s.
- b. *Leslie’s, I met a friend of ...
- c. *Leslie, I met a friend of ...’s.

The appropriate generalization about possessor extraction is thus expressed by an English-specific constraint that disallows determiners (more precisely local objects whose HEAD value is *determiner*) as members of the values of SLASH. This correctly rules out (42b), where a trace would otherwise be possible as the object of *of*. Example (42c), on the other hand, is excluded on the same grounds as (40c), as it also involves a trace that is the complement of a nonsubstantive category, and hence is a violation of the Trace Principle.

The second further point we wish to raise concerns the treatment of modifiers like those in (43): 15

(43)  
- a. When did Pat eat dinner?
- b. When do you think Pat ate dinner?
- c. On Saturday, Dana will go to Spain.
- d. On Saturday, I think Dana will go to Spain.

On the basis of examples like these, it is commonly assumed that the grammar of traces must allow adverbial expressions to participate in unbounded dependency constructions, presumably binding a trace. This conclusion, if accepted, would require us to modify the Trace Principle in some way so as to allow noncomplement traces.

But the view that adverbs participate in unbounded dependency constructions, as noted by Cattell (1978), might be called into question by contrasts like the following:

(44)  
- a. When do you/they feel (that) Dana should get promoted?
- b. When do you/they think (that) Pat should get promoted?

The examples in (44) are all ambiguous—the sentence-initial adverb appears to modify either the matrix clause or the embedded clause. The matrix modification reading requires a slightly unusual contextualization, for example, a sort of psychiatric discussion of someone’s thoughts, beliefs, or feelings, but nonetheless is available in principle. This ambiguity would be predicted by a UDC analysis where the adverbial trace can appear in either matrix or embedded clause position.

However, as noted by Cattell, the examples in (45) appear to be unambiguous, allowing only matrix modification. This apparent lack of ambiguity is quite mysterious from the perspective of any UDC analysis of ‘dislocated’ adverbials, for, as examples like (46) show, the verbs in (45) cannot be said to disallow extraction in general:

(46)  
- a. Which position do you deny (that) Pat is qualified for ...?
- b. Which problems do they doubt (that) Dana can solve ...?
- c. This class, she doubts (that) Sandy will attend ...
- d. This country, I deny (that) Chris will find true happiness in ...

Cattell (1978: 62) also suggests that simply negating a sentence is sufficient to preclude the possibility of embedded modification, as the following contrast illustrates:

(47)  
- a. Why do they think/suppose/say/imagine (that) she killed him?
- b. Why don’t they think/suppose/say/imagine (that) she killed him?

He concludes that a syntactic solution to the contrast between (44) and (45) is unlikely.

Cattell argues further that there is a semantic basis for the contrasts in question here. The matrix verbs in (44) and (47a) differ from those in (45) (and the negated forms in (47b)) in terms of their ability to function as parenthetical tags, as illustrated in (48)–(49):

(48)  
- a. When should Dana get promoted, do you/they feel?
- b. When should Pat get promoted, do they think?
- c. Where will she camp, do you/they believe?

15. We are particularly indebted to Georgia Green, Larry Horn, Polly Jacobson, and Mark Liberman for helpful discussions about adverbial extraction.
d. On Tuesdays, Sandy will attend an aerobics class, she thinks.

e. In this country, Chris will find true happiness, I feel.

(49)  
a. *When should Pat get promoted, do you deny?  
b. *When should Dana get promoted, do they doubt?  
c. *When should Dana get promoted, don't they think?  
d. *On Tuesdays, Sandy will attend an aerobics class, she denies.  
e. *In this country, Chris will find true happiness, I doubt.

It seems that the ability to perform a parenthetical-like function may play an important role in the explanation of the ambiguity differences in (44)–(45).

The problem of providing an exact characterization of the distinction that is relevant to an account of these data is by no means trivial, however. For example, the correlation between parenthetical tags and the possibility of embedded modification may be imperfect. Nonetheless, on the basis of facts such as these, one might attempt to develop a traceless analysis of adverb extraction, where there is no true modification of the embedded clause.16 Although Cattell (1978: 66ff.) seeks to explicate the relevant semantic and contextual properties in terms of the notion volunteered stance, he offers no mechanism to make sense of his observed correlations and gives no precise formulation of any alternative analysis.

However, the essential ingredients for such an alternative analysis were suggested by Liberman (1973). Building on unpublished work by Bresnan (1968), Liberman offered an analysis of examples like those in (50) and (51):

(50) One more step and I’m afraid I’ll have to shoot you.

(51) One more beer and I’m afraid it looks like I’m going to have to drive you home.

The problem noted by Liberman is the following. The semantic analysis of sentences of the form One more X and Y is basically that of a conditional ‘if r then Q,’ with contextual information interacting with the content of one more X to determine P, and the content of Y determining Q. But in examples like these, Q is determined not by Y, but rather by the complement of Y, with the highest clause in Y not making a direct contribution to Q, even though it is clearly an essential part of the matrix syntactic environment. Let us refer to such cases as ‘annotations,’17 distinguishing matrix annotations like those considered by Bresnan and Liberman from parenthetical annotations of a familiar sort.

16. For additional arguments against treating sentence-initial adverbials in terms of trace binding in GB, see Hegarty 1990.
17. We owe this term to Mark Liberman (personal communication, 1990).

Liberman’s solution to the problem of matrix annotations, cast in a generative semantics–like framework, posited a transformational rule that inserts annotative material into the matrix syntactic environment, thus deriving the second conjunct in (50) (I’m afraid I’ll have to shoot you) from a semantic structure like that of I’ll have to shoot you, with the additional material (I’m afraid) external to the underlying semantic structure of the clause. The semantic structure for the entire sentence (50) then, once contextual material is supplied, is essentially that of ‘if you take one more step, then I’ll have to shoot you,’ with the ‘I’m afraid’ annotation being just that—an annotation external to the semantic content.

Of course the insight of Liberman’s analysis could be preserved even if one discards the transformational assumptions of his presentation. The essential claim being made, one that could plausibly be expressed equally well with lexical rules, is that a large class of expressions (e.g. think, feel, and believe) may receive an annotative interpretation when they occur in syntactic matrix position. For example, a verb like think, even while it occurs in the environment shown in (52a), may give rise to an interpretation like the one sketched in (52b):

(52) a. \[ \begin{array}{c}
\text{think} \\
\text{NP} \\
\text{VP} \\
\end{array} \]

This analysis would thus allow an interpretation of (52a) that can be paraphrased by They like you, I think or They, I think, like you. The paraphrase is likely to be only approximate, however, given that there is no reason to believe that matrix and parenthetical annotations have identical interpretations (or constraints).

Once structure-interpretation mismatches of this sort are accepted, the Bresnan/Liberman-style of analysis allows a biclausal, adverb-initial sentence like (53) to be assigned two interpretations in simple consequence of the two possible interpretations of the matrix embedding environment (I feel/think):

18. We use the attribute QUALIFICATION as an expository convenience, in an attempt to finesse the host of murky conceptual and representational problems that surround the construction of a precise theory of annotations, a theory that would presumably include an account of various prosodic phenomena and discourse particles in addition to matrix and parenthetical annotations.
The ambiguity of adverb-initial sentences like (53), which can be analyzed simply in terms of Schema 5 (the head-adjunct schema), is thus explained without any appeal to adverbal traces. If an anaphoric interpretation is imposed, then the content of $s[fin]$ will be the content of the complement, and the modifier in (53) will appear to modify the embedded clause. On the other hand, if the familiar type of interpretation is assigned to (53), then the adverb will modify the psa whose relation is determined by the matrix verb. In this latter case, the verb’s complement will function as an argument of the relation of the matrix verb.

By contrast, the lack of ambiguity in examples like (54) follows directly from the well-known semantic/pragmatic fact that verbs like deny and doubt cannot function anaphortically, as illustrated in (55):¹⁹

\[
\begin{align*}
(54) & \quad a. \text{ When did they deny (that) Pat should get promoted?} \\
& \quad b. \text{ Where did you deny (that) Chris will find true happiness.} \\
(55) & \quad a. * \text{ When should Pat get promoted, do you deny?} \\
& \quad b. * \text{ When should Dana get promoted, do they?}
\end{align*}
\]

Attractive though the Bresnan/Liberman-style of analysis may seem, we doubt that it can ultimately tell the whole story about adverbial extraction in English. Although it seems clear that there is some relation between parentheticality and the ambiguity differences in (44)–(45), the possibility of adverbial extraction from an embedded clause cannot always be tied to the parentheticality of the matrix environment. Examples like (56) illustrate this point:²⁰

\[
\begin{align*}
(56) & \quad a. \text{ When their parents are in town next week, I doubt that the twins will attend any lectures.} \\
& \quad b. \text{ When their parents are in town next week, I don’t for a moment think that the twins will attend any lectures.} \\
& \quad c. \text{ When he was president, Reagan denies that he ever received kickbacks from the PLO.}
\end{align*}
\]

¹⁹ The analysis being considered here shares properties with that developed independently by Hegarty (1990), though there are minor differences in assumptions about the relevant data. Hegarty, for example, assumes that the presence of a that complementizer prevents the possibility of embedded modification.

²⁰ Here we are indebted to Charles Lee and Chris Manning for persistent construction of counterexamples.

Unlike the essentially similar examples in (45), which appear to differ only inasmuch as they are wh-questions rather than topicalizations, the examples in (56) allow the extracted adverbials to be interpreted as embedded modifiers. Given the grammaticality of such examples, it would seem to be an unavoidable conclusion that the grammar of English must allow matrix adverbials to modify semantically embedded complement clauses. There may be complex semantic and pragmatic factors contributing to the difficulty in assigning a complement-modifying interpretation to Cattell’s examples in (45), but in the right context even these may prove to be acceptable.²¹ The grammar of English cannot exclude adverbial extractions from the class of unbounded dependencies.

   In this section we will not revise our analysis of UDCs so as to allow for the possibility of such embedded modification. Rather, we will take this matter up again in Chapter 9, once certain modifications of the mechanism for introducing INHER | SLASH dependencies have been presented.

²¹ As an example of a semantic constraint, we would suggest that only those sential adjuncts can be extracted that are monotonic in the sense of preserving the entailments of the sentences to which they adjoin. This explains, for example, why probably cannot be extracted, even across a bridge verb: *Probably, Kim thought that Sandy stole the books (on the embedded modification reading).

Pragmatic constraints that bear on the extraction of (inter alia) adjuncts are exemplified by Kroch’s (1989) proposal to block certain ‘long’ movements (Rizzi 1990; Cinque 1990; Szabolcsi and Zwarts 1991) by presupposition failure. On this approach, with which we are sympathetic, the unacceptability of much-discussed examples like (i), involving extraction of wh-amount phrases out of complement questions, stems from the difficulty of constructing a context in which the presuppositions of the question are met:

\[
\begin{align*}
\text{(i)} & \quad \# \text{How much did you wonder whether the book cost? [cf. That much, I wonder whether any book could cost.]} \\
\text{(ii)} & \quad \# \text{How did you wonder why Sandy worded the letter? [cf. That rudely, I wonder why anyone would ever word a letter.]} \\
\text{(iii)} & \quad \# \text{How did Kim not know who fixed the disk drive? [cf. With nothing but a crowbar and a ballpeen hammer, I don’t know who could fix this disk drive.]} \\
\end{align*}
\]

Thus (i) is acceptable only under the highly improbable presupposition that there exists a price $p$ (say, $59.95$) such that the addressee wonders whether the book in question costs $p$. Indeed, we believe this approach can be extended to cover a rather wide range of other types of extractions, both of adjuncts and of ‘nonreferential complements,’ as exemplified in (ii) and (iii):
4.5 Parasitic Gaps

In section 4.3 we considered double gapped examples where the two traces are independent in the sense that they are bound at different places in the structure. However, it can also happen that two or more traces in the same sentence are bound in unison, as shown by the examples in (57):  

(57)  
   a. That was the rebel leader who, rivals of ——, shot ——. 
   b. [Those boring old reports], Kim filed —— without reading ——. 
   c. Which of our relatives, should we send snapshots of ——, to ——? 
   d. [Someone as vain as Sandy], would be easy to sell pictures of —— to ——. 

In examples like this, nothing special needs to be said. These are simply cases where two distinct traces happen to structure-share their LOCAL values. For example, the relative clause in (57a) will be assigned the structure sketched in (58):

(58)

\[
S \quad \text{[INHER SLASH (\_)]} \\
\quad \lfloor \text{LOCAL [\_]} \rfloor S \quad \text{[INHER SLASH (\_)]} \\
\quad \lfloor \text{TO-BIND SLASH (\_)} \rfloor who \\
NP \quad \text{[INHER SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor \\
NP \quad \text{[INHER SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor \\
\text{PP} \quad \text{[INHER SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor rivals \\
P \quad \text{[LOCAL [\_]} \text{SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor \\
NP \quad \text{[LOCAL [\_]} \text{SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor \\
\text{VP} \quad \text{[INHER SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor shot \\
P \quad \text{[LOCAL [\_]} \text{SLASH (\_)]} \\
\quad \lfloor \text{INHER SLASH (\_)} \rfloor \\
qf \\
\]

22. We would like to thank Janet Fodor, Polly Jacobson, and Bob Levine for extremely valuable discussions about the analysis of parasitic gaps.

The key point here is that since the two traces have the same LOCAL structure, tagged [\_], the subject NP and the VP have the same INHER \ SLASH value, namely the singleton set containing the local structure [\_]. It follows from the Nonlocal Feature Principle (stated in (16) above) that the INHER \ SLASH value of the lower S must be the same also. When this single SLASH value element is identified with the local structure of the relative pronoun who, both traces are bound in unison.

It is well-known that certain traces that are coindexed with another trace can occur in positions where a trace could not otherwise appear, as long as the other trace could have occurred independently. Such traces are known as parasitic gaps, and the other, nonparasitic gap is said to license the parasitic gap. In cases like this we will refer to the licensing gap as the host gap. For example, in (58), the trace within the subject is the parasite, and the object trace is the host. This is because a trace could appear independently as an object but not within the subject, as the facts in (59) show:

(59)  
   a. *That was the rebel leader who, rivals of ——, shot the British consul. 
   b. That was the rebel leader who, agents of foreign powers shot ——.

Likewise, it has often been assumed (Huang 1982; Gazdar et al. 1985) that examples like (60a) are ungrammatical and hence that in the case of sentences like (60b), the trace within the adjunct is a parasite and the object trace is its host:

(60)  
   a. [Those boring old reports], Kim went to lunch without reading ——. 
   b. [Those boring old reports], Kim filed —— without reading ——.

GKPS propose a general account of facts such as these: SLASH is both a FOOT feature (one that must be instantiated onto the mother of a local tree if it is instantiated onto a daughter—in accordance with their FOOT Feature Principle) and a HEAD feature (one that must be instantiated onto a head daughter if it is instantiated onto a mother in a local tree—in accordance with their HEAD Feature Convention). The interaction of these two principles of GPSG is sketched in (61):

23. In spite of the fact that examples such as (60a) have been treated as ungrammatical in much of the syntactic literature, they are clearly better than ones like (59a). For many speakers, the authors included, examples like (60a) seem completely acceptable.
Unbounded Dependency Constructions

(61) a. Rule: \( V[\text{BAR} \ 2] \rightarrow NP, H[\text{BAR} \ 1] \)

b. 
\[
\begin{array}{c}
\text{V} \\
\text{NP} \\
\text{SLASH NP} \\
\text{BAR 1, SLASH NP} \\
\text{BAR 2, SLASH NP} \\
\text{rivals of} \\
\text{shot the British consul}
\end{array}
\]

c. 
\[
\begin{array}{c}
\text{V} \\
\text{NP} \\
\text{SLASH NP} \\
\text{BAR 1, SLASH NP} \\
\text{BAR 2, SLASH NP} \\
\text{rivals of} \\
\text{shot}
\end{array}
\]

d. 
\[
\begin{array}{c}
\text{V} \\
\text{NP} \\
\text{SLASH NP} \\
\text{BAR 1, SLASH NP} \\
\text{BAR 2, SLASH NP} \\
\text{agents of foreign} \\
\text{powers} \\
\text{shot}
\end{array}
\]

Given the rule in (61a), the structure in (61b) satisfies the FOOT Feature Principle (the precursor of HPSG’s Nonlocal Feature Principle), but not the Head Feature Convention (the precursor of HPSG’s Head Feature Principle), because the head daughter (the \( V[\text{BAR} \ 1] \)) lacks the nonempty specification for the HEAD feature SLASH that is borne by the mother. Hence (59a) and examples like it where only a nonhead daughter contains a gap are predicted to be ungrammatical. The structures in (61c) and (61d), on the other hand, satisfy both relevant grammatical principles. Since SLASH is a HEAD feature, instantiating a nonempty SLASH specification onto the mother necessitates instantiation of the very same specification on the head daughter, and nothing prevents instantiating the very same specification onto the nonhead daughter, as in (61c). In this way, the GKPS treatment of SLASH as both HEAD and FOOT feature renders the presence of \([\text{SLASH NP}] \) on the subject NP parasitic on the simultaneous presence of \([\text{SLASH NP}] \) on the head \( V \). The GKPS account of contrasts like (60) is similar.

The GKPS analysis of extraction phenomena is perhaps best-known for its explanation of “across-the-board” exceptions to Ross’s (1967) Coordinate Structure Constraint (CSC), a long-standing source of difficulty for transformational analyses of coordination and extraction.24 Briefly, because each conjunct in a local coordinate structure is treated as a head daughter, it follows from GKPS’s Foot Feature Principle that an instantiated nonempty SLASH specification on any daughter will also be present on the mother. And if a nonempty SLASH specification is instantiated on the mother of any structure, then that same specification must be instantiated on the head daughter(s) (in accordance with the Head Feature Convention). Since all conjuncts are taken to be head daughters, it follows that if any conjunct is slashed, then all conjuncts must be slashed. Since being slashed corresponds to containing a gap, this consequence is equivalent to the claim that all conjuncts will contain a gap if any conjunct does, which is precisely Ross’s CSC, together with its across-the-board exceptions. The GKPS analysis of filler-gap constructions thus derives the CSC (with its across-the-board exceptions) and the phenomenon of parasitic gaps from a single underlying principle: the principle that SLASH is both a HEAD feature and a FOOT feature.

Elegant though this theory may appear, it suffers from at least one very serious difficulty. As noted by Sells (1984: 307ff.), Swedish differs from English in allowing extractions out of subjects that are not parasitic on the presence of another gap in the sentence. Thus examples like (62), whose structure is identical in all relevant respects to the one in (61b) (which we saw to be ungrammatical in English), are grammatical in Swedish (subject to certain pragmatic conditions):

(62) den deckare som de sista sidorna i hade
that detective novel that the last pages in had
kommit bort
come away
‘that detective novel whose last pages had come away . . .’

But Swedish observes the CSC, subject to the very same across-the-board exceptions as English. The GKPS theory, where SLASH is crucially classified as a HEAD feature, apparently predicts that a language like Swedish is impossible. If SLASH is a HEAD feature in Swedish, Sells argues, then examples like (62) should not be grammatical, but they are. And if SLASH is not a HEAD feature in Swedish, then Swedish should not obey the CSC, but it does. Either way, the claim that SLASH is doubly classified incorrectly links the phenomenon of parasitic gaps with the CSC and must be abandoned.

We propose, following Sells, to treat the CSC as deriving from a constraint that is particular to coordinate structures, and that in no way involves the assumption that conjuncts are head daughters in those structures. This analysis, which we discuss in section 4.6, leaves the matter of parasitic gaps to an inde-

24. For a variety of as yet unanswered arguments demonstrating the superiority of nontransformational analyses of coordination, see Gazdar, Pullum, Sag, and Wasow 1982.
pendent analysis that must deal with cross-linguistic differences such as the one observed by Sells.

But what then is the basis for the analysis of parasitic gaps? One answer to this question might be the introduction of a second universal principle, which we shall refer to as the SLASH Inheritance Principle. This principle, stated in (63), places a further constraint on the inheritance of SLASH value elements:

(63) **SLASH Inheritance Principle (SIP):**

Every member of the INHER | SLASH set on a headed constituent must be inherited from (i.e. belong to the INHER | SLASH set of) a daughter that is either (a) (strictly) subcategorized by a substantive head, or (b) the head.

SIP would play much the same role in HPSG that is played by Subjacency in GB, in the sense that it places a condition on each step of the unbounded dependency. Readers familiar with GB will also recognize the close similarity of SIP to Huang’s (1982) Condition on Extraction Domains, or to Kayne’s (1983) Connectedness Condition (especially as modified by Engdahl (1983)). The main difference is our clause (b) allowing inheritance from the head daughter. This is needed, for example, in order to pass SLASH values up from VP to S. (In then-current GB, of course, this was unnecessary because traces move straight from the topic position of one S to the topic position of the next S up, without having to be passed up one node at a time.)

Let us now return to the question about parasitic gaps that we raised earlier without answering: *Why should traces that would otherwise be impossible become grammatical when licensed by another trace in a normal trace position?* That is, how do we account for the contrast between the grammatical sentence (57a), repeated here as (64a), and the ungrammatical (59a), repeated here as (64b)?

(64)  

a. That was the rebel leader who, [rivals of —, shot —].

b. *That was the rebel leader who, [rivals of —, shot the British consul].

The contrast in (64) is predicted by the interaction of the Nonlocal Feature Principle with the Trace Principle and SIP. For reference, this set of principles is recapitulated in (65):

(65) **Summary of HPSG trace theory (preliminary version):**

a. Nonlocal Feature Principle: For each nonlocal feature, the INHERITED value on the mother is the union of the INHERITED values on the daughters minus the TO-BIND value on the head daughter.

b. Trace Principle (English): Every trace must be strictly subcategorized by a substantive head.

c. SLASH Inheritance Principle: Every member of the INHERITED |
Both of the traces in (66a) are strictly subcategorized, and hence the Trace Principle is satisfied. But what about SIP? The key node to consider is the top node, whose INHER \slash SLASH set contains exactly one element, labelled 1. This node satisfies SIP since the element 1 is also contained in the INHER \slash SLASH set of the VP head daughter. The fact that this element also appears in the INHER \slash SLASH set of the subject has no bearing on the grammaticality: all the conditions of the theory are satisfied even though the subject is not strictly subcategorized. Thus (66a) is predicted to be grammatical.

But now consider (66b). The trace is strictly subcategorized, so that the Trace Principle is satisfied. The problem is with SIP. Again, the key node to consider is the top one. As in (66a), the INHER \slash SLASH set contains one element, labelled 1. But now SIP is not satisfied, for the only daughter that this element is inherited from is the subject. But the subject, of course, is neither strictly subcategorized nor the head. Hence we predict that (66b) is ungrammatical. To put it informally, the gap from the subject in (66a) is parasitic on, or licensed by, the independently permissible object gap, but the one in (66b) is not.

It should also be noted that this analysis of parasitic gaps avoids a further defect of the GKPS analysis pointed out independently by Pollard (1985), Hulke and Levine (1987a), and Jacobson (1987). The GKPS analysis, which relies crucially on the stipulation that SLASH is both a HEAD feature and a FOOT feature, fails to predict the following contrast:

(67)  
\begin{itemize}
  \item a. *Who did my talking to ___ bother Hilary?
  \item b. Who did my talking to ___ bother ___?
\end{itemize}

The problem for the GKPS analysis is that nothing makes a gap within the subject of the S(+INV) dependent on the presence of a gap in the VP, since the latter is not a phrasal head.

On our analysis, illustrated in (68), the relevant structure (i.e. the lower S) is an instance of Schema 3:

(68)

Here, the subject of the inverted auxiliary verb is not strictly subcategorized and hence can contain a gap only if the VP complement (which is strictly subcategorized) also contains a gap, as guaranteed by SIP. Thus the indicated local tree in (68) is ruled out by our theory (but not by the theory of GKPS), and examples like (67a) are correctly deemed ungrammatical.

A number of researchers working within the tradition of phrase structure grammar have pointed out a set of examples that are potentially problematic for the theory of SLASH inheritance we have been considering. For example, contrasts like those in (69) and (70) are discussed by Farkas et al. (1983), by Sells (1984: 309), and by Jacobson (1984: 415):

(69)  
\begin{itemize}
  \item a. *Who did you consider friends of ___ angry at Sandy?
  \item b. Who did you consider friends of Sandy angry at ___?
  \item c. Who did you consider friends of ___ angry at ___?
\end{itemize}

(70)  
\begin{itemize}
  \item a. *Here's the jerk that I expected my pictures of ___ to bother you.
  \item b. Here's the jerk that I expected my pictures of you to bother ___.
  \item c. Here's the jerk that I expected my pictures of ___ to bother ___.
\end{itemize}

These researchers have argued that examples like (69)–(70) behave like the parasitic gap examples we have been considering; but in (69)–(70) neither the GKPS analysis nor the one based on SIP correctly predicts the observed contrasts. This is illustrated in (71), the structure of the relevant piece of (70a):

(71)  

25. If nonpredicative prepositions are treated as intransitive (i.e. the prepositional object is the only member of the preposition's SUBCAT list), then we have not yet explained why the trace that follows of in (66) is strictly subcategorized. In fact, as noted in section 4.4, we have not yet provided a schema for introducing prepositions that take an object but no subject. We return to this problem in Chapter 9.
The deviance of (71) follows from nothing in the theory we have been considering: the object NP from which the mother inherits is strictly subcategorized and hence should by itself legitimate SLASH inheritance. And there is nothing about the infinitival VP here (since it is not the head) that would predict that inheriting from it is sufficient to restore acceptability, as in (72), the structure of the relevant piece of (70c):

\[
(72) \quad \text{VP} \quad \text{[INHER | SLASH (I)]}
\]

\[
\text{V} \quad \text{[INHER | SLASH (I)]} \quad \text{[INHER | SLASH (I)]}
\]

\[
\text{NP} \quad \text{expected} \quad \text{my pictures of} \quad \text{to bother}
\]

Within the theory of complementation presented in Chapter 3, the problem posed by (69) is the same in all relevant respects.

The solution to this problem that is suggested by Farkas et al. (1983), Jacobson (1983, 1987), and Sells (1984) builds on ideas of Kuno (1973), who argued that English has a constraint legislating that no medial constituent may contain a gap (i.e. no constituent that is nonfinal in its phrase may be ‘incomplete’). Kuno’s formulation of this constraint did not allow for parasitic gaps (as noted by Jacobson (1987: 415)), but the desired effect—to allow medial phrases containing gaps just in case they are followed by some phrase that also contains a gap—can be obtained by formulating Kuno’s (Clause Nonfinal) Incomplete Constituent Constraint as a linear precedence rule like (73). (We have introduced a minor modification of the Farkas et al./Sells/Jacobson proposal to induce compatibility with other aspects of the HPSG analysis of UDCs.)

\[
(73) \quad \text{INCOMPLETE CONSTITUENT CONSTRAINT:}
\]

\[
[\text{INHER | SLASH empty-set}] < [\text{INHER | SLASH nonempty-set}]
\]

The effect of (73) is to block structures like (71), where a slashed daughter precedes an unslashed daughter, but to allow structures like (72), where a medial slashed daughter precedes another slashed daughter. This formulation of the Incomplete Constituent Constraint is intended to be a complete theory of parasitic gaps, effectively replacing the GKPS theory of SLASH as a head feature or any other alternative such as SIP.

Note that (73) also correctly predicts the behavior of parasitic gaps in sentential structures, whether they are inverted, as in (74a), or un inverted, as in (74b):

\[
(74) \quad \text{a. Parasitic Gaps}
\]

\[
\text{NP} \quad \text{[INHER | SLASH (NP)]}
\]

\[
\text{V [+INV]} \quad \text{did} \quad \text{my talking to} \quad \text{bother} \quad \text{* Kim /}
\]

\[
\text{(who)}
\]

\[
(74) \quad \text{b. Parasitic Gaps}
\]

\[
\text{NP} \quad \text{[INHER | SLASH (NP)]}
\]

\[
\text{NP} \quad \text{[INHER | SLASH (NP)]}
\]

\[
\text{V} \quad \text{[INHER | SLASH (NP)]}
\]

\[
\text{(who)} \quad \text{my talking to} \quad \text{bothered} \quad \text{* Kim /}
\]

\[
\text{(who)}
\]

In both these examples, the structures are correctly predicted to be well-formed just in case the VP daughter is slashed.

One important difference should be noted from the outset between the account of parasitic gaps based on the SLASH Inheritance Principle and the one based on the Incomplete Constituent Constraint that we have just sketched. The former account rules out extraction from adverbial modifiers (when not accompanied by extraction from a head daughter), while the latter theory allows these.\textsuperscript{26} In light of the grammaticality of a large number of examples of this type, for example, those in (75), the facts would seem to provide strong arguments against most extant analyses of parasitic gaps (including that of GKPS), and strong support for the Farkas et al./Sells/Jacobson reanalysis based on linear precedence (LP) constraints:

\[
(75) \quad \text{a. That's the symphony that Schubert died [without finishing __].}
\]

\[
\text{b. Which room does Julius teach his class [in __] ?}
\]

\[
\text{c. Who did you go to Girona [in order to meet __]? (Hegarty 1990)}
\]

\[
\text{d. What kind of wagon did they use to ride to school [in __]?}
\]

\[
\text{e. How many of the book reports did the teacher smile [after reading __]?}
\]

\[
\text{f. This is the blanket that Rebecca refuses to sleep [without __].}
\]

\textsuperscript{26} See n. 23.
Whatever constraints may affect the possibility of extraction from adverbial modifiers like these, they are considerably more subtle than the absolute constraints on grammatical structures considered thus far.27

There are, however, a number of problems facing any attempt to explain English parasitic gaps in terms of the Incomplete Constituent Constraint. First, as noted by Sells (1984: 319–320), there is a potential problem having to do with traces, which, although analyzed as slashed constituents, fail to obey the LP rule in (73):

(76) Which council members did you [persuade ___ to support the resolution]?

Sells’s solution is to stipulate that [+NULL] elements are immune to LP rules; Jacobson’s (1987: 416) is to eliminate the analysis of traces as slashed constituents, a proposal we also explore in Chapter 9.

A second problem is that the formulation in (73) disallows structures like the one in (77b), which are required in order to account for the grammaticality of examples like (77a):

(77) a. Those boring old reports, Kim wrote critiques of ___ without falling asleep.
   b. [INHER SLASH NP]
      VP
         [INHER SLASH NP]
      VP
         [INHER SLASH ___]
   wrote critiques of ___ without falling asleep

We might seek to remedy this problem by modifying (73) so that slashed daughters are not forbidden to precede all nonslashed daughters, but only unslashed complement daughters.28

27. As pointed out to us by Joan Bresnan, similar observations were made in unpublished work by Rothstein (1981). See also Grosz 1972 for some relevant discussion.

28. For a presentation of a linear precedence theory whose rules may make reference to notions such as ‘complement,’ see Sag 1987 and P&S-87, Chapter 7. For a subtly different reformulation of this constraint with similar consequences, see Sells 1984: 315ff.

(81) [INHER SLASH NP]
   VP
      [INHER SLASH ___]
      PP
         [INHER SLASH ___]
   [INHER SLASH ___]
   argue with ___ about politics

29. Sells and Jacobson both offer examples of equi verbs that they claim obey the same parasitic gap behavior as raising verbs like *consider* and *expect*. However, we do not consider examples like (i) less acceptable than Fodor’s example (116c):

(i) I don’t know which children you ordered the parents of ___ to stop disturbing the teachers.
Assuming that the data are as we have just described them to be, there is a straightforward solution possible within HPSG. The basic intuition is very simple: all the cases of true parasitic gaps are contained within subjects. That is, assuming (1) that extraction from adjuncts is in principle grammatical and (2) that incomplete nonsubject phrases like those discussed by Hukari and Levine and Fodor are grammatical, then the only cases of SLASH inheritance to be blocked are those where a SLASH value has been inherited from a subject (where this notion must include ‘raised’ subjects like those in (69a) and (70a)) without that SLASH value also being inherited from another daughter. The facts are then predicted by replacing SIP with the following constraint: 31

(84) Subject Condition:

The initial element of a lexical head’s SUBCAT list may be slashed only if that list contains another slashed element.

The Subject Condition predicts the familiar contrast in (85):

(85) a. *That was the rebel leader who rivals of ___ assassinated the British consul.
    b. That was the rebel leader who rivals of ___ assassinated ___.

This is because only in (85b) does the SUBCAT list of the verb assassinated satisfy (84). Similarly, the contrast between (86a) and (86b) (discussed above as a problem for the GKPS analysis of parasitic gaps) is accounted for:

(86) a. *Who did my talking to ___ bother Hillary?
    b. Who did my talking to ___ bother ___?

In this case, only in (86b) do the SUBCAT lists of did and bother satisfy (84). By stating the Subject Condition in terms of SUBCAT lists, the superficial difference between inverted and uninvited clauses is correctly ignored. 32

The analysis of the Farkas et al./Sells/Jacobson examples in (69) and (70) is

31. It should be noted that our Subject Condition is weaker than the condition of the same name in Chomsky 1973. In Chapter 9, we propose a minor reformulation of (84).
32. We have discovered a number of speakers who accept examples like (85a) and (86a). Such varieties could be described simply by eliminating the Subject Condition. For some of these speakers, however, there appears to be a contrast between (86a), b), both of which are judged to be acceptable, and (i) and (ii), which are judged to be unacceptable:

(i) Who did rivals of ___ assassinate the British Consul?
(ii) Who did rivals of ___ assassinate ___?

At present, we have no account of this variety.
more subtle. Here the Subject Condition crucially affects the SUBCAT list of the embedded verb heading the VP complement, rather than that of the matrix verb. Because raising is analyzed as structure sharing between a SUBCAT element and the unexpressed subject of another SUBCAT element (section 3.5), it follows that a raised slashed object will always give rise to an unexpressed slashed subject that must satisfy the Subject Condition, as shown in (87):

In (87), the subject of bother, labelled [2], is a synsem object with a nonempty INHER|SLASH value, whereas [3] is a synsem object with an empty INHER|SLASH value. The SUBCAT list of bother in (87) thus violates the Subject Condition. The minimally different structure in (88), on the other hand, produces no such violation:

Here, the fact that the object of bother is slashed renders that verb's SUBCAT list compatible with (84).

In addition, we now have an account of the differences between English and Swedish noted by Sells (1984: 311ff.): Swedish simply lacks the Subject Condition. In consequence of this fact, not only are examples like (62) above grammatical in Swedish, but so are examples like (89), whose analogs in English violate that constraint:

The analysis just sketched, perhaps surprisingly, also provides an account of the behavior of Hukari and Levine's (1991)'certain heros' examples:

This should be compared with the configurationally similar (82), where the higher VP is headed by an equi verb, rather than a raising verb. As a consequence, the subject SUBCAT element on the head verb (make) of the lower VP is merely coindexed with the matrix object and therefore does not share its nonempty SLASH value.

34. As pointed out to us by Manfred Sailer and Tilman Höhle (personal communication, 1992), our Subject Condition wrongly rules out the extraction of objects of raising-to-object verbs, e.g. I wonder who he expected to win. We provide a solution to this problem in Chapter 9, section 5.1, footnote 38.
Examples like these are argued by Hukari and Levine to be problematic for the Incomplete Constituent Constraint proposed by Jacobson. The problem is that, on virtually any analysis of adjectives like easy (see section 4.3), the AP very easy to listen to ___ (in spite of containing a slashed VP) is itself slashed. Hence a VP like (91), which must be part of the structure of the grammatical (90a, b), contains a slashed NP followed by an unslashed phrase, in violation of the Incomplete Constituent Constraint:

(91)

\[ \begin{array}{c}
\text{VP} \\
\left[ \text{INHER} | \text{SLASH} \{ 1 \} \right]
\end{array} \]

\[ \begin{array}{c}
\text{NP} \\
\left[ \text{INHER} | \text{SLASH} \{ 1 \} \right]
\end{array} \]

\[ \begin{array}{c}
\text{AP} \\
\left[ \text{INHER} | \text{SLASH} \{ 1 \} \right]
\end{array} \]

\[ \text{finds long stories about ___} \]

\[ \text{very easy to listen to ___} \]

Why are such examples grammatical? The theory we have sketched provides a clear answer to this question: because adjectives like easy subcategorize for a slashed complement, a raised, slashed object like the one in (91) will be identified with an unexpressed subject that is less oblique than a slashed element, as shown in (92):

(92)

\[ \begin{array}{c}
\text{VP} \\
\left[ \text{INHER} | \text{SLASH} \{ 1 \} \right]
\end{array} \]

\[ \begin{array}{c}
\text{AP} \\
\left[ \text{INHER} | \text{SLASH} \{ 1 \} \right]
\end{array} \]

\[ \begin{array}{c}
\text{NP} \\
\left[ \text{INHER} | \text{SLASH} \{ 1 \} \right]
\end{array} \]

\[ \text{finds long stories about ___} \]

\[ \text{very ___} \]

\[ \text{easy to listen to ___} \]

Since the synsem objects labelled [2] and [3] in (92) are both slashed, it follows that the SUBCAT list of easy in (92) satisfies the Subject Condition.35

We conclude this discussion of parasitic gaps by considering the facts in (93):

(93)

\[ a. \] I never know \{ which topics \}, jokes about ___; are likely to offend people.

\[ b. \] I never know \{ which people \}, jokes about ___; are likely to offend ___.

\[ c. \] People that sensitive, I never know \{ which topics \}, jokes about ___; are likely to offend ___.

Like all theories of parasitic gaps with which we are familiar, our account predicts the contrast between (93a) and (93b). In (93a) we have an unlicensed subject-internal gap; but in (93b) the subject-internal gap is licensed by the coindexed object gap. However, nothing in our theory requires that the licensing gap be coindexed with the parasitic gap. Thus our theory also correctly predicts the perhaps surprising grammaticality of (93c).

In sum, a more empirically and theoretically adequate account of parasitic gaps is provided by preserving the essence of the GKPS account (constraints on the inheritance of SLASH specifications in local structures), but abandoning 35. A remaining puzzle, pointed out to us by Polly Jacobson, is that phrases containing easy-type APs also ameliorate sentences containing subject-internal gaps, as in (i) and (ii):

\[ \text{(i) } \] There are certain heroes that \{ long stories about ___ | are always very easy to listen to ___ \}.

\[ \text{(ii) } \] There are certain heroes that \{ long stories about ___ | are too boring to listen to ___ \}.

Examples like these are counterexamples to the GKPS account of parasitic gaps (SLASH is present on the nonhead daughter of S (the subject), but not on the phrasal head (the VP) as well as the account stated in terms of the Incomplete Constituent Constraint. Unfortunately, they are also counterexamples to our theory, inasmuch as they violate the Subject Condition. This is because, even though the SUBCAT lists of the embedded predicates (easy and boring) in these examples are well-formed, it turns out that in both cases the SUBCAT list of the auxiliary are is not. For instance, in (i) this SUBCAT list is of the form (iii):

\[ \text{(iii) } \] (NP[INHER | SLASH {1}]), AP[INHER | SLASH {1}]

Thus the gap in the AP is already bound, and therefore (on our account) should not be able to license the subject-internal gap. A possible solution to this problem might be to restrict the applicability of the Subject Condition to role-assigned subjects. This change has no effect on our analysis of (91), inasmuch as we assume that the subject of easy is itself role-assigned (see footnote 8).
the claim that SLASH is a head feature. As we have seen, there are a variety of problems facing the attempt to explain the behavior of parasitic gaps in terms of Kuno’s Incomplete Constituent Constraint, and these appear to be adequately accounted for by the Subject Condition. The trace theory we have been led to by these considerations is summarized in (94):

(94) Summary of HPSG trace theory (final version):
   a. Nonlocal Feature Principle: For each nonlocal feature, the INHERITED value on the mother is the union of the INHERITED values on the daughters minus the TO-BIND value on the head daughter.
   b. Trace Principle (English): Every trace must be strictly subcategorized by a substantive head.
   c. Subject Condition (English): A lexical head’s SUBCAT list may contain a slashed subject only if it also contains another slashed element.

In Chapter 9, we will return to these matters and consider a revision of this analysis wherein traces, and hence the Trace Principle, are eliminated entirely, their effect being derived from lexical rule application.

4.6 More on Island Constraints

4.6.1 Constraints on Coordinate Structure Extraction

In order to explain the deviance of examples like (95)–(96), Ross proposed that transformations were subject to the Coordinate Structure Constraint (Ross 1967: (4.84)) stated in (97):

(97) Coordinate Structure Constraint
   In a coordinate structure,
   (a) no conjunct may be moved,
   (b) nor may any element contained in a conjunct be moved out of that conjunct.

Following Grosu (1973), we may refer to (97a, b) as the Conjunction Constraint and the Element Constraint, respectively.

Although Ross stipulated (97) as a constraint on variables in transformational rules, subsequent proposals have attempted to derive these generalizations from other principles, as is desirable. The analysis presented above in fact already entails the Conjunction Constraint. Under virtually any assumptions about the nature of coordinate structures, it is the mother of the coordinate structure that is (strictly) subcategorized in examples like (95), not the individual conjuncts. Since a conjunct is never subcategorized for in our theory, it can never be realized as a trace—because of the Trace Principle.

The Element Constraint is another matter altogether. Its factual correctness has been challenged by Goldsmith (1985) and Lakoff (1986), who cite examples like (98) as fully acceptable:

(98) a. How many lakes can we [destroy — and not arouse public antipathy]?
   b. How many kinds of tequila has he [snack off to Mexico, sampled —, and come back the same day without telling anyone]?
   c. Concerts that short, you can leave work early, hear the entirety of —, and still be back at the job before anyone notices you are gone.

Goldsmith’s discussion of examples like (98a) suggests that these exceptions to the Element Constraint might be explained in terms of a semantically coherent class of subordinate modifiers that have only the appearance of coordinate structures. Lakoff, on the other hand, goes so far as to suggest that Element Constraint effects are to be explained entirely on semantic and pragmatic grounds, rather than in terms of a grammatical constraint. The theory of UDCs we have outlined here is perfectly consistent with such an approach.

But suppose, for the sake of argument, that Lakoff is incorrect in denying the correctness of the Element Constraint. How could the Element Constraint be incorporated into our theory of UDCs?

As we have already seen (section 4.5), our HPSG treatment of coordinate structures differs from the GKP analysis in at least two fundamental respects: (1) coordinate structures (in English) are unheaded (cf. P&S-87, p. 56); (2) SLASH is not treated as a HEAD feature. The identities that must hold among

36. We should note that we have made no attempt here to incorporate the suggestion (Chomsky 1982; Cinque 1990) that all parasitic gaps are pronominal in nature. To do so would require substantial modification of the Nonlocal Feature Principle. A modification along these lines (e.g. replacing the set union condition by a condition allowing daughters to bear distinct but coindexed INHER | SLASH values if one is pronominal), though intriguing, is beyond the scope of the present study. We thank Paul Postal for discussion of this point.

37. Goodall (1987) attempts to reduce the Conjunction Constraint to Principle C of the binding theory. But this is both too weak and insufficiently general.
conjunctions thus do not follow from the HFP or from any principle of trace theory (summarized in (94) at the end of the previous section), nor do they follow from any other principles external to coordination theory. Aside from the Conjunct Constraint, which, as we have seen, is derived from the Trace Principle, it is left to the theory of coordination to derive whatever further constraints affect coordinate structures.

The consequence of adopting the principle in (99) would be a strong version of coordination theory: 38

(99) **Coordination Principle (strong version):**

In a coordinate structure, the CATEGORY and NONLOCAL value of each conjunct daughter is identical to that of the mother.

The theory embodied in (99) is strong in that it forces complete categorial identity between the coordinate mother and each conjunct daughter, thereby leaving it to the theory of ellipsis (which might allow *is* or *Kim is* to be omitted from the left periphery of noninitial conjunctions) to provide an account of examples of unlike category coordination like (100):

(100) Kim is a Republican and proud of it.

The strong version of the Coordination Principle makes a number of correct predictions, including obligatory agreement of CASE in coordinate nominals, but not obligatory agreement of person, number, and gender. These predictions follow because CASE is treated as a HEAD feature (and hence is part of the CATEGORY value that must be shared by all conjuncts), while person, number, and gender are treated as features of indices (Chapter 2), and hence are not constrained by (99).

In addition, (99) derives the Element Constraint, as it entails that the INHER/SLASH value of each conjunct is identical to that of the coordinate mother. In fact, (99) also has the effect that the INHER[REL] and INHER[QUE] values of each conjunct must be identical to those of the mother, thus predicting contrasts like the following:

(101) a. Here's the student, [(whose, mother and whose, father) both attended the soccer match].
    b. *Here's the student, [(Hilary and whose, father) both attended the soccer match].

(102) a. I know [(whose mother and whose father) got married in the Poconos].
    b. *I know [(whose mother and his father) got married in the Poconos].

38. The similarity between this principle and the Conjoint Realization Principle of Gazdar and Pullum 1982 should be noted.

The theory of coordination just outlined seems too strong, however, inasmuch as it disallows examples like (103):

(103) a. Francis arrived late today but will be on time tomorrow.
    b. Leslie likes that picture and is trying to buy it.

Here the two conjoined VPs have differing values for the feature AUX, and hence are incorrectly predicted to be ungrammatical.

One approach to this problem would be to accept the central claim of the coordination analysis in Sag et al. 1985, namely that coordinate structures may involve "archicategories," or partially specified feature structures. If such partial structures are allowed, 39 then we may revise the Coordination Principle as follows:

(104) **Coordination Principle (weak version):**

In a coordinate structure, the CATEGORY and NONLOCAL value of each conjunct daughter is subsumed by (is an extension of) that of the mother.

This weak version of the Coordination Principle avoids the dilemma of coordinated auxiliary and nonauxiliary VPs by allowing the coordinate mother to be unspecified for AUX in structures like (105):

(105) \[ \text{VP} \left\{ \text{fin, -AUX} \right\} \]

And structures like this are able to appear in any environment where no condition is imposed on the AUX value of the VP. The formulation in (104) also

39. The use of partially specified feature structures (or, to put it more technically, feature structures that are merely well-typed, but not necessarily resolved or even totally well-typed) to model linguistic objects raises issues of a foundational nature that we cannot address here. Thus far, we have assumed that all linguistic entities are total objects, in the sense that each feature appropriate for a given entity has a value specified; partial feature structures have arisen only as partial models (or incomplete descriptions) of such entities. The kind of coordination analysis under consideration here requires a fundamental philosophical shift, inasmuch as it becomes necessary to countenance linguistic entities (namely, the mothers in coordinate structures) that are inherently partial; at the same time, presumably, we would want to continue to require that 'normal' linguistic entities be total objects. We do not pretend to have laid the necessary foundations for such an analysis.
allows a straightforward account (without appeal to ellipsis) of examples like *Kim is a Republican and proud of it.*

The principle in (104), though weaker than (99), is stronger than it might appear. It guarantees that whenever a syntactic environment imposes some condition on a phrase in a given position X, that condition is respected by every conjunct of a coordinate structure in position X. For example, as we saw in Chapter 3, the SUBCAT specification of a raising verb like believe contains a post-object complement (X in (106)) that is an infinitival VP whose SUBCAT element is required (by the Raising Principle) to be structure-shared with the SYNSEM value of the object NP:

\[
\begin{array}{c}
\text{VP} \longrightarrow \text{S}[	ext{fin}] \\
\text{[SUBCAT (1|NP)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \longrightarrow \text{S}[	ext{fin}] \\
\text{[SUBCAT (1|NP)]} \\
\end{array}
\]

In consequence of (104), if X is realized as a coordinate structure, whatever is required to be true of X (in consequence of the SUBCAT list of believe and the Subcategorization Principle) must also be true of each conjunct, as shown in (107):

\[
\begin{array}{c}
\text{VP} \longrightarrow \text{S}[	ext{fin}] \\
\text{[SUBCAT (NP)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \longrightarrow \text{S}[	ext{fin}] \\
\text{[SUBCAT (NP)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \longrightarrow \text{S}[	ext{fin}] \\
\text{[SUBCAT (NP)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{VP} \longrightarrow \text{S}[	ext{fin}] \\
\text{[SUBCAT (NP)]} \\
\end{array}
\]

Thus the examples in (108) are correctly ruled out:

\[
\begin{array}{l}
\text{(108)} \quad a. *\text{Jessie believes Tracy to be happy and walks}. \\
\text{b. *Jessie believes Tracy happy and to be healthy}. \\
\text{c. *Jessie believes Tracy am walking to the store and that I left}. \\
\end{array}
\]

In much the same way, this weak version of the Coordination Principle still entails the Element Constraint. The HFP and the information specified in Schema 6 together require that in structures like (109), the head daughter must be a finite S that is further specified as INHER|SLASH (1):

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

The Coordination Principle in turn guarantees that each conjunct of a coordinate head of (109) also be so specified, as illustrated in (110):

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

This rules out all Element Constraint violations, as desired, including those involving the features REL and QUE.

In this section, we have explored the consequences of various formulations of coordination theory relevant to the analysis of constraints on coordinate structure extraction. While the exact set of facts to be accounted for remains somewhat unclear (because of the uncertain status of the Element Constraint), we have shown how our analysis can be adapted to derive the Element Constraint or not, all the while deriving the Conjunct Constraint from the Trace Principle, which is independently motivated, as we have seen.

4.6.2 Some Complex Noun Phrases

The fact that filler-gap dependencies cannot penetrate into relative clauses has long been known. To account for this fact, illustrated by the ungrammatical examples in (111), Ross (1967) stipulated that (a certain class of) transformations were subject to his Complex NP Constraint (CNPC), which barred movement out of a clause adjoined to a nominal constituent:

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]

\[
\begin{array}{c}
\text{S}[	ext{fin}] \\
\text{[INHER|SLASH (1)]} \\
\end{array}
\]
(111)  
  a. *Which book do you know [the person [who wrote —]]?  
  b. *Which book do you know [the person [who Allison talked to  
  — about —]]?

Ross's theory, like many subsequent theories that attempt to derive the CNPC  
as a theorem, predicts that factive (fact that) clauses are also extraction is-  
lands. However, once one looks past the definiteness factors that are at work in  
examples like (112), it is not clear that there should be any grammatical con-  
straint barring extraction from a factive complement:

(112)  
  a. ?*Here is the book that Leslie denied [the claim [that Kris had  
  written —]].  
  b. ?*Which book did Dana make [the suggestion [that we should  
  read —]]?

In particular, examples like (113), which are ruled out in some theories of  
UDCs, including Ross's, seem to be impeccable:

(113)  
  a. Which rebel leader, would you favor [a proposal [that the CIA  
  assassinate —]]?  
  b. Which Middle East country, did you hear [rumors [that we had  
  infiltrated —]]?

If nothing further is said, our analysis of filler-gap dependencies predicts that  
factive clauses allow extraction dependencies of this sort. We assume that  
such clauses are optionally subcategorized for by nouns like fact, claim, and  
rumor, with which they form an N constituent. Assuming this complement  
analysis is correct, then there is no grammatical principle that prevents the  
inheritance of nonempty INHER | SLASH values sketched in (114):

(114)  

40. This fact is noted by Maling and Zaeon (1982). See also Chang and McCloskey  
(1983: 706) and Chomsky's (1986b) treatment in terms of L-marking.

As just noted, this appears to be the right prediction, modulo nongrammatical  
factors such as definiteness.

The remaining data that are usually discussed under the rubric of the Com-  
plex NP Constraint concern the impossibility of extracting out of (certain kinds  
of) relative clauses. We will take up that matter in section 5.2.4, where we  
present in detail our analysis of relative clauses in a variety of languages.

4.7 Conclusion

In this chapter, we have developed a theory of unbounded dependencies in  
English that provides the basis for an account of topicalization constructions,  
extraction in relative and interrogative clauses, and tough-constructions. The  
theory in turn provides an account of constraints on the distribution of traces,  
so-called that-trace facts, and the behavior of parasitic gaps. The essence of  
our account is the interaction of two principles of universal grammar (the Non-  
local Feature Principle and the Trace Principle) and a language-particular con-  
straint on SUBCAT lists (the Subject Condition).

As we have seen, these principles incorporate important insights from di-  
verse frameworks, including GKPS's Foot Feature Principle and other GPSG  
proposals due to Jacobson, Farkas et al., Sells, Fodor, and Hukari and Levine,  
as well as other principles proposed within GB, most notably Chomsky's ECP.  
Although we have incorporated insights such as these, we have also indicated  
why we believe our analysis to be both empirically more adequate and conceptu-  
ally simpler than previous attempts. Yet the integrative approach we have  
taken in this domain illustrates well the synthetic methodology that we believe  
is most productive in modern syntactic research.